

Lecture Notes
Linear Algebra

Prof. Boris Houska

Contents

1	Linear Equations	3
2	Matrices	63
3	Determinants	133
4	Vector Spaces	194
5	Linear Transformations	239
6	Euclidean Spaces	306
7	Eigenvalues and Eigenvectors	373
8	Quadratic Forms	472
9	Singular Value Decomposition	508

Chapter 1

Linear Equations

Linear Equations

- Introduction
- Scalar Linear Equations
- Linear Equations with Two Variables
- General Linear Equation Systems
- Gaussian Elimination
- Structure of the Solution Set
- Summary

Contents

- Introduction
- Scalar Linear Equations
- Linear Equations with Two Variables
- General Linear Equation Systems
- Gaussian Elimination
- Structure of the Solution Set
- Summary

Objectives

In this lecture we will learn about

- basic set notation,
- scalar linear equations,
- linear equations with two variables,
- matrices and vectors, and
- general linear equations and their solutions.

Notation

- We try to get used to set notation early on.
- We use the symbol $\mathbb{R}$ to denote the set of real numbers.
- If we write

$$X = \{1, 2, 3\},$$

this means: X is a set, consisting of the numbers 1, 2, and 3.

- If we write

$$X = \{ x \in \mathbb{R} \mid 3x = 2 \},$$

we read: “ X is the set of real numbers x that satisfy $3x = 2$.”

Notation

- The empty set is denoted by $\emptyset$.
- If X and Y are sets, the set of pairs (x, y) with $x \in X$ and $y \in Y$ is denoted by $X \times Y$.
- For example, if $X = \{1, 2, 3\}$ and $Y = \{4, 5\}$, we have

$$X \times Y = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}.$$

- If all elements of a set X are also elements of a set Y , we write $X \subseteq Y$.
- For example, $\{1, 3\} \subseteq \{1, 2, 3, 5\}$.

Contents

- Introduction
- **Scalar Linear Equations**
- Linear Equations with Two Variables
- General Linear Equation Systems
- Gaussian Elimination
- Structure of the Solution Set
- Summary

Scalar Linear Equation

Scalar linear equations have the form

$$ax = b.$$

- Here, a, b are given numbers, usually real numbers $a, b \in \mathbb{R}$.
- We are looking for all numbers x for which this equation is true.
- This means that we want to find the solution set

$$X = \{ x \in \mathbb{R} \mid ax = b \}.$$

Scalar Linear Equation

There are three cases:

- Case 1: we have $a \neq 0$. In this case: $X = \{b/a\}$.
- Case 2: we have $a = 0$ and $b = 0$. In this case: $X = \mathbb{R}$.
- Case 3: we have $a = 0$ but $b \neq 0$. In this case: $X = \emptyset$.

Contents

- Introduction
- Scalar Linear Equations
- **Linear Equations with Two Variables**
- General Linear Equation Systems
- Gaussian Elimination
- Structure of the Solution Set
- Summary

Two Linear Equations

Let us consider the linear equation system

$$ax + by = e$$

$$cx + dy = f.$$

- Here, $a, b, c, d, e, f \in \mathbb{R}$ are given real numbers.
- We are looking for all pairs (x, y) for which both equations are true.
- In other words, we want to find the solution set

$$X = \{ (x, y) \in \mathbb{R} \times \mathbb{R} \mid ax + by = e \text{ and } cx + dy = f \}.$$

Two Linear Equations

Example:

$$2x + 3y = 10$$

$$4x + 5y = 18$$

Strategy:

- Solve the first equation with respect to x ,

$$x = 5 - \frac{3}{2}y.$$

- Substitute this relation into the second equation,

$$4\left(5 - \frac{3}{2}y\right) + 5y = 18 \implies y = 2 \implies X = \{(2, 2)\}.$$

Two Linear Equations

Let us try to do the same for the general case

$$ax + by = e$$

$$cx + dy = f.$$

- If $a \neq 0$, we can solve the first equation with respect to x ,

$$x = \frac{e - by}{a}.$$

- And we substitute this result into the second equation

$$c \left(\frac{e - by}{a} \right) + dy = f \implies (ad - bc)y = af - ce.$$

Two Linear Equations (cont.)

- Thus, if $a \neq 0$ and $ad - bc \neq 0$, we find

$$y = \frac{af - ce}{ad - bc}$$

and

$$\begin{aligned} x &= \frac{e - by}{a} = \frac{e}{a} - \frac{b}{a} \frac{af - ce}{ad - bc} \\ &= \frac{ead - ebc - abf + bce}{a(ad - bc)} = \frac{de - bf}{ad - bc}. \end{aligned}$$

- Thus, if $a \neq 0$ and $ad - bc \neq 0$, the solution set is given by

$$X = \left\{ \left(\frac{de - bf}{ad - bc}, \frac{af - ce}{ad - bc} \right) \right\}.$$

Two Linear Equations (cont.)

- Interesting observation: as long as $ad - bc \neq 0$,

$$x = \frac{de - bf}{ad - bc} \quad \text{and} \quad y = \frac{af - ce}{ad - bc}$$

always solves both linear equations, also if $a = 0$.

- Just substitute the above expressions into the equations

$$ax + by = e$$

$$cx + dy = f$$

to prove this. (Exercise!)

- Thus, in summary, as long as $ad - bc \neq 0$, the solution set is given by

$$X = \left\{ \left(\frac{de - bf}{ad - bc}, \frac{af - ce}{ad - bc} \right) \right\}.$$

Two Linear Equations (cont.)

- It remains to discuss the case that $ad - bc = 0$.
- In this case, we can subtract c times the first equation in

$$ax + by = e$$

$$cx + dy = f$$

from a times the second equation. This yields

$$\underbrace{(bc - ad)}_{=0} y = af - ce$$

- Thus, if $ad - bc = 0$ and $af - ce \neq 0$, then $X = \emptyset$.

Two Linear Equations (cont.)

What if $ad - bc = 0$ and $af - ce = 0$? There are again several cases:

- If $a \neq 0$, the solution set is a “line”,

$$X = \left\{ \left(\frac{e - by}{a}, y \right) \mid y \in \mathbb{R} \right\}.$$

- If $c \neq 0$, the “same line” is given by,

$$X = \left\{ \left(\frac{f - dy}{c}, y \right) \mid y \in \mathbb{R} \right\}.$$

- If $a = c = 0$, we have to satisfy two scalar equations:

$$by = e \quad \text{and} \quad dy = f.$$

Two Linear Equations: Summary

Case	X
$ad - bc \neq 0$	$\left\{ \left(\frac{de-bf}{ad-bc}, \frac{af-ce}{ad-bc} \right) \right\}$
$ad - bc = 0, af - ce = 0, a \neq 0$	$\left\{ \left(\frac{e-by}{a}, y \right) \mid y \in \mathbb{R} \right\}$
$ad - bc = 0, af - ce = 0, c \neq 0$	$\left\{ \left(\frac{f-dy}{c}, y \right) \mid y \in \mathbb{R} \right\}$
$a = c = 0, b = e = 0, d \neq 0$	$\left\{ \left(x, \frac{f}{d} \right) \mid x \in \mathbb{R} \right\}$
$a = c = 0, d = f = 0, b \neq 0$	$\left\{ \left(x, \frac{e}{b} \right) \mid x \in \mathbb{R} \right\}$
$a = b = c = d = e = f = 0$	$\mathbb{R} \times \mathbb{R}$
All other cases	$\emptyset$

General Observations

- The term “ $ad - bc$ ” seems to be very important:

The linear equation system

$$ax + by = e$$

$$cx + dy = f$$

has a unique solution if and only if $ad - bc \neq 0$.

- The solution set is either a plane, or a line, or a singleton, or empty.

There are no other possibilities.

Contents

- Introduction
- Scalar Linear Equations
- Linear Equations with Two Variables
- **General Linear Equation Systems**
- Gaussian Elimination
- Structure of the Solution Set
- Summary

General Linear Equation System

Let us consider the general linear equation system

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m.$$

- There are $m \in \mathbb{N}$ equations and
- there are $n \in \mathbb{N}$ variables, namely, $x_1, x_2, \dots, x_n$.
- In general, we might have $m \neq n$.

Matrices

In order to avoid writing out all equations one often merely collects the coefficient matrix,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ & \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}.$$

Here, a matrix is simply a “rectangular table of numbers”.

Matrices

Sometimes, we don't need the last column. In this case, one considers the thinner matrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Matrices

Definitions:

- A matrix with 1 column is called a “vector” or “column vector”.
- A matrix with 1 row is called a “row vector”.
- A matrix m rows and n columns is called a “ $(m \times n)$ -matrix”.

Examples:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad (4 \ 5 \ 6 \ 7), \quad \text{and} \quad \begin{pmatrix} 3 & 8 & 1 \\ 7 & 5 & 17 \end{pmatrix},$$

are a vector of length 3, a row vector of length 4, and a (2×3) -matrix.

Special Matrices

Definition: A matrix $(m \times n)$ -matrix is called

1. a square matrix if $m = n$,
2. a diagonal matrix if $m = n$ and all non-diagonal elements are zero,
3. an upper triangular matrix if all entries below the diagonal are 0.

Examples:

$$\underbrace{\begin{pmatrix} 1 & 4 & 7 \\ 2 & 0 & 9 \\ 1 & 0 & 3 \end{pmatrix}}_{\text{square}}, \quad \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}}_{\text{diagonal}}, \quad \text{and} \quad \underbrace{\begin{pmatrix} 3 & 8 & 1 & 2 \\ 0 & 5 & 17 & 4 \\ 0 & 0 & 7 & 9 \end{pmatrix}}_{\text{upper triangular}}.$$

Convention: The diagonal of a matrix always refers in this course always to the elements a_{ii} on the main diagonal, not $a_{i,n+1-i}$.

Row Echelon Form

Definition: A matrix $(m \times n)$ -matrix is in (sorted) row echelon form if

1. the first non-zero element in each row, called leading entry (LE), is 1,
2. each LE is in a column (one) to the right of the previous row's LE,
3. zero rows, if any, are below rows having a non-zero element.

$$\begin{pmatrix} 1 & a_{12} & \dots & a_{1r} & a_{1,r+1} & \dots & a_{1n} \\ 0 & 1 & \dots & a_{2r} & a_{2,r+1} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & a_{r,r+1} & \dots & a_{rn} \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \quad \text{with } 1 \leq r \leq m.$$

Row Echelon Form

Remarks:

- The matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 6 \\ 0 & 0 & 1 & 5 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is in row echelon form, but not using the above column sorting.

- The matrix

$$\begin{pmatrix} 1 & 3 & 2 & 4 & 6 \\ 0 & 1 & 0 & 5 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is in sorted row echelon form (our way to sort things here).

- Textbooks: different authors use different conventions...

Block Matrix Notation

- It is sometimes convenient to write matrices in block notation.
- For example, the matrix

$$\left(\begin{array}{cccc|ccc} 1 & a_{12} & \dots & a_{1r} & a_{1,r+1} & \dots & a_{1n} \\ 0 & 1 & \dots & a_{2r} & a_{2,r+1} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & a_{r,r+1} & \dots & a_{rn} \\ \hline 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{array} \right) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$$

has four blocks: A upper triangular, B general, $C = 0$ and $D = 0$.

General Linear Equation System

Let us come back to our general linear equation system

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m.$$

- Assume that the coefficient matrix is in sorted row echelon form.
- If we don't have $b_{r+1} = \dots = b_m = 0$, there is no solution.
- Thus, let us also assume that we have $b_{r+1} = \dots = b_m = 0$.

Triangular Linear Equation System

Under the above assumptions, our equation system has the form

$$\begin{aligned}x_1 + a_{12}x_2 + \dots + a_{1r}x_r + a_{1,r+1}x_{r+1} + \dots + a_{1n}x_n &= b_1 \\x_2 + \dots + a_{2r}x_r + a_{2,r+1}x_{r+1} + \dots + a_{2n}x_n &= b_2 \\&\vdots \\x_r + a_{r,r+1}x_{r+1} + \dots + a_{rn}x_n &= b_r.\end{aligned}$$

Solution Strategy:

- Start with any numbers $x_{r+1}, \dots, x_n \in \mathbb{R}$.
- Solve the r -th equation with respect to x_r .
- Substitute into x_r into $(r-1)$ -th equation and solve w.r.t. x_{r-1} .
- Keep iterating backwards, until $x_r, x_{r-1}, \dots, x_1$ is found.

Triangular Linear Equation System

Summary

- The linear equation system

$$x_1 + a_{12}x_2 + \dots + a_{1r}x_r + a_{1,r+1}x_{r+1} + \dots + a_{1n}x_n = b_1$$

$$x_2 + \dots + a_{2r}x_r + a_{2,r+1}x_{r+1} + \dots + a_{2n}x_n = b_2$$

$$\vdots$$

$$x_r + a_{r,r+1}x_{r+1} + \dots + a_{rn}x_n = b_r$$

has for any given numbers $x_{r+1}, \dots, x_n \in \mathbb{R}$, a unique solution for $x_1, x_2, \dots, x_r \in \mathbb{R}$. It can be found by running the iteration

$$x_k = b_k - a_{k,k+1}x_{k+1} - \dots - a_{kn}x_n$$

backwards; that is, for $k = r$, $k = r - 1$, and so on, until $k = 1$.

Contents

- Introduction
- Scalar Linear Equations
- Linear Equations with Two Variables
- General Linear Equation Systems
- **Gaussian Elimination**
- Structure of the Solution Set
- Summary

General Linear Equation System

Gaussian elimination is a procedure for transforming the system

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

into a linear equation system in (sorted) row echelon form.

Elementary Row Operations

Recall: coefficient matrix of the linear equation system denoted by

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ & \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}.$$

The solution set does not change if

1. we add one row to another, or if
2. we multiply one row elementwise with a scalar $s \neq 0$, or if
3. we permute two rows.

Column Permutations

- We may rename the variables in an equation system.
- For example, instead of solving

$$\left\{ \begin{array}{rcl} ax + by & = & e \\ cx + dy & = & f \end{array} \right\}, \quad \text{we could solve} \quad \left\{ \begin{array}{rcl} ay + bx & = & e \\ cy + dx & = & f \end{array} \right\}$$

as long as we keep in mind that we exchange the role of x and y .

- This is the same as permuting columns of the coefficient matrix

$$\left(\begin{array}{ccc} a & b & e \\ c & d & f \end{array} \right), \quad \text{Permute columns 1 and 2} \longrightarrow \left(\begin{array}{ccc} b & a & e \\ d & c & f \end{array} \right)$$

- In general, we could permute the i -th and j th column.

Gaussian Elimination

Main idea:

- Apply elementary row operations and column permutations to

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right).$$

until the left block matrix “ A ” is in sorted row echelon form.

Theorem:

- We can bring the matrix A in row echelon form by applying a finite number of elementary row operations and column permutations.

Gaussian Elimination

Strategy:

1. If $A \neq 0$, permute rows and columns until $a_{11} \neq 0$; else, we're done.
2. For all $i \geq 2$ subtract $\frac{a_{i1}}{a_{11}}$ times the first row from the i -th row. Then

$$\left(\begin{array}{c|ccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ \hline 0 & c_{22} & \dots & c_{2n} & d_2 \\ & \vdots & & \vdots & \vdots \\ 0 & c_{m2} & \dots & c_{mn} & d_m \end{array} \right).$$

3. Set $A \leftarrow C$ and $b \leftarrow d$. Go to Step 1 until all blocks are processed.

Proof of the theorem: the above strategy always leads to a row echelon form, no conditions on the coefficients needed!

Gaussian Elimination

Example:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{array} \right)$$

Subtract 2 times the first row from the second row.

Gaussian Elimination

Example:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 3 & 6 & 9 & 12 \end{array} \right)$$

Subtract 3 times the first row from the third row.

Gaussian Elimination

Example:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 10 \\ \hline 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{array} \right)$$

Subtract 2 times the second row from the third row.

Gaussian Elimination

Example:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 10 \\ \hline 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Scale the second row.

Gaussian Elimination

Example:

$$\left(\begin{array}{cc|cc} 1 & 4 & 7 & 10 \\ 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 0 & 0 \end{array} \right)$$

This matrix is in row echelon form!

Gaussian Elimination

Example–Summary: The linear equation system

$$x + 4y + 7z = 10$$

$$2x + 5y + 8z = 11$$

$$3x + 6y + 9z = 12$$

is equivalent to the linear equation system

$$x + 4y + 7z = 10$$

$$y + 2z = 3.$$

For any given $z \in \mathbb{R}$, we can use a backward iteration to find

$$y = 3 - 2z \quad \text{and} \quad x = z - 2.$$

Gaussian Elimination

Example–Summary: The linear equation system

$$x + 4y + 7z = 10$$

$$2x + 5y + 8z = 11$$

$$3x + 6y + 9z = 12$$

has the solution set

$$X = \left\{ \left(\begin{array}{c} z - 2 \\ 3 - 2z \\ z \end{array} \right) \middle| z \in \mathbb{R} \right\}.$$

It can be visualized as a “line” in a three dimensional coordinate system, with coordinates x , y , and z .

Gauss-Jordan Elimination

- The goal of a Gaussian elimination is to perform elementary row operations and column permutations until the given matrix is in sorted row echelon form.
- In order to find the actual solution of the linear equation system, we then still have to the recursive backward substitution.
- Alternatively, one can perform the backward substitution directly on the coefficient scheme. This means that we additionally eliminate all non-zero elements above the leading 1s in the sorted row-echelon form. The corresponding methods is known under the name Gauss-Jordan elimination.

Reduced Row Echelon Form

- A matrix is in reduced row echelon form if it is in sorted row echelon form and, additionally, all entries above the leadings 1s are 0:

$$\left(\begin{array}{cccc|ccc} 1 & 0 & \dots & 0 & a_{1,r+1} & \dots & a_{1n} \\ 0 & 1 & \ddots & \vdots & a_{2,r+1} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & a_{r,r+1} & \dots & a_{rn} \\ \hline 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{array} \right)$$

Gauss-Jordan Elimination

Example

- In our previous example, we have used Gaussian elimination to write our coefficient matrix in sorted row echelon form

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Instead of doing a backward substitution, we can subtract 4 times the second row from the first row:

$$\begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \implies X = \left\{ \begin{pmatrix} -2 + z \\ 3 - 2z \\ z \end{pmatrix} \mid z \in \mathbb{R} \right\}.$$

Contents

- Introduction
- Scalar Linear Equations
- Linear Equations with Two Variables
- General Linear Equation Systems
- Gaussian Elimination
- **Structure of the Solution Set**
- Summary

Vectors

- The solution of a linear equation system is a set of n -tuples,

$$(x_1, x_2, \dots, x_n).$$

- The set of real-valued n -tuples is denoted by $\mathbb{R}^n$.
- We can sort the n -tuple into a column, called “vector”,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

- The subtle differences between n -tuples and vectors will be discussed a bit later in this course (see our lecture on vector spaces).

Sums of Vectors

- The sum of two vectors (or n -tuples) is defined by

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

- Thus, the $+$ operator maps two elements of the set $\mathbb{R}^n$ to another element of $\mathbb{R}^n$. That is,

$$+ : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n.$$

Product of Real Number and Vector

- The product of an element $C \in \mathbb{R}$ with a vector is defined by

$$C \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} Cx_1 \\ Cx_2 \\ \vdots \\ Cx_n \end{pmatrix}$$

- Thus, the $\cdot$ operator maps an element of $\mathbb{R}$ and elements of $\mathbb{R}^n$ to another element of $\mathbb{R}^n$. That is,

$$\cdot : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n.$$

- In practice, we often don't write out the " $\cdot$ ".

Homogeneity

Definition

- The linear equation system

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

is called homogeneous if $b_1 = b_2 = \dots = b_m = 0$.

Observation

- Homogeneous linear equation systems have always at least one solution. Namely, $x_1 = x_2 = \dots = x_m = 0$.

Difference of Solutions

Observation:

- Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be solutions of a linear equation system with coefficient matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right).$$

Then, $x - y$ is a solution of the homogeneous linear equation system with coefficient matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & 0 \\ & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & 0 \end{array} \right).$$

Homogeneous Linear Equations

Consequence:

- If $x \in \mathbb{R}^n$ is a solution of the linear equation system with coefficient matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right),$$

then all other solutions can be found by adding a solution of its associated homogeneous linear equation system to x .

Homogeneous Linear Equations

Observations:

- Homogeneous linear equation systems are easier to analyze than inhomogenous ones.
- If x and y are solutions of a homogeneous linear equation system, then $x + y$ is also a solution.
- Similarly, if x is a solution, then Cx is a solution, for all $C \in \mathbb{R}$.

Proof: one of our homework exercises!

Linear Subspace

Definition:

A non-empty set $S \subseteq \mathbb{R}^n$ is called a linear subspace of $\mathbb{R}^n$ if

1. we have $x + y \in S$ for all $x, y \in S$, and
2. we have $Cx \in S$ for all $x \in S$ and all $C \in \mathbb{R}$.

Theorem:

- If S is a linear subspace of $\mathbb{R}^n$, then $0 \in S$.
- The solution set of a homogeneous linear equation system with m equations and n variables is a linear subspace of $\mathbb{R}^n$.

Affine Subspace

- The sum of a given $x \in \mathbb{R}^n$ and a linear subspace $S \subseteq \mathbb{R}^n$,

$$x + S \stackrel{\text{def}}{=} \{ x + y \mid y \in S \},$$

is called an affine subspace of $\mathbb{R}^n$.

- Together with the previous comments, we find that the solution set of a general linear equation system with m equations and n variables is either empty or an affine subspace of $\mathbb{R}^n$.
- In principle, we have now understood the basic properties of the solution set of linear equations. More details about the subspace S and its “dimension”, will be introduced in Lecture 2.

Affine Subspace

Example:

- Recall that the solution set of the linear equations

$$x + 4y + 7z = 10$$

$$2x + 5y + 8z = 11$$

$$3x + 6y + 9z = 12$$

is given by $x = -2 + z$ and $y = 3 - 2z$, where z can be any number.

- For $z = 0$, we find the solution triple $(-2, 3, 0)$.
- Thus, the solution set has the form

$$X = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix} + S \quad \text{with} \quad S = \left\{ z \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \mid z \in \mathbb{R} \right\}.$$

Contents

- Introduction
- Scalar Linear Equations
- Linear Equations with Two Variables
- General Linear Equation Systems
- Gaussian Elimination
- Structure of the Solution Set
- **Summary**

Summary

In this lecture we have learned about:

- basic set notation,
- definition of vectors and matrices,
- the (sorted) row echelon form of a matrix,
- solving general linear equation systems by Gaussian elimination, and
- the affine subspace structure of the solution set of linear equations.

Chapter 2

Matrices

Matrices

- Introduction
- Linear Dependence and Independence
- Rank of a Matrix
- Matrix Multiplication
- Invertible Matrices
- Elementary Matrices
- Summary

Contents

- Introduction
- Linear Dependence and Independence
- Rank of a Matrix
- Matrix Multiplication
- Invertible Matrices
- Elementary Matrices
- Summary

Objectives

In this lecture we will learn about the basic properties of matrices.

Topics include:

- linear dependence and independence
- rank of a matrix
- multiplication of matrices
- elementary matrices, and
- invertible matrices.

Contents

- Introduction
- **Linear Dependence and Independence**
- Rank of a Matrix
- Matrix Multiplication
- Invertible Matrices
- Elementary Matrices
- Summary

Linear Dependence

Definition:

- The m -tuple of vectors $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ is called linearly dependent if there exist numbers $C_1, C_2, \dots, C_m \in \mathbb{R}$, which are not all 0, such that

$$C_1 a_1 + C_2 a_2 + \dots C_m a_m = 0.$$

Caution:

- The a_i s are vectors, not components of a vector!

Linear Independence

Definition:

- The m -tuple of vectors $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ is called linearly independent if they are not linearly dependent.

Example:

- The vectors

$$a_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad a_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

are linearly independent: the equations $C_2 = 0$ and $C_1 + C_2 = 0$ admit only the trivial solution $C_1 = C_2 = 0$.

Linear Independence

More examples:

- The vectors

$$a_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and} \quad a_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

are linearly dependent, since $a_1 - a_2 + a_3 = 0$.

- The vectors

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad a_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are linearly independent. Why?

Unit Vectors

Definition:

- The vectors $e_1, e_2, \dots, e_n \in \mathbb{R}^n$, given by

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

are called the unit vectors of $\mathbb{R}^n$.

Observations:

- If $x \in \mathbb{R}^n$ is any vector, then $x = x_1 e_1 + x_2 e_2 + \dots + x_n e_n$.
- The unit vectors are linearly independent.

Maximum Number of Independent Vectors

- Recall from Lecture 1: the homogeneous linear equation system

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = 0$$

always has a solution at $x_1 = x_2 = \dots = x_n = 0$.

- If we have $m < n$, we additionally know that there is at least one solution for x that is not the zero vector. This follows by writing the coefficient matrix in row echelon form and using that $m < n$ means that we have more variables than equations.

Maximum Number of Independent Vectors

- Next, let us introduce the column vectors

$$v_1 \stackrel{\text{def}}{=} \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad v_2 \stackrel{\text{def}}{=} \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad \dots, \quad v_n \stackrel{\text{def}}{=} \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

- Then the homogeneous linear equation system can be written as

$$x_1 v_1 + x_2 v_2 + \dots + x_n v_n = 0.$$

And, if $m < n$, this equation has a solution for x that is not all zeros.

- Consequently, if $m < n$, any collection of n vectors $v_1, \dots, v_n \in \mathbb{R}^m$ is linearly dependent!

Maximum Number of Independent Vectors

Theorem:

- Any collection of $n + 1$ or more vectors in $\mathbb{R}^n$ is linearly dependent.

Definition:

- A maximal collection of linearly independent vectors of a linear subspace $S \subseteq \mathbb{R}^n$ is called a basis of S .

Remark:

- Concerning the special case $S = \mathbb{R}^n$, the unit vectors $e_1, \dots, e_n$ are linearly independent. They are also a maximal collection of independent vectors, since any collection of $n + 1$ vectors is linearly dependent. Thus, the unit vectors are a basis of $\mathbb{R}^n$.

Basis and Dimension

Theorem:

- Let $S \subseteq \mathbb{R}^n$ be a linear subspace of $\mathbb{R}^n$ and $b_1, \dots, b_r \in S$ a basis.

Let $c_1, \dots, c_s \in S$ be another basis of S . Then we have $r = s$.

Proof:

- Since $b_1, \dots, b_r$ is a maximal collection of independent vectors of S , the vectors $b_1, \dots, b_r, c_i$ are linearly dependent. Thus, there exist numbers $a_{1i}, a_{2i}, \dots, a_{ri}, a_{r+1,i} \in \mathbb{R}$, not all zero, such that

$$a_{1i}b_1 + a_{2i}b_2 + \dots + a_{ri}b_r + a_{r+1,i}c_i = 0.$$

Additionally, we know that $a_{r+1,i} \neq 0$, since the b_i s are a basis.

Hence, we may assume $a_{r+1,i} = -1$, after scaling appropriately.

Basis and Dimension

Proof (continued):

- The above relation can be summarized as

$$c_i = a_{1i}b_1 + a_{2i}b_2 + \dots + a_{ri}b_r \quad (1 \leq i \leq s)$$

- Suppose that $s > r$, then the homogeneous linear equation system

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1s}x_s = 0$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2s}x_s = 0$$

$$\vdots$$

$$a_{r1}x_1 + a_{r2}x_2 + \dots + a_{rs}x_s = 0$$

has at least one non-zero solution x .

Basis and Dimension

Proof (continued):

- The previous constructions imply that

$$\begin{aligned}x_1c_1 + x_2c_2 + \dots + c_sx_s &= x_1a_{11}b_1 + x_1a_{21}b_2 + \dots + x_1a_{r1}b_r \\&\quad \vdots \\&\quad + x_sa_{1s}b_1 + x_sa_{2s}b_2 + \dots + x_sa_{rs}b_r \\&= 0.\end{aligned}$$

But this contradicts our assumptions that the c_i s are a basis.

- Thus, we must have $s \leq r$. But if $s < r$, we exchange the roles of b and c and get again a contraction. Consequently, we must have $s = r$ and this completes our proof!

Basis and Dimension

Definition:

- The above theorem implies that every linear subspace $S \subseteq \mathbb{R}^n$ has a basis and that the number of basis vectors is unique. In the following, we call this unique number of basis vectors of S the dimension of S .

Consequence:

- The dimension of $\mathbb{R}^n$ is n , because the n unit vectors are a basis.

Basis and Dimension

Example:

- In Lecture 1 we have shown that the solution set of

$$x + 4y + 7z = 10$$

$$2x + 5y + 8z = 11$$

$$3x + 6y + 9z = 12$$

is given by

$$X = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} + S \quad \text{with} \quad S = \left\{ z \cdot \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix} \mid z \in \mathbb{R} \right\}.$$

- The vector with coefficients $(-1, -2, 1)$ is a basis vector of the linear subspace S . Consequently, S has dimension 1.

Spans of Vectors

- Let $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ be given vectors.
- The span of these vectors is defined as the set of linear combinations,

$$\text{span}(a_1, a_2, \dots, a_m) \stackrel{\text{def}}{=} \{C_1 a_1 + \dots + C_m a_m \mid C_1, \dots, C_m \in \mathbb{R}\}.$$

- Obviously, $S = \text{span}(a_1, a_2, \dots, a_m)$ is a linear subspace of $\mathbb{R}^n$.
- If $a_1, a_2, \dots, a_r \in \mathbb{R}^n$ and $b_1, b_2, \dots, b_s \in \mathbb{R}^n$ are given vectors, then

$$\text{span}(a_1, a_2, \dots, a_r) \subseteq \text{span}(b_1, b_2, \dots, b_s)$$

if and only if each of the a_i s is a linear combination of the b_j s.

- If $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ are linearly independent vectors, then they are a basis of the linear subspace $S = \text{span}(a_1, a_2, \dots, a_m)$; $\dim(S) = m$.

Contents

- Introduction
- Linear Dependence and Independence
- **Rank of a Matrix**
- Matrix Multiplication
- Invertible Matrices
- Elementary Matrices
- Summary

Row and Column Rank

Definition:

- The dimension of the linear subspace that is spanned by the row vectors of a matrix is called its row rank.
- The dimension of the linear subspace that is spanned by the column vectors of a matrix is called its column rank.

The goal of this section is to show that the row- and column rank of any given matrix coincide:

$$\text{“row rank} = \text{column rank”}$$

Row and Column Rank

Example:

- Consider the matrix

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

- The first and second row vectors are linearly independent. Thus, the row rank of this matrix is equal to 2.
- The first and the last column are unit vectors and, consequently, linearly independent. But the second column vector is the sum of the first and the last column vector. Thus, the column rank of this matrix is also equal to 2.

Elementary Operation

Observation: neither the row- nor the column rank changes if

1. we add one row to another, or if
2. we multiply one row elementwise with a scalar $s \neq 0$, or if
3. we permute two rows, or if
4. we add one column to another, or if
5. we multiply one column elementwise with a scalar $s \neq 0$, or if
6. we permute two columns.

Operations 1-3 are called elementary row operations.

Operations 4-6 are called elementary column operations.

Complete Gaussian Elimination

- In Lecture 1, we have discussed Gaussian elimination. It uses elementary row-operations and column permutations to bring any given matrix in the row-echelon form

$$\left(\begin{array}{cccc|ccc} 1 & a_{12} & \dots & a_{1r} & a_{1,r+1} & \dots & a_{1n} \\ 0 & 1 & \dots & a_{2r} & a_{2,r+1} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & a_{r,r+1} & \dots & a_{rn} \\ \hline 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{array} \right) .$$

Complete Gaussian Elimination

- By subtracting multiples of first column from all other columns:

$$\left(\begin{array}{cccc|ccc} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & a_{2r} & a_{2,r+1} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & a_{r,r+1} & \dots & a_{rn} \\ \hline 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{array} \right) .$$

Complete Gaussian Elimination

- Repeating this process for all subsequent rows yields:

$$\left(\begin{array}{cccc|ccc} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \hline 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{array} \right) = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array} \right).$$

- This matrix has row rank r and column rank r (there are r ones on the diagonal or E_r).

Rank

Theorem:

- The row rank r of a matrix coincides with its column rank. Therefore, in the following, we simply call the number r the rank of the matrix.

Important:

- The rank of a matrix is invariant under elementary operations. In particular, it doesn't matter in which order we perform these operations during a Gaussian- or Gauss-Jordan elimination process. The final result is always the same.

We will meet many “invariants” in this course, in many different situations. They are our friends!

Linear Equation Systems

Consequence:

- Consider a linear equation system with coefficient matrix

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

- The linear equation system has at least one solution x if and only if

$$\text{rank}(A) = \text{rank}\left(\left(A \mid b \right)\right).$$

- The set of solutions has the form $X = x + S$, where S is a linear subspace of dimension $\dim(S) = n - \text{rank}(A)$.

Contents

- Introduction
- Linear Dependence and Independence
- Rank of a Matrix
- **Matrix Multiplication**
- Invertible Matrices
- Elementary Matrices
- Summary

Sums of Matrices

- Let A and B be $(m \times n)$ -matrices, with coefficients a_{ij} and b_{ij} .
- We define the sum $C = A + B$ componentwise. The coefficients c_{ij} of C are given by

$$c_{ij} = a_{ij} + b_{ij} \quad (1 \leq i \leq m, 1 \leq j \leq n).$$

- This definition could also be interpreted as a generalization of our definition of the sum of vectors, which is here applied to either all row or all column vectors of a matrix.

Product of a Scalar and a Matrix

- The product of a scalar $\alpha \in \mathbb{R}$ and a matrix A is defined to be a matrix whose (i, j) -th coefficient is αa_{ij} .
- Such products of a scalar and a matrix are compatible with the former sum operation in the sense that we have

$$\alpha(A + B) = \alpha A + \alpha B$$

$$(\alpha + \beta)A = \alpha A + \beta A$$

$$(\alpha\beta)A = \alpha(\beta A)$$

for all matrices A, B of compatible dimension, and all $\alpha, \beta \in \mathbb{R}$.

Scalar Product of Vectors

Definition:

- The standard scalar product of two vectors $a, b \in \mathbb{R}^n$ is defined by

$$\langle a, b \rangle \stackrel{\text{def}}{=} a_1b_1 + a_2b_2 + \dots + a_nb_n.$$

Properties. The following relations holds for all $a, b, c \in \mathbb{R}^n$:

1. $\langle a, b \rangle = \langle b, a \rangle$ (symmetry)
2. $\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$ (linearity)
3. $\langle a, a \rangle \geq 0$ (positivity)

Caution: The scalar product has nothing to do with the product of a scalar and a vector/matrix!

Matrix Products

Definition:

- Let A be a $(m \times n)$ -matrix and B an $n \times l$ -matrix. Then the product of A and B is defined to be the $(m \times l)$ -matrix C , whose coefficients satisfy

$$c_{i,j} = \sum_{k=1}^n a_{i,k} b_{k,j}.$$

- We write $C = AB$.

Remark:

- The (i, j) -th entry of C is the scalar product of the i -th row vector of A with the j -th column vector of B .

Matrix Products

Example 1:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 5 & 8 \end{pmatrix}$$

Example 2:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 11 & 7 \end{pmatrix}$$

The matrix-matrix product is not commutative.

That is, in general, we have $AB \neq BA$.

Matrix Products

Theorem: The following properties holds for all matrices A, B, C of compatible dimensions:

1. $(AB)C = A(BC)$ (associativity)
2. $(A + B)C = AC + BC$ and $A(B + C) = AB + AC$ (distributivity)

Proof:

- Associativity follows by comparing the entries on the left and right,

$$\sum_j \left(\sum_k a_{ik} b_{kj} \right) c_{jl} = \sum_{j,k} a_{ik} b_{kj} c_{jl}$$

versus

$$\sum_k a_{ik} \left(\sum_j b_{kj} c_{jl} \right) = \sum_{j,k} a_{ik} b_{kj} c_{jl}.$$

Matrix Products

Proof (continued):

- Distributivity follows from linearity of the scalar product. Namely, if a_i and b_i are the i -th column vectors of A and B and c_j the j -th column vector of C , then

$$\begin{aligned} ((A + B)C)_{ij} &= \langle a_i + b_i, c_j \rangle \\ &= \langle a_i, c_j \rangle + \langle b_i, c_j \rangle \\ &= (AC)_{ij} + (BC)_{ij}. \end{aligned}$$

The proof of the second distributivity law is then analogous.

Linear Equation System as Matrix Product

- The case that A is an $(m \times n)$ -matrix and x a $(n \times 1)$ -matrix is of particular importance. The matrix product of A and x has the form

$$Ax = \begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{pmatrix}$$

- Thus, our general linear equation system can be written in the form

$$Ax = b.$$

- The $(n \times 1)$ -matrix x can also be regarded as a (column) vector.

Matrix-Vector Product

- The above expression for Ax could also be written as

$$Ax = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

In this form, we can see that Ax is always an element of the linear subspace that is spanned by the column vectors of A .

Two Ways to Look at the Matrix Product

- If we write $B = (b_1, \dots, b_s)$ in column format, we have

$$AB = (Ab_1, \dots, Ab_s)$$

Thus, the span of the columns of AB (= column space) is contained in the column space of A . Consequently, we have

$$\text{rank}(AB) \leq \text{rank}(A).$$

- We could also denote the rows of A by $a_1, \dots, a_r$. Then the rows of AB are given by $a_1B, a_2B, \dots, a_rB$. Thus, the span of the rows of AB (=row space) is contained in the row space of B . Thus, we have

$$\text{rank}(AB) \leq \text{rank}(B).$$

Matrix Transpose

Definition:

- The transpose of a matrix A is obtained by exchanging rows and columns. That is,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}^T \stackrel{\text{def}}{=} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ & & \vdots & \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

Properties:

- We have $\text{rank}(A^T) = \text{rank}(A)$, since row rank = column rank.

Matrix Transpose

Remarks

- We have $(AB)^{\top} = B^{\top}A^{\top}$, because

$$((AB)^{\top})_{ij} = (AB)_{ji} = \sum_k a_{jk}b_{ki} = \sum_k b_{ki}a_{jk} = (B^{\top}A^{\top})_{ij}.$$

- A square matrix A is called symmetric if $A = A^{\top}$.
- For example, the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$$

is a symmetric matrix.

Stacked Vectors

- Let $a, c \in \mathbb{R}^n$ and $b, d \in \mathbb{R}^m$ be given by vectors.
- We can stack the vectors a and b as well as c and d ,

$$\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^{n+m} \quad \text{and} \quad \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^{n+m}$$

- The standard scalar product of these stacked vectors satisfies

$$\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right\rangle = \sum_{i=1}^n a_i c_i + \sum_{i=1}^m b_i d_i = \langle a, c \rangle + \langle b, d \rangle.$$

Block Matrices

- We can form block matrices by stacking their row or column vectors.
- If A , B , C , and D are matrices of compatible dimensions, we have

$$\begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = AC + BD.$$

- The dimensions are compatible, if
 1. A and B have the same number of rows,
 2. C and D have the same number of columns,
 3. the number of columns of A equals the number of rows of C , and
 4. the number of columns of B equals the number of rows of D .

Block Matrices

Summary

We can form block matrix products by treating block entries as if they were scalars, provided that:

1. we do not alter the order of matrix multiplications, and
2. we ensure that all dimensions are compatible.

Example

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}.$$

Contents

- Introduction
- Linear Dependence and Independence
- Rank of a Matrix
- Matrix Multiplication
- **Invertible Matrices**
- Elementary Matrices
- Summary

Unit Matrix

- The $(n \times n)$ -unit matrix is denoted by

$$E = E_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

- Its column and row vectors are the unit vectors of $\mathbb{R}^n$.
- Consequently, it satisfies $\text{rank}(E_n) = n$.

Left and Right Inverse

Definition: Let $B \in \mathbb{R}^{m \times n}$ be a given $(m \times n)$ -matrix.

- We say that $A \in \mathbb{R}^{n \times m}$ is a left inverse of B if $AB = E_n$.
- We say that $C \in \mathbb{R}^{n \times m}$ is a right inverse of B if $BC = E_m$.

Observations:

- The matrix E_n is both a left and a right inverse of itself, $E_n E_n = E_n$.
- If A is a left inverse of B and C a right inverse of B , then

$$A = AE_m = A(BC) = (AB)C = E_n C = C.$$

Thus, if A has a left and a right inverse, these matrices coincide.

Left and Right Inverse

More Observations:

- If $AB = E_n$ and $BC = E_m$, we have

$$n = \text{rank}(E_n) = \text{rank}(AB) \leq \text{rank}(B) \leq \min(m, n)$$

$$\text{and } m = \text{rank}(E_m) = \text{rank}(BC) \leq \text{rank}(B) \leq \min(m, n).$$

Hence, if a matrix has a left- and a right inverse, it must be a square matrix, $m = n$.

Left and Right Inverse

Uniqueness:

- Let us assume that A has a left- and right inverse B , such that

$$AB = BA = E \quad \text{with} \quad E = E_n = E_m \quad (\text{since } m = n; \text{ see above}).$$

If C is another matrix that satisfies $AC = CA = E$, then we can repeat the above argument, recalling that

$$C = CE = C(AB) = (CA)B = EB = B.$$

Thus, the left-and-right inverse B of a matrix A , if it exists, is unique.

Invertible Matrix

Definition:

- A matrix A is called invertible if there exists a matrix B such that

$$AB = BA = E.$$

Notation:

- We have already discussed that if there exist a matrix B with $AB = BA = E$, then B is unique. Therefore, we call this unique matrix the inverse of A . It is denoted by

$$B = A^{-1}.$$

- The above equation then reads $A^{-1}A = AA^{-1} = E$.

Characterization of Invertible Matrices

Theorem:

- An $(n \times n)$ -matrix A is invertible if and only if it has full rank; that is, $\text{rank}(A) = n$.

Proof:

- If A has full rank, the equation systems

$$Ab_i = e_i$$

have solutions b_i for all unit vectors e_i . Hence, the matrix $B = (b_1, \dots, b_n)$ is a right inverse. The same argument can be applied to A^T to find a left inverse. As we have already shown that the left and right inverses of A must coincide, A is invertible.

Characterization of Invertible Matrices

Proof (continued):

- Reversely, if A is invertible, there exists a matrix B with $AB = E$.

But then

$$n = \text{rank}(E) = \text{rank}(AB) \leq \text{rank}(A).$$

Thus, A must have full rank, recalling that the maximum possible rank of an $(n \times n)$ -matrix is n .

Computing Left- and Right Inverse Matrices

- For a given matrix $A \in \mathbb{R}^{m \times n}$, we can use Gaussian elimination to solve the equation systems $Ab_i = e_i$. If solutions exist, the matrix

$$B = (b_1, \dots, b_n)$$

is a right inverse (if no solutions exist, no right inverse exists).

- Similarly, we can attempt solving the equations $A^T b_i = e_i$ to search for a left inverse, $B = (b_1, \dots, b_m)^T$.
- In practice, we perform elementary row operations on the matrices $(A \ E_n)$ or $(A^T \ E_m)$ until a row echelon form (or reduced row echelon form) is found.

Computing Invertible Matrices

Example

- Let us search for a right inverse of the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

- We perform elementary row operations on the matrix

$$\begin{aligned} (A \ E) &= \begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{pmatrix}. \end{aligned}$$

Computing Invertible Matrices

Example (continued)

- Thus, the inverse of the matrix A is given by

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

- Let us also double-check that our result is really correct:

$$AA^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E,$$
$$A^{-1}A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E.$$

Linear Equation Systems

- If A is an invertible $(n \times n)$ -matrix, the linear equation system

$$Ax = b$$

has a unique solution. This is because we have

$$n = \text{rank}(A) \leq \text{rank}((A \ b)) \leq n,$$

which implies that the solution set has dimension 0. Specifically, since multiplying the above equation with A^{-1} from the left yields

$$A^{-1}b = A^{-1}(Ax) = (A^{-1}A)x = Ex = x,$$

the solution set is given by $X = \{A^{-1}b\}$.

Contents

- Introduction
- Linear Dependence and Independence
- Rank of a Matrix
- Matrix Multiplication
- Invertible Matrices
- **Elementary Matrices**
- Summary

Elementary Operations

Let us recall our three elementary row operations:

Operation 1: add one row to another

Operation 2: multiply one row elementwise with a scalar $s \neq 0$

Operation 3: permute two rows

Question: can we express these elementary row operations as
matrix products?

Operation 1: Shearing

Definition:

- The $(n \times n)$ -matrix

$$T_1 \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

is called the elementary shearing matrix. Note that T_1 coincides with the unit matrix E except for the additional “1” in the first row, second column.

Operation 1: Shearing

- Multiplication of a matrix A with T_1 from the left yields:

$$\begin{aligned} T_1 A &= \begin{pmatrix} 1 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} + a_{21} & a_{12} + a_{22} & \dots & a_{1m} + a_{2m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}. \end{aligned}$$

- Thus, multiplying with T_1 from left adds the second to the first row.

Operation 1: Shearing

- We could also multiply a matrix A with T_1 from the right:

$$\begin{aligned} T_1 A &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{11} + a_{12} & \dots & a_{1n} \\ a_{21} & a_{21} + a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m1} + a_{m2} & \dots & a_{mn} \end{pmatrix}. \end{aligned}$$

- Multiplying with T_1 from right adds the first to the second column.

Operation 2: Scaling

Definition:

- Let $s \neq 0$ be a given scalar. The associated $(n \times n)$ -matrix

$$D_{1,s} \stackrel{\text{def}}{=} \begin{pmatrix} s & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

is called the elementary scaling matrix.

Properties:

- Obviously, multiplication with $D_{1,s}$ from left (right) can be used to scale the first row (column) of a given matrix A with the scalar $s \neq 0$.

Operation 3: Permutations

Definition:

- We define for any given $1 < k \leq n$ an $(n \times n)$ -matrix P_{1k} , whose coefficients are given by

$$(P_{1k})_{ij} = \begin{cases} 1 & \text{if } 1 \neq i = j \neq k \\ 1 & \text{if } i = 1 \text{ and } j = k \\ 1 & \text{if } i = k \text{ and } j = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The P_{1k} s are called the elementary permutation matrices.

Operation 3: Permutations

Example:

- For $n = 3$ and $k = 2$, we have

$$P_{1,2} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Properties:

- Multiplication with P_{1k} from left permutes first and k -th row.
- Multiplication with P_{1k} from right permutes first and k -th column.
- If $i \neq 1$, $j \neq 1$, and $i \neq j$, we can permute the i -th and j -th row/column by first multiplying with P_{1i} , then with P_{1j} , and then once more with P_{1i} from the left/right; $P_{ij} = P_{1i}P_{1j}P_{1i}$.

Operation 3: Permutations

Examples:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 5 & 4 & 6 \\ 8 & 7 & 9 \end{pmatrix}.$$

Elementary Operations

Summary:

- Generally, shearing, scaling, and permutation matrices are called elementary matrices.
- All elementary row operations can be realized by multiplying with elementary matrices from the left.
- All elementary column operations can be realized by multiplying with elementary matrices from the right.

To be more precise:

- It is enough to define elementary scaling and shearing, T_1 and $D_{1,s}$, only for the first row/column, since the P_{1k} s are defined for all k .

But caution: different authors use different conventions!

Some Remarks

- We will discuss permutations in much more detail later in this course; first in Lecture 3 in the context of constructing determinants, and then in the context of more general groups, see Lecture 4.
- Permutations of the form $P_{ij} = P_{1i}P_{1j}P_{1i}$, which exchange the i -th and j -th row/column of a matrix, are also called transpositions.
- It should be clear that any permutation of rows/columns can be realized by a sequence of such transpositions.

Inverse Elementary Operations

Observations:

- The scaling $D_{1,s}$ can be inverted by scaling with $s^{-1} = \frac{1}{s}$,

$$D_{1,s}^{-1} = D_{1,s^{-1}}.$$

- Elementary permutations can be inverted, too,

$$P_{1k}^{-1} = P_{1k}.$$

- A less obvious observation is that the inverse of an elementary shearing matrices can be written as a product of a scaling matrix, an elementary shearing matrix, and another scaling matrix:

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

Elementary Decomposition

Theorem:

- Every invertible matrix is a product of a finite number of elementary matrices.

Proof:

- Use complete Gaussian elimination to write any invertible matrix A as,

$$M_1 M_2 \dots M_r A N_1 N_2 \dots N_s = E,$$

for some elementary matrices $M_1, M_2, \dots, M_r$ and $N_1, N_2, \dots, N_s$.

- Inverses of elementary matrices are products of elementary matrices.

Thus,

$$A = M_r^{-1} \dots M_2^{-1} M_1^{-1} N_s^{-1} \dots N_2^{-1} N_1^{-1}$$

is a product of elementary matrices, too. This completes our proof.

Contents

- Introduction
- Linear Dependence and Independence
- Rank of a Matrix
- Matrix Multiplication
- Invertible Matrices
- Elementary Matrices
- **Summary**

Summary

This lecture has discussed very important concepts in Linear Algebra:

- Linear Dependence and Independence
- Basis and Dimension of Linear Subspaces
- Complete Gaussian Elimination
- Row Rank = Column Rank = Rank of a Matrix
- Matrix-Matrix Multiplication and its Properties
- Unit Matrices, Invertible Matrices
- Elementary Matrices: Shearing, Scaling, and Permutation

We will revisit these concepts in Lectures 4 and 5, in the context of general vector spaces.

Chapter 3

Determinants

Determinants

- Introduction
- Determinants of (2×2) -Matrices
- Permutations
- General Determinants
- Computation of Determinants
- Volume of Parallelepipeds
- Summary

Contents

- Introduction
- Determinants of (2×2) -Matrices
- Permutations
- General Determinants
- Computation of Determinants
- Volume of Parallelepipeds
- Summary

Objectives

In this lecture we will learn about the definition and basic properties of determinants. Topics include:

- determinants of (2×2) -matrices
- odd and even permutations
- alternating multilinear functions and Leibniz' rule
- product formulas for determinants
- Cramer's rule
- the volume formula for parallelepipeds

Contents

- Introduction
- **Determinants of (2×2) -Matrices**
- Permutations
- General Determinants
- Computation of Determinants
- Volume of Parallelepipeds
- Summary

The (2×2) Case

Recall from Lecture 1:

- The linear equation system

$$ax + by = e$$

$$cx + dy = f$$

has a unique solution if and only if $ad - bc \neq 0$.

- If this conditions is satisfied, the unique solution is given by

$$x = \frac{de - bf}{ad - bc} \quad \text{and} \quad y = \frac{af - ce}{ad - bc}.$$

The (2×2) Case

Matrix Notation from Lecture 2:

- The above linear equation system can be written in the form

$$Az = B \quad \text{with} \quad A \stackrel{\text{def}}{=} \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad z \stackrel{\text{def}}{=} \begin{pmatrix} x \\ y \end{pmatrix}, \quad B \stackrel{\text{def}}{=} \begin{pmatrix} e \\ f \end{pmatrix}.$$

- If $ad - bc \neq 0$, the unique solution is the vector

$$z = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{de-bf}{ad-bc} \\ \frac{af-ce}{ad-bc} \end{pmatrix} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \underbrace{\begin{pmatrix} e \\ f \end{pmatrix}}_{=B}.$$

The (2×2) Case

Comparison of Coefficients:

- Compare the expressions

$$z = A^{-1}B \quad \text{and} \quad z = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} B$$

- If $ad - bc \neq 0$, these expressions are valid for any vector $B \in \mathbb{R}^2$.
- In particular, we could substitute any basis vectors, such as the unit vectors, $B = e_1$ or $B = e_2$. Thus, we find

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

- The above method is called a comparison of coefficients.

The (2×2) Case

Definition: The determinant of the (2×2) -matrix

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

is defined as $\det(A) \stackrel{\text{def}}{=} ad - bc$.

Theorem

- A (2×2) -matrix A is invertible if and only if $\det(A) \neq 0$. Moreover, if it is invertible, we have

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

The (2×2) Case

Properties of the determinant of (2×2) matrices:

- We have $\det(E) = 1 \cdot 1 - 0 \cdot 0 = 1$.
- We have

$$\begin{aligned} \det \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right) &= \det \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix} \\ &= (ae + bg)(cf + dh) - (af + bh)(ce + dg) \\ &= (ad - bc)(eh - fg) = \det \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) \det \left(\begin{pmatrix} e & f \\ g & h \end{pmatrix} \right). \end{aligned}$$

Thus, $\det(AB) = \det(A)\det(B)$ for 2×2 -matrices A and B .

The (2×2) Case

Properties of the determinant of (2×2) matrices:

- The determinant is linear in the first column (with $\alpha \in \mathbb{R}$)

$$\begin{aligned} \det \left(\begin{pmatrix} \alpha a + e & b \\ \alpha c + f & d \end{pmatrix} \right) &= (\alpha a + e)d - b(\alpha c + f) \\ &= \alpha(ad - bc) + (ed - bf) \\ &= \alpha \det \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) + \det \left(\begin{pmatrix} e & b \\ f & d \end{pmatrix} \right). \end{aligned}$$

- A similar computation yields linearity in all rows and columns.

The (2×2) Case

Properties of the determinant of (2×2) matrices:

- We also have

$$\det \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) = ad - bc = \det \left(\begin{pmatrix} a & c \\ b & d \end{pmatrix} \right).$$

Thus, we have $\det(A) = \det(A^T)$ for 2×2 matrices A .

- If A is invertible, the relation $\det(AB) = \det(A)\det(B)$ implies

$$\det(A^{-1}) = \frac{1}{\det(A)} \quad \text{and} \quad \det(ABA^{-1}) = \det(B)$$

for 2×2 matrices A and B .

The (2×2) Case

Properties of the determinant of (2×2) elementary matrices:

- Shearing:

$$\det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1.$$

- Scaling:

$$\det \begin{pmatrix} s & 0 \\ 0 & 1 \end{pmatrix} = s.$$

- Permutation of two rows:

$$\det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1.$$

Contents

- Introduction
- Determinants of (2×2) -Matrices
- **Permutations**
- General Determinants
- Computation of Determinants
- Volume of Parallelepipeds
- Summary

Permutations

- A permutation $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ is a bijjective map from the set of numbers $1, 2, 3, \dots, n$ to the same set.
- We can collect the function values in a table

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}.$$

In the second row, each number from 1 to n shows up exactly ones.

- A composition of two permutations σ and τ is denoted by $\sigma \circ \tau$.
- The set S_n of permutation of this form is often called the permutation group or symmetry group of order n .

Permutations

Example:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 4 & 5 & 1 \end{pmatrix},$$

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 2 & 4 & 1 \end{pmatrix},$$

$$\implies \sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 5 & 3 \end{pmatrix}$$

$$\text{but } \tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 1 & 5 \end{pmatrix}$$

Permutations

Properties:

- All $\sigma, \tau, \omega \in S_n$ satisfy $\sigma \circ (\tau \circ \omega) = (\sigma \circ \tau) \circ \omega$ (associativity)
- The identity map, also called neutral element,

$$e = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 1 & 2 & 3 & \dots & n \end{pmatrix},$$

is a permutation that satisfies $e \circ \sigma = \sigma \circ e = \sigma$ for all $\sigma \in S_n$.

- Since any $\sigma \in S_n$ is a bijection, there exists a $\tau \in S_n$ such that

$$\sigma \circ \tau = \tau \circ \sigma = e.$$

The permutation τ is called the inverse of σ , denoted by $\tau = \sigma^{-1}$.

Transpositions

Definition:

- A transposition is a special type of permutation in S_n that permutes exactly two numbers, but leaves all other numbers constant.
- In other words, $\tau \in S_n$ is a transposition if there are exactly two numbers, $i, j \in \{1, \dots, n\}$ with $i \neq j$, such that

$$\tau(i) = j, \quad \tau(j) = i \quad \text{and} \quad \tau(k) = k \quad \text{for } k \neq i, j.$$

Permutations and Transpositions

Theorem:

- The number of elements of S_n is given by $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$.
- The number of transpositions in S_n is given by $\binom{n}{2} = \frac{n(n-1)}{2}$.
- Every transposition $\tau \in S_n$ satisfies $\tau \circ \tau = e$; that is $\tau^{-1} = \tau$.
- Every permutation $\sigma \in S_n$, with $n \geq 2$, is the composition of a finite number of transpositions,

$$\sigma = \tau_1 \circ \tau_2 \circ \tau_3 \circ \dots \circ \tau_m, \quad m < \infty.$$

Permutations and Transpositions

Proof:

- The first three statements follow directly from the above definitions.
- Regarding the fourth statement, we use an induction over n .
- **Induction Start:** for $n = 2$, the set S_2 only contains two elements: the identity map e and a transposition τ , but $e = \tau \circ \tau$. Thus, the statement is true for $n = 2$.
- **Induction Assumption:** Assume that we can write any $\sigma \in S_n$ as a composition of a finite number of transpositions.

Permutations and Transpositions

Proof (continued):

- **Induction Step:** If $\sigma \in S_{n+1}$ is given, there are two possibilities. The first case is that $\sigma(n+1) = n+1$. In this case, σ merely permutes the numbers $1, \dots, n$, which is already covered by our induction assumption. And the second case is that $\sigma(n+1) \neq n+1$. In this, we can first apply a transposition τ that exchanges $n+1$ and $\sigma(n+1)$, such that we are back to the first case, and write

$$\tau \circ \sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_m \implies \sigma = \tau \circ \tau_1 \circ \tau_2 \circ \dots \circ \tau_m.$$

Thus, σ is a composition of a finite number of transpositions. This completes our proof.

Permutations and Transpositions

- The decomposition of permutations into transpositions is not unique, for example

$$e = \tau \circ \tau = \tau \circ \tau \circ \tau \circ \tau.$$

- Nevertheless, the following statement still holds:

Theorem

- Let $\sigma \in S_n$, $n \geq 2$. Then the number

$$\operatorname{sgn}(\sigma) \stackrel{\text{def}}{=} (-1)^m$$

does not depend on how we decompose σ into transpositions,

$$\sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_m.$$

The number $\operatorname{sgn}(\sigma)$ is called the sign of the permutation σ .

Permutations and Transpositions

Proof:

- Let $\mathbb{Z}$ denote the set of integers. We define $\Delta : \mathbb{Z}^n \rightarrow \mathbb{Z}$ as

$$\Delta(x_1, x_2, \dots, x_n) \stackrel{\text{def}}{=} \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

Note that the product has $\binom{n}{2}$ terms. Any transposition τ satisfies

$$\Delta(x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(n)}) = -\Delta(x_1, x_2, \dots, x_n).$$

Hence, if σ is a composition of m transposition, we have

$$\Delta(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = (-1)^m \Delta(x_1, x_2, \dots, x_n).$$

As $\Delta \neq 0$, the statement of the theorem follows.

Permutations and Transpositions

Remarks:

- The previous theorem implies that there exists a unique map

$$\operatorname{sgn} : S_n \rightarrow \{\pm 1\}$$

such that

1. $\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma) \cdot \operatorname{sgn}(\tau)$ for all $\sigma, \tau \in S_n$, and
 2. $\operatorname{sgn}(\tau) = -1$ for all transpositions τ .
- Concerning the case $n = 1$, we additionally define $\operatorname{sgn}(e) \stackrel{\text{def}}{=} 1$.
 - A permutation σ is called even if $\operatorname{sgn}(\sigma) = 1$.
 - Otherwise, it is called odd.

Permutation Matrices

- Every permutation $\sigma \in S_n$ can be associated with a $(n \times n)$ -matrix P_σ that permutes the unit vectors $e_1, \dots, e_n$, such that

$$P_\sigma e_i = e_{\sigma(i)} \quad (1 \leq i \leq n).$$

- This means that the $\sigma(i)$ -th row of P_σ has to be equal to e_i , or, equivalently, the i -th column of P_σ has to be equal to $e_{\sigma(i)}$,

$$\begin{aligned} P_\sigma &= \left(e_{\sigma^{-1}(1)}, e_{\sigma^{-1}(2)}, \dots, e_{\sigma^{-1}(n)} \right)^\top \\ &= \left(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)} \right). \end{aligned}$$

- Thus, we have $P_\sigma^{-1} = P_{\sigma^{-1}} = P_\sigma^\top$.
- Also note that $P_\sigma P_\tau = P_{\sigma \circ \tau}$ for $\sigma, \tau \in S_n$.

Contents

- Introduction
- Determinants of (2×2) -Matrices
- Permutations
- **General Determinants**
- Computation of Determinants
- Volume of Parallelepipeds
- Summary

Determinants

- Recall that we had defined the determinant of (2×2) -matrices as

$$\det \left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right) = a_{11}a_{22} - a_{12}a_{21}.$$

- Note that the product terms in this expression have the form

$$a_{1\sigma(1)}a_{2\sigma(2)}$$

with $\sigma \in S_2$ being a permutation. The sign in front of this term happens to be the sign of this permutation.

- Now, our rough plan is to generalize this observation in order to introduce a determinant of $(n \times n)$ -matrices.

Determinants

Definition:

- The determinant of an $(n \times n)$ -matrix A is defined as

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- Here, the $a_{i,j}$ s are the coefficients of A .
- And S_n denotes the set of permutations of order n .

Remark:

- Historically, the above equation is known under the name “Leibniz’ formula” (Gottfried Leibniz, 1646–1716).

Determinants

- The Leibniz' formula has a couple of disadvantages.
- First of all, it is not apparent why the above definition of determinants is supposed to be useful.
- Moreover, expressions like " $ad - bc$ " for (2×2) -matrices are certainly elegant, but similar explicit expressions for larger matrices are not.
- For example, for a (3×3) -matrix, we get $3! = 6$ terms:

$$\begin{aligned}\det(A) = & a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}\end{aligned}$$

- For a (10×10) -matrix, we get $10! = 3,628,800$ terms.

Alternating Multilinear Form

- Let $a_1, a_2, \dots, a_n \in \mathbb{R}^n$ denote the columns of a matrix A .
- A function $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is called linear in the i -th column if

$$g(x) \stackrel{\text{def}}{=} f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$$

is a linear function; that is

$$g(x + y) = g(x) + g(y)$$

$$\text{and} \quad g(\alpha x) = \alpha g(x)$$

for all $x, y \in \mathbb{R}^n$ and all $\alpha \in \mathbb{R}$.

Alternating Multilinear Form

Definition: A function $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is called an alternating multilinear form if the following conditions hold.

1. The function f is linear in the i -th column.
2. If we permute two columns of A , $f(A)$ is changing its sign,

$$f(a_1, \dots, a_i, \dots, a_j, \dots, a_n) = -f(a_1, \dots, a_j, \dots, a_i, \dots, a_n)$$

for $1 \leq i < j \leq n$.

Characterization of Determinants

Theorem:

- Every alternating multilinear form $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ satisfies

$$f(A) = \det(A)f(E).$$

Proof:

- Let e_i denote the i -th unit vector. Then the i -th column of A can be written as

$$a_i = a_{i1}e_1 + a_{i2}e_2 + \dots + a_{in}e_n.$$

Characterization of Determinants

Proof (continued):

- By exploiting multilinearity of f , we have

$$\begin{aligned} f(a_1, \dots, a_i, \dots, a_n) &= a_{i1}f(a_1, \dots, e_1, \dots, a_n) \\ &\quad + a_{i2}f(a_1, \dots, e_2, \dots, a_n) \\ &\quad \vdots \\ &\quad + a_{in}f(a_1, \dots, e_n, \dots, a_n) \end{aligned}$$

- As we can do the same for the other columns, we know $f(A)$, if we know how to evaluate f for matrices whose columns are unit vectors.

Characterization of Determinants

Proof (continued):

- Since f is an alternating multilinear form, it vanishes as soon as two columns are equal.
- Hence, it is enough to evaluate f for matrices whose columns are pairwise different unit vectors. But such matrices can be transformed into the unit matrix, E , by permuting columns.
- By collecting all terms, we find that

$$f(A) = \det(A)f(e_1, e_2, \dots, e_n) = \det(A)f(E).$$

- This completes our proof!

Properties of Determinants

Consequences:

- There is only one alternating multilinear form f with $f(E) = 1$:

$$f(A) = \det(A).$$

- Consider the function $h(B) = \det(AB)$. By using that

$$AB = (Ab_1, Ab_2, \dots, Ab_n)$$

we find that h is an alternating multilinear form in the columns of B .

Thus, the following product rule for determinants holds:

$$h(B) = \det(B)h(E) \implies \det(AB) = \det(A)\det(B).$$

Properties of Determinants

Theorem:

- For a square matrix A , we have $\det(A) = \det(A^\top)$.

Proof:

- The function $f(A) = \det(A^\top)$ can be interpreted as an alternating multilinear form of the rows of A , which is unique upon requiring $f(E) = 1$. If $\tau = \sigma^{-1}$ is the inverse permutation of $\sigma \in S_n$, we have

$$a_{\tau(1)1}a_{\tau(2)2} \cdots a_{\tau(n)n} = a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

In other words, the terms occurring in the definition of $\det(A)$ match the terms of $\det(A^\top)$.

Properties of Determinants

- Another interesting property of determinants is the formula

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

where i can be any row index of the $(n \times n)$ -matrix A and $D_{i,j}$ denotes the determinant of the $(n-1) \times (n-1)$ -matrix that is obtained by removing the i -th row and j -th column of A . This formula simply follows by checking that the expression $\sum_{j=1}^n (-1)^{i+j} a_{ij} D_{ij}$ is again a normed alternating multilinear form in A , which must then coincide with $\det(A)$ due to the above theorem about the characterization of determinants.

Properties of Determinants

Example

$$\begin{aligned}\det \left(\begin{pmatrix} 0 & \mathbf{1} & \mathbf{2} \\ 3 & 4 & 5 \\ 6 & 7 & 0 \end{pmatrix} \right) &= -\mathbf{1} \cdot \det \left(\begin{pmatrix} 3 & 5 \\ 6 & 0 \end{pmatrix} \right) \\ &\quad + \mathbf{2} \cdot \det \left(\begin{pmatrix} 3 & 4 \\ 6 & 7 \end{pmatrix} \right) \\ &= -\mathbf{1} \cdot (-30) + \mathbf{2} \cdot (21 - 24) \\ &= 24.\end{aligned}$$

Cramer's Rule

- The above formula can also be written in a slightly different form,

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} D_{ij} = \sum_{j=1}^n a_{ij} b_{ji}$$

$$\text{with } b_{ji} \stackrel{\text{def}}{=} (-1)^{i+j} D_{ij}.$$

- The matrix B with the coefficients $b_{ij} = (-1)^{i+j} D_{ji}$ is called the complementary matrix (or adjoint matrix) of A .
- Additionally, for $i \neq k$, we have $\sum_{j=1}^n a_{ij} b_{jk} = 0$. This implies

$$AB = \det(A)E.$$

Cramer's Rule

Theorem

- The matrix A is invertible if and only if $\det(A) \neq 0$.
- If A is invertible, then we have

$$A^{-1} = \frac{B}{\det(A)},$$

where B is the complementary matrix of A .

Remark:

- The latter equation is known as Cramer's rule.
(Gabriel Cramer, Swiss mathematician, 1704–1752)

Determinants of Block Matrices

Theorem

- Consider a block matrix of the form

$$M = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

with square matrices A and D . Then we have

$$\det(M) = \det(A)\det(D).$$

Proof:

- For the special case $A = E$, the proof of the theorem should be clear: just develop the determinant with respect to the first n columns.

Determinants of Block Matrices

Proof (continued):

- Next, we consider the alternating multilinear form

$$f(A) = \det \left(\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \right)$$

for given matrices B and D . Since we must have

$f(A) = \det(A)f(E)$, the statement of the theorem follows, by reducing it to the above special case:

$$f(A) = \det(A) \det \left(\begin{pmatrix} E & B \\ 0 & D \end{pmatrix} \right) = \det(A) \det(D).$$

Contents

- Introduction
- Determinants of (2×2) -Matrices
- Permutations
- General Determinants
- **Computation of Determinants**
- Volume of Parallelepipeds
- Summary

Computation of Determinants

- Neither Leibniz' formula nor Cramer's rule are really practical for computing determinants of large matrices.
- We can, however, use Gaussian elimination. For this aim, we recall that any invertible matrix A can be written as a product of elementary row operations $M_1, M_2, \dots, M_k$ (see Lecture 2),

$$A = M_1 M_2 \dots M_k.$$

- Thus, by using the product rule for determinants, we find

$$\det(A) = \det(M_1)\det(M_2)\dots\det(M_k).$$

But the determinant of elementary matrices is easy to compute!

Determinants of Elementary Matrices

- Shearing (develop first row):

$$\det \begin{pmatrix} 1 & e_1^\top \\ 0 & E \end{pmatrix} = 1.$$

- Scaling (develop first row):

$$\det \begin{pmatrix} s & 0 \\ 0 & E \end{pmatrix} = s.$$

- Permutation of two rows (since the sign of transpositions is -1):

$$\det(P) = -1.$$

Computation of Determinants

- There are many variants of the above computation method.
- Obviously, we could also allow elementary column operations.
- Instead of a complete Gaussian elimination it is then also enough to work out the row echelon form R of A , such that

$$M_1 M_2 \dots M_k A N_1 N_2 \dots N_l = R$$

for some elementary matrices $M_1 M_2 \dots M_k$ and $N_1 N_2 \dots N_l$.

- We have $\det(R) = 1$ if A has full rank and $\det(R) = 0$ otherwise.
- Finally, the product rule for determinants yields

$$\det(A) = \frac{\det(R)}{\det(M_1)\det(M_2)\dots\det(M_k)\det(N_1)\det(N_2)\dots\det(N_l)}.$$

Contents

- Introduction
- Determinants of (2×2) -Matrices
- Permutations
- General Determinants
- Computation of Determinants
- **Volume of Parallelepipeds**
- Summary

Line Segments

- Any vector $a \in \mathbb{R}^n$ can be associated with a line segment,

$$L(a) \stackrel{\text{def}}{=} [0, a] \stackrel{\text{def}}{=} \{ Ca \mid C \in [0, 1] \} \subseteq \mathbb{R}^n.$$

Caution: $L(a)$ is not a linear subspace of $\mathbb{R}^n$.

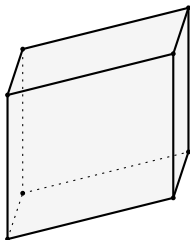
- If $A = (a_1, \dots, a_n)$ has linearly independent column vectors $a_i \in \mathbb{R}^n$,

$$\begin{aligned} \mathcal{P}(A) &\stackrel{\text{def}}{=} L(a_1) + \dots + L(a_n) \\ &\stackrel{\text{def}}{=} \{ x_1 + \dots + x_n \mid x_i \in L(a_i) \text{ for } 1 \leq i \leq n \} \end{aligned}$$

is called a parallelepiped (or, for $n = 2$, a parallelogram).

Parallelepiped

Example:



Visualization of a parallelepiped in $\mathbb{R}^3$.

All six facets are parallelograms.

Unit Cubes

- The special parallelepiped that is obtained for $A = E$,

$$\mathcal{P}(E) = L(e_1) + \dots + L(e_n) = \{ x \in \mathbb{R}^n \mid 0 \leq x_i \leq 1 \},$$

is called the n -dimensional unit cube.

- Note that the 1-dimensional unit cube is a unit interval.
- Similarly, the 2-dimensional unit cube is a unit square.
- In Euclidean spaces with standard scalar product, one usually defines the n -volume of a unit cube to be 1. We write

$$\text{Vol}_n(\mathcal{P}(E)) = 1.$$

Linear Transformations

- Let $X \subseteq \mathbb{R}^n$ be a given set and $B \in \mathbb{R}^{n \times n}$ a given matrix. Then the associated linear transformation of X is defined as

$$BX \stackrel{\text{def}}{=} \{ Bx \mid x \in X \}.$$

- Specifically, for parallelepipeds, we have

$$\begin{aligned} B\mathcal{P}(A) &= BL(a_1) + \dots + BL(a_n) \\ &= L(Ba_1) + \dots + L(Ba_n) = \mathcal{P}(BA), \end{aligned}$$

where the last equation follows from our definition of matrix products.

Scaling

- If D is an invertible diagonal matrix, the associated parallelepiped

$$D\mathcal{P}(E) = \mathcal{P}(D),$$

is a cuboid. Its standard (Euclidean) n -volume is given by

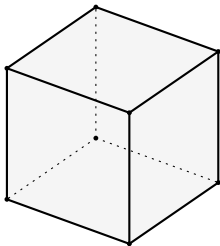
$$\text{Vol}_n(\mathcal{P}(D)) = |d_{11}| \cdot |d_{22}| \cdot \dots \cdot |d_{nn}| = |\det(D)|.$$

- More generally, if A is an invertible matrix and D an invertible diagonal matrix, the standard n -volume satisfies

$$\text{Vol}_n(\mathcal{P}(DA)) = \text{Vol}_n(D\mathcal{P}(A)) = |\det(D)| \cdot \text{Vol}_n(\mathcal{P}(A)).$$

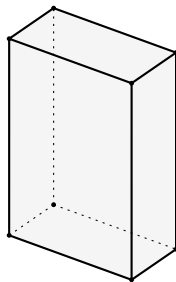
Visualization of a Diagonal Scaling Operation

Unit Cube



Volume = 1

Cuboid



Volume = $|d_{11}| \cdot |d_{22}| \cdot |d_{33}|$

Permutation

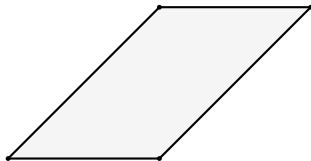
- If $\sigma \in S_n$ is a transposition and P_σ the map $x \rightarrow P_\sigma x$ “mirrors” the vector by exchanging two of its coefficients.
- Thus, $\mathcal{P}(P_\sigma A) = P_\sigma \mathcal{P}(A)$ is a mirror image of $\mathcal{P}(A)$.
- The standard n -volume is invariant under such mirroring operations. Consequently, we have

$$\text{Vol}_n(\mathcal{P}(P_\sigma A)) = \text{Vol}_n(\mathcal{P}(A)) = |\det(P_\sigma)| \cdot \text{Vol}_n(\mathcal{P}(A)).$$

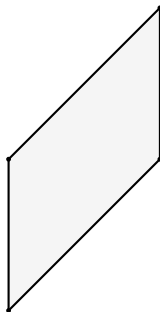
The latter equation holds, because we have $|\det(P_\sigma)| = 1$.

Visualization of a Mirroring Operation

X



$P_\sigma X$



The volume of a set X remains invariant under permutation of coordinates.

Shearing

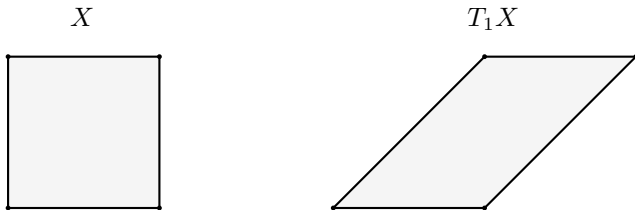
- We recall from Lecture 2 that the elementary shearing matrix is

$$T_1 = \begin{pmatrix} 1 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- The parallelepiped $T_1\mathcal{P}(A)$ is obtained by shearing $\mathcal{P}(A)$. This means that all points $\mathcal{P}(A)$ in the subspace $S = \text{span}(e_2, \dots, e_n)$ are kept fixed while all other points are shifted parallel to S , by a distance proportional to their perpendicular distance from that subspace.
- Note that n -volume is invariant with respect to such shearing,

$$\text{Vol}_n(\mathcal{P}(T_1A)) = \text{Vol}_n(\mathcal{P}(A)) = |\det(T_1)| \cdot \text{Vol}_n(\mathcal{P}(A)).$$

Visualization of a Shearing Operation



The volume of a set remains invariant under shearing.

Volume Formula

Theorem

- Let A be an invertible $(n \times n)$ -matrix. The standard (Euclidean) n -volume of the parallelepiped in $\mathbb{R}^n$ is given by

$$\text{Vol}_n(\mathcal{P}(A)) = |\det(A)|.$$

Proof

- As established in Lecture 2, any invertible matrix A can be written as product of elementary matrices, namely, scaling-, permutation-, and shearing matrices. The statement of the theorem follows then from the elementary geometric considerations on the previous slides and an application of the product rule for determinants.

What is a Volume?

Side Remark (not part of this course)

- In this course, we avoid to introduce a precise definition of what we mean by the standard n -volume Vol_n , because such a discussion would lead far beyond the scope of this course. We only briefly mention that it is—at least in principle—possible to introduce very general classes of sets $X \subseteq \mathbb{R}^n$, for example, Lebesgue measurable sets, whose Lebesgue measure, Vol_n , indeed satisfies

$$\text{Vol}_n(\mathcal{P}(E)) = 1 \quad \text{and} \quad \text{Vol}_n(AX) = |\det(A)| \cdot \text{Vol}_n(X).$$

In the current course, we will **not** discuss such Lebesgue measures. We refer to courses on Analysis instead.

Contents

- Introduction
- Determinants of (2×2) -Matrices
- Permutations
- General Determinants
- Computation of Determinants
- Volume of Parallelepipeds
- **Summary**

Summary

In this lecture we have learned about

- permutations and transpositions,
- the general definition of a determinant using Leibniz' formula,
- alternating multilinear forms and characterization of determinants,
- computation rules for determinants: products, transpose, block matrix
- Cramer's rule for inverse matrices,
- practical computation of determinants using elementary matrices, and
- the volume formula for parallelepipeds.

Chapter 4

Vector Spaces

Vector Spaces

- Introduction
- Groups and Homomorphisms
- Real and Complex Numbers
- Vector Spaces
- Summary

Contents

- Introduction
- Groups and Homomorphisms
- Real and Complex Numbers
- Vector Spaces
- Summary

Objectives

In this lecture we will learn about the definition and basic properties of vector spaces. Topics include:

- basic definition of a group
- the general linear group (revisiting matrix multiplication)
- the imaginary number $\mathbf{i}$ and the group $\{\pm 1, \pm \mathbf{i}\}$.
- real and complex numbers
- general real and complex vector spaces

Contents

- Introduction
- Groups and Homomorphisms
- Real and Complex Numbers
- Vector Spaces
- Summary

Groups

Definition: A group $(G, *)$ is a set G with a binary operation

$$* : G \times G \rightarrow G$$

such that

1. we have $(a * b) * c = a * (b * c)$ for all $a, b, c \in G$ (associativity),
2. there exists a neutral element $e \in G$ such that

$$e * a = a * e = a \quad \text{for all } a \in G.$$

3. there exists for every $a \in G$ an inverse element $b \in G$ such that

$$a * b = b * a = e.$$

Groups

Remarks on Notation:

- The symbol for the binary operation “ $*$ ” could also be replaced by other symbols such as “ $+$ ”, “ $\cdot$ ”, “ $\perp$ ”, etc.. If everything is clear from the context, this symbol is typically left away. For example, we could just write “ ab ” instead of “ $a * b$ ”.
- Since the associativity law holds, we simply write “ abc ” instead of “ $(ab)c$ ” or “ $a(bc)$ ”.

Groups

Inverse Element:

- The inverse element is unique: if we had $ab = ba = e$ and $ac = ca = e$, then

$$c = ec = (ba)c = b(ac) = be = b.$$

- The unique inverse of a is usually denoted by a^{-1} . An exception, however, is made if the symbol for the binary operation is “+”. In this special case, we denote the inverse element by $-a$.

Groups

More Remarks:

- Independent of how we bracket our expressions, we have

$$(a_1 a_2 \dots a_n)^{-1} = a_n^{-1} a_{n-1}^{-1} \dots a_1^{-1}.$$

- In general, the order of the expressions matter, $ab \neq ba$.
- A group G is called Abelian/commutative if $ab = ba$ for all $a, b \in G$.
- If we use “multiplicative notation”, we may define the expressions

$$a^n \stackrel{\text{def}}{=} \underbrace{aa \dots a}_{n \text{ times}} \quad \text{and} \quad a^{-n} \stackrel{\text{def}}{=} (a^{-1})^n.$$

They satisfy $a^m a^n = a^{m+n}$ and $(a^n)^m = a^{mn}$, with $n, m \in \mathbb{Z}$.

Groups

Examples:

- The group $(\{1\}, \cdot)$ has only one element, namely, the number 1. The binary operation is the standard multiplication.
- The group $(\{\pm 1\}, \cdot)$ has two elements: 1 and -1 . The binary operation is the standard multiplication.
- The group $(\mathbb{R}, +)$ consists of the set of real numbers and addition. In this context, the neutral element is denoted by 0.
- The group $(\mathbb{R}^\bullet, \cdot)$ consists of the set of real numbers without 0 and multiplication. In this context, the neutral element is denoted by 1.

The above examples are all examples for Abelian groups!

Groups

More Examples:

- Let M be any set and let $\text{Bij}(M)$ be the set of bijective functions from M to M . Then $(\text{Bij}(M), \circ)$ is a group, where $\circ$ denotes the composition of functions. The identity map is the neutral element, and inverse functions are the inverse elements. If M has more than 2 elements, this group is not commutative.
- If M has a finite number of elements, $|M| = n$, bijections on M corresponds to the set of permutations (see Lecture 3), and we write

$$\text{Bij}(M) \cong S_n$$

to indicate that the groups $(\text{Bij}(M), \circ)$ and $(S_n, \circ)$ are, in this case, isomorphic. We'll discuss the exact meaning of "isomorphic" below.

Groups

More Examples:

- We have already met another very important group in Lecture 2.

Namely, the set $\text{GL}(n, \mathbb{R})$ of invertible $(n \times n)$ -matrices over the real numbers $\mathbb{R}$ is a group with respect to the matrix multiplication. In this case, the neutral element is the unit matrix E , recalling that

$$EA = AE = A$$

for all invertible square matrices A . The inverse element is the inverse matrix A^{-1} , recalling that

$$A^{-1}A = AA^{-1} = E.$$

The group $\text{GL}(n, \mathbb{R})$ is called the general linear group (over $\mathbb{R}$).

Imaginary Numbers

- The equation $x^2 = -1$ has no real-valued solution. This problem can be fixed by introducing an “imaginary unit” $\mathbf{i}$ that satisfies

$$\mathbf{i}^2 = -1.$$

- If we additionally define the product $-\mathbf{i} \stackrel{\text{def}}{=} (-1)\mathbf{i}$, the set $G = \{\pm 1, \pm \mathbf{i}\}$ can be regarded as a commutative (multiplicative) group, where 1 is the neutral element and the relations

$$(-1)(-1) = 1 \quad \text{and} \quad (-\mathbf{i})\mathbf{i} = 1$$

relate the inverse elements.

- The elements of G are all powers of $\mathbf{i}$, as

$$\mathbf{i}^1 = \mathbf{i}, \quad \mathbf{i}^2 = -1, \quad \mathbf{i}^3 = -\mathbf{i}, \quad \text{and} \quad \mathbf{i}^4 = 1.$$

Cancellation

Theorem

- Let $(G, \cdot)$ be a group. If $a, b, c \in G$ satisfy $ac = bc$, then $a = b$.
Similarly, if $ca = cb$, then $a = b$.

Proof (we show right cancellation only, left is analogous)

- Existence of Inverse*: there exists $c^{-1} \in G$ with $cc^{-1} = e$.
- Associativity*: if $ac = bc$, it follows that $(ac)c^{-1} = (bc)c^{-1}$ implies

$$a(cc^{-1}) = b(cc^{-1}) \implies ae = be.$$

- Neutral Element*: since $ae = a$ and $be = b$, we have $a = b$.

All group axioms are needed.

Otherwise, the above proof does not work!

Group Homomorphisms

Definition

- A homomorphism from a group $(G, *)$ to a group $(H, \circ)$ is a map of the form $f : G \rightarrow H$ such that for all $a, b \in G$:

$$f(a * b) = f(a) \circ f(b).$$

Notation Examples:

- If G and H are additive, the above equation becomes

$$f(a + b) = f(a) + f(b).$$

- Similarly, if G is multiplicative and H additive:

$$f(ab) = f(a) + f(b).$$

Group Homomorphisms

Theorem

- Any homomorphism $f : G \rightarrow H$ maps the neutral element of G to the neutral element of H .

Proof:

- Let e be the neutral element in G . It satisfies $ee = e$. Thus,

$$f(e) = f(ee) = f(e)f(e).$$

After multiplying this equation with $f(e)^{-1}$ on both sides, we conclude that $f(e)$ must be the neutral element of H .

Isomorphisms

Definition

- A bijective homomorphism is called an isomorphism.
- Two groups G and H are called isomorphic if there is an isomorphism of the form $f : G \rightarrow H$.

Remarks:

- The composition of group homomorphisms, $G \rightarrow H \rightarrow J$, is again a group homomorphism. The same is true for isomorphisms.
- If $f : G \rightarrow H$ is an isomorphism, then so is its inverse map.
- Isomorphic groups can be considered as “essentially equal” in the sense that G can be “copied” into H by applying f (and vice versa).

Subgroups

Definition

- A non-empty subset $H \subseteq G$ is called a subgroup of G if

$$a, b \in H \implies ab \in H \quad \text{and} \quad a^{-1} \in H.$$

Remarks:

- From $a^{-1} \in H$ it follows $e = aa^{-1} \in H$; and H is then also a group.
- The canonical injection $\iota : H \rightarrow G$, given by $\iota(a) = a$, is a group homomorphism.
- If $f : G \rightarrow H$ is a group homomorphism, then the image set $f(G)$ is a subgroup of H . If f is injective, G and $f(G)$ are isomorphic.

Subgroups

Examples for Subgroups:

- The group $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$.
- The set $\mathbb{R}_{>0}$ of strictly positive real numbers is a subgroup of $(\mathbb{R}^\bullet, \cdot)$.
- The set of even permutations of order n is a subgroup of S_n .
- The set $\{\pm 1\}$ is a subgroup of the group $(\{\pm 1, \pm i\}, \cdot)$.

This is practical, since checking the subgroup conditions is often less work than checking all group axioms.

Subgroups

Examples for Homomorphisms:

- Recall that we can associate with any permutation $\sigma \in S_n$ a permutation matrix

$$P_\sigma \stackrel{\text{def}}{=} \left(e_{\sigma^{-1}(1)}, e_{\sigma^{-1}(2)}, \dots, e_{\sigma^{-1}(n)} \right)^\top,$$

where the e_i s are the unit vectors of $\mathbb{R}^n$. The map

$$\sigma \rightarrow P_\sigma$$

is an injective group homomorphism from S_n to $\text{GL}(n, \mathbb{R})$.

Consequently, the set of permutation matrices is a subgroup of $\text{GL}(n, \mathbb{R})$. It is isomorphic to the group S_n .

Subgroups

More Examples for Homomorphisms:

- The map $\text{sgn} : S_n \rightarrow \{\pm 1\}$ is a group homomorphism (see Lecture 3).
- The map $x \rightarrow e^x$ is an isomorphism between $(\mathbb{R}, +)$ and $(\mathbb{R}_{>0}, \cdot)$.
The associated inverse map is the natural logarithm.
- The determinant $\det : \text{GL}(n, \mathbb{R}) \rightarrow \mathbb{R}^\bullet$ is a group homomorphism from $\text{GL}(n, \mathbb{R})$ to $(\mathbb{R}^\bullet, \cdot)$.

Contents

- Introduction
- Groups and Homomorphisms
- **Real and Complex Numbers**
- Vector Spaces
- Summary

Fields

Definition: A field is a triple $(F, +, \cdot)$ consisting of a set F and two operations, called, addition and multiplication, such that:

1. $(F, +)$ is a commutative group with neutral element $0 \in F$,
2. $(F^\bullet, \cdot)$ is a commutative group with neutral element $1 \in F$,

$$F^\bullet \stackrel{\text{def}}{=} \{a \in F \mid a \neq 0\},$$

3. we have $a(b + c) = ab + bc$ (distributivity).

The most important field is given by $(\mathbb{R}, +, \cdot)$.

Complex Numbers

- In some parts of this course, we need complex numbers.
- We have already introduced the imaginary unit $\mathbf{i} = \sqrt{-1}$.
- Informally, complex numbers are expressions of the form

$$a + b\mathbf{i},$$

where $a, b \in \mathbb{R}$ are real numbers.

- The set of such pairs (a, b) is denoted by $\mathbb{C} \stackrel{\text{def}}{=} \mathbb{R} \times \mathbb{R}$.
- Its associated addition and multiplication rules are

$$\begin{aligned}(a, b) + (c, d) &\stackrel{\text{def}}{=} (a + c, b + d) \\ \text{and} \quad (a, b)(c, d) &\stackrel{\text{def}}{=} (ac - bd, ad + bc)\end{aligned}$$

Complex Numbers

Theorem

- The set of complex numbers, $\mathbb{C} \stackrel{\text{def}}{=} \mathbb{R} \times \mathbb{R}$, together with its associated addition and multiplication rules,

$$(a, b) + (c, d) \stackrel{\text{def}}{=} (a + c, b + d),$$

$$\text{and} \quad (a, b)(c, d) \stackrel{\text{def}}{=} (ac - bd, ad + bc),$$

is a field.

This theorem is important: we can work with complex number as if they were real numbers—as long as we only do basic addition and multiplication operations.

Complex Numbers

Notation

- In the following, we directly use the standard notation. This means we write $a + b\mathbf{i}$ instead of (a, b) .
- Moreover, one usually associates numbers of the form $(a, 0)$ directly with a real number: we write “ a ” instead of “ $a + 0\mathbf{i}$ ” or “ $(a, 0)$ ”.
- For example, we write 0 rather than $(0, 0)$ or $0 + 0\mathbf{i}$.
- Similarly, we write 1 rather than $(1, 0)$ or $1 + 0\mathbf{i}$.

Complex Numbers

Proof of the Theorem:

- Check that $(\mathbb{C}, +)$ is a commutative group with neutral element 0:

Commutativity is clear. Additionally, we use that

$$\begin{aligned}((a + b\mathbf{i}) + (c + d\mathbf{i})) + (e + f\mathbf{i}) &= ((a + c + e) + (b + d + f)\mathbf{i}) \\ &= (a + b\mathbf{i}) + ((c + d\mathbf{i}) + (e + f\mathbf{i}))\end{aligned}$$

$$0 + (a + b\mathbf{i}) = a + b\mathbf{i}$$

$$\text{and} \quad (a + b\mathbf{i}) + (-a - b\mathbf{i}) = 0.$$

Complex Numbers

Proof (continued):

- Similarly, $(\mathbb{C}^\bullet, \cdot)$ is a commutative group with neutral element 1:

Commutativity is again clear. And we quickly verify

$$\begin{aligned}((a + b\mathbf{i})(c + d\mathbf{i}))(e + f\mathbf{i}) &= (ace - bde - adf - bcf) \\ &\quad + (acf + ade + bce - bdf)\mathbf{i} \\ &= (a + b\mathbf{i})((c + d\mathbf{i})(e + f\mathbf{i})) \\ 1(a + b\mathbf{i}) &= a + b\mathbf{i} \\ \text{and } 1 &= (a + b\mathbf{i}) \left(\frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}\mathbf{i} \right).\end{aligned}$$

Complex Numbers

Proof (continued):

- Finally, distributivity needs to be verified:

$$\begin{aligned}(a + b\mathbf{i})((c + d\mathbf{i}) + (e + f\mathbf{i})) &= (a(c + e) - b(d + f)) \\ &\quad + (b(c + e) + a(d + f))\mathbf{i} \\ &= (ac - bd) + (bc + ad)\mathbf{i} \\ &\quad + (ae - bf) + (be + af)\mathbf{i} \\ &= (a + b\mathbf{i})(c + d\mathbf{i}) \\ &\quad + (a + b\mathbf{i})(e + f\mathbf{i}).\end{aligned}$$

- This completes our proof.

Complex Numbers

Remarks:

- Our formula for the inverse element looks a bit complicated:

$$\frac{1}{a + b\mathbf{i}} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}\mathbf{i}.$$

- In order to simplify it, we define the complex conjugate of $z = a + b\mathbf{i}$,

$$\bar{z} \stackrel{\text{def}}{=} a - b\mathbf{i}.$$

- This has the advantage that $z\bar{z} = a^2 + b^2$. Then, we can just write

$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}}.$$

Complex Numbers

More Remarks:

- We additionally define $|z| \stackrel{\text{def}}{=} \sqrt{z\bar{z}} \geq 0$.
- The expression $|z|$ is called the absolute value of z .
- In analogy to real valued z , we have the triangle inequality

$$|z + w| \leq |z| + |w|.$$

- Any complex number $z = a + b\mathbf{i} \in \mathbb{C}$ has a square-root in $\mathbb{C}$. This can be checked by setting $c = \sqrt{a^2 + b^2}$, $a, b \in \mathbb{R}$, because we have

$$\left(\sqrt{\frac{c+a}{2}} + \sqrt{\frac{c-a}{2}} \mathbf{i} \right)^2 = a + b\mathbf{i}.$$

Quadratic Equations

- Since every complex number z has a square-root in $\mathbb{C}$, we can also show that the quadratic equation

$$0 = z^2 + pz + q$$

with $p, q \in \mathbb{C}$ has solutions $z \in \mathbb{C}$. Namely, there exists an $r \in \mathbb{C}$ with $r^2 = p^2 - 4q$. Consequently, $z_1 = -\frac{p+r}{2}$ and $z_2 = -\frac{p-r}{2}$ satisfy

$$z^2 + pz + q = (z - z_1)(z - z_2).$$

- We will generalize the above result in Lecture 7, where we will discuss the fundamental theorem of algebra. It asserts that any polynomial of degree $n \geq 1$ can be expressed as a product of linear factors in $\mathbb{C}$.

Contents

- Introduction
- Groups and Homomorphisms
- Real and Complex Numbers
- **Vector Spaces**
- Summary

Vector Spaces

Recall from Lecture 2:

- We have introduced the set of real n -tuples $\mathbb{R}^n$.
- The elements of this set are vectors with n -coefficients.
- The sum, $x + y$, of two vectors $x, y \in \mathbb{R}^n$ is defined componentwise.
- The product Cx with a $C \in \mathbb{R}$ is defined componentwise, too.

Goal of this lecture:

- We also want to allow the case that $C \in \mathbb{C}$ is a complex number.
- We want to allow that the vectors x and y are more general objects.

But we still want to make sense of sums, $x + y$, and products, Cx .

Vector Spaces

Definition

- Let F be a field (here: $\mathbb{R}$ or $\mathbb{C}$). The triple $(V, +, \cdot)$ with maps

$$+ : V \times V \rightarrow V \quad \text{and} \quad \cdot : F \times V \rightarrow V$$

is called a vector space if $(V, +)$ is a commutative group and

$$1a = a$$

$$C(a + b) = Ca + Cb$$

$$(C + D)a = Ca + Da$$

$$\text{and} \quad C(Da) = (CD)a$$

for all $a, b \in V$ and all $C, D \in F$.

Vector Spaces

Important Examples:

- Most important example for this course: $F = \mathbb{R}$ and $V = \mathbb{R}^n$.
- Second most important example: $F = \mathbb{C}$ and $V = \mathbb{C}^n$.
- A set $W \subseteq V$ is called a linear subspace of V if

$$a, b \in W \quad \text{and} \quad C \in F \quad \implies \quad a + b \in W \quad \text{and} \quad Ca \in W.$$

Obviously, W is then also a vector space over F .

- Thus, the solution set of a linear equation system with coefficients in $\mathbb{R}$ is a vector space over $\mathbb{R}$. (see Lectures 1 and 2)

Vector Spaces

Other Examples:

- Vector spaces come up in zillions of applications.
- These applications are often not directly related to Linear Algebra I.
- Still, it eventually helps to have examples from Analysis I in mind:
 1. The set of continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ is a vector space.
 2. The set of differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ is a vector space.
 3. The set of polynomials $f : \mathbb{R} \rightarrow \mathbb{R}$ is a vector space.

Vector Spaces

Many of the concepts that we learned in Lecture 2 can be generalized to arbitrary vector spaces. In particular, we recall the notions of

- **linear subspaces**
- **linear dependence and independence**
- **basis and dimension**

On the following slides, we go over these core concepts once more, step-by-step, only a little bit more general than last time.

Linear Subspaces

Completely analogous to Lectures 1 and 2:

- the span of given vectors $a_1, \dots, a_m \in V$,

$$\text{span}(a_1, \dots, a_m) \stackrel{\text{def}}{=} \{ C_1 a_1 + \dots + C_m a_m \mid C_1, \dots, C_m \in F \},$$

is a linear subspace of V and, consequently, a vector space.

Only new aspect:

- A vector space V is called finitely generated if there exist $a_1, \dots, a_m \in V$ such that $V = \text{span}(a_1, \dots, a_m)$.

Linear Dependence and Independence

Completely analogous to Lectures 2:

- the vectors $a_1, \dots, a_m \in V$ are called linearly dependent if there exist numbers $C_1, \dots, C_m \in F$, not all equal to 0, such that

$$C_1 a_1 + \dots + C_m a_m = 0.$$

Otherwise, they are called linearly independent.

- Finitely generated vector space can have at most finitely many linearly independent elements.

Basis and Dimension

Completely analogous to Lectures 2:

- The basis of a vector space is a maximal system of linearly independent vectors.
- In a finitely generated vector space, all bases consist of an equal number of linearly independent vectors.

This number is called the dimension of the vector space.

Only new aspect:

- If a vector space is not finitely generated, it is called an infinite dimensional vector space.

Example

Polynomials

- Let V be the vector space of polynomials with degree ≤ 2 , over $\mathbb{R}$.
- This means that V is the set of functions of the form

$$p(x) = ax^2 + bx + c$$

with real-valued coefficients $a, b, c \in \mathbb{R}$.

- The vector space V is finitely generated by the monomials

$$e_1(x) = 1, \quad e_2(x) = x, \quad \text{and} \quad e_3(x) = x^2.$$

- Namely, we can write $p = ce_1 + be_2 + ae_3$.

Example

Polynomials

- Because the monomials

$$e_1(x) = 1, \quad e_2(x) = x, \quad \text{and} \quad e_3(x) = x^2.$$

are linearly independent and because every polynomial of degree ≤ 2 can be written in the form $p = ce_1 + be_2 + ae_3$, the elements $e_1, e_2, e_3 \in V$ are a basis.

- Consequently, the dimension of V is equal to $\dim(V) = 3$.
- More generally, we can use an analogous argument to prove that the vector space of polynomials of degree $\leq n$ has dimension $n + 1$.

Contents

- Introduction
- Groups and Homomorphisms
- Real and Complex Numbers
- Vector Spaces
- **Summary**

Summary

- In this lecture we have learned about some of the most basic concepts in Algebra, such as groups and group homomorphisms.
- We've introduced general fields, but we've focused on real and complex numbers.
- Finally, the main section of this lecture has been on vector spaces.
- The analysis of general vector spaces is very similar to studying $\mathbb{R}^n$.
- But: general vector spaces can be infinite dimensional.
- The focus of this course is on finitely generated vector spaces.
- Important examples of vector spaces are $\mathbb{R}^n$, $\mathbb{C}^n$, and their linear subspaces, but also other spaces such as the set of polynomials.

Chapter 5

Linear Transformations

Linear Transformations

- Introduction
- Linear Transformations
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- Summary

Contents

- Introduction
- Linear Transformations
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- Summary

Objectives

In this lecture we will learn about the definition and basic properties of linear maps between vector spaces. Topics include:

- linear maps between vector spaces and their relation to matrices
- kernels, images, and isomorphisms
- basis of a vector space, and change of basis
- characterization of the set of linear maps between vector spaces

Contents

- Introduction
- **Linear Transformations**
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- Summary

Linear Transformations

- Let A be an $m \times n$ matrix with coefficients in $F = \mathbb{R}$ or $F = \mathbb{C}$.
- We consider the map

$$x \rightarrow Ax$$

that maps the $(n \times 1)$ matrix x to the $(m \times 1)$ matrix Ax .

- If we associate the set of $(n \times 1)$ -matrices with F^n , we can regard

$$f_A : F^n \rightarrow F^m \quad \text{with} \quad f_A(x) \stackrel{\text{def}}{=} Ax$$

as a function from the vector space $V = F^n$ to $W = F^m$.

Linear Transformations

- Let us recall once more that the equation $b = f_A(x) = Ax$ means that

$$b_i = \sum_j a_{ij}x_j.$$

- In the context of linear equation systems, asking whether $Ax = b$ has a solution is the same as asking whether b is in the image set of the function f_A ,

$$\text{Im}(f_A) \stackrel{\text{def}}{=} f_A(V) = \{f_A(x) \mid x \in V\} = \{Ax \mid x \in V\} \subseteq W.$$

- Functions like f_A are called linear transformations.

Linear Transformations

Definition: A function of the form $f : V \rightarrow W$, from a vector space V to a vector space W (over the same field F), is called a linear transformation (or linear function or linear map) if

1. $f(x + y) = f(x) + f(y)$ for all $x, y \in V$, and
2. $f(Cx) = Cf(x)$ for all $x \in V$ and all $C \in F$.

Remark

- For $F = \mathbb{R}$, $V = \mathbb{R}^n$, and $W = \mathbb{R}^m$, the function $f_A(x) = Ax$, with an $(m \times n)$ -matrix A , is such a linear transformation.

Linear Transformations

Observations and Remarks:

- If $f : V \rightarrow W$ is a linear transformation and $a_1, a_2, \dots, a_r \in V$, then

$$f(C_1a_1 + C_2a_2 + \dots + C_ra_r) = C_1f(a_1) + C_2f(a_2) + \dots + C_rf(a_r).$$

This follows directly from the above definition.

- Let $0_F \in F$, $0_V \in V$, and $0_W \in W$ be the zeros of F , V , and W .

Every linear transformation $f : V \rightarrow W$ satisfies

$$f(0_V) = f(0_F 0_V) = 0_F f(0_V) = 0_W.$$

- In the following, however, we will simply write $f(0) = 0$.

Linear Transformations

Example 1

- Let W be a linear subspace of the vector space V .
- The natural inclusion

$$\iota : W \rightarrow V, \quad \iota(x) = x$$

is a simple but important example for a linear transformation.

Linear Transformations

Example 2

- Let $a_1, a_2, \dots, a_n$ be given elements of a vector space V .
- The map

$$F^n \rightarrow V, \quad x \rightarrow x_1 a_1 + x_2 a_2 + \dots + x_n a_n$$

is a linear transformation from F^n to V .

- Note that the vectors $a_1, a_2, \dots, a_n$ are a generating system of V ,
 $V = \text{span}(a_1, a_2, \dots, a_n)$, if and only if the above map is surjective.

Linear Transformation of Basis Vectors

Theorem

- Let $e_1, e_2, \dots, e_n$ be a basis of a vector space V .
- Let W be another vector space. Assume $b_1, b_2, \dots, b_n \in W$ are given.
- Then there exists a unique linear transformation $f : V \rightarrow W$ with

$$f(e_i) = b_i \quad (1 \leq i \leq n).$$

Proof

- Pick any $x = C_1e_1 + \dots + C_ne_n \in V$. Obviously, we have to set

$$f(x) = f(C_1e_1 + \dots + C_ne_n) = C_1b_1 + \dots + C_nb_n.$$

There is no other option. And this map is also linear.

Linear Transformation of Basis Vectors

Remark

- For the case that $V = \mathbb{R}^n$ (or $V = F^n$ if we want to keep it general), we can use the unit vectors $e_1, e_2, \dots, e_n$ as basis.
- Then there exists for every n -tuple of vectors $a_1, a_2, \dots, a_n \in W$ exactly one linear map f with $f(e_i) = a_i$.
- This linear map is given by

$$f(x) = x_1 a_1 + x_2 a_2 + \dots + a_n x_n.$$

This is exactly our “Example 2” (see two slides above).

Linear Transformation of Basis Vectors

Remark

- Let us come back to the linear map $f_A(x) = Ax$.
- If we apply the map f_A to the unit vectors, e_i , it returns the column vectors, a_i , of the matrix A .

Example

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies f_A(e_1) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}$$
$$\text{and } f_A(e_2) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}$$

If we know all the columns of A , we also know what A is.

Contents

- Introduction
- Linear Transformations
- **Kernels and Images**
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- Summary

Image

Definition

- The image of a linear transformation $f : V \rightarrow W$ is denoted by

$$\text{Im}(f) = f(V) = \{ f(x) \mid x \in V \}.$$

Observation

- The set $\text{Im}(f)$ is a linear subspace of W , because we have

$$b_1 = f(a_1) \quad \text{and} \quad b_2 \in f(a_2) \quad \implies \quad \begin{cases} b_1 + b_2 = f(a_1 + a_2) \\ Cb_1 = f(Ca_1). \end{cases}$$

Thus, if b_1 and b_2 are in $\text{Im}(f)$, then so are $b_1 + b_2$ and Cb_1 .

Image and Rank

Image of a Matrix

- Let $a_1, a_2, \dots, a_n \in F^m$ be the columns of a $(m \times n)$ -matrix A .
- The image of $f_A(x) = Ax$, given by,

$$\text{Im}(f_A) \stackrel{\text{def}}{=} \{ Ax \mid x \in F^n \} = \text{span}(a_1, \dots, a_n) \subseteq F^m$$

is the column space of A (sometimes also called range space of A).

- Since $W = F^m$ is finite dimensional and $\text{Im}(f_A)$ is a linear subspace of W , it has a finite dimension. We recall that this dimension is called the rank of A (see Lecture 2),

$$\text{rank}(A) = \dim(\text{Im}(f_A)).$$

Kernel

Definition

- The kernel of a linear transformation $f : V \rightarrow W$ is the set of elements of V that are mapped to 0,

$$\ker(f) \stackrel{\text{def}}{=} \{ x \in V \mid f(x) = 0 \}.$$

Observation

- Note that $\ker(f)$ is a linear subspace of V , because we have

$$f(a_1) = 0 \quad \text{and} \quad f(a_2) = 0 \quad \implies \quad \begin{cases} f(a_1 + a_2) = 0 \\ f(Ca_1) = 0. \end{cases}$$

Null Space and Nullity

Remarks

- The kernel of $f_A(x) = Ax$ for a given $(m \times n)$ -matrix A ,

$$\ker(f_A) \stackrel{\text{def}}{=} \{ x \in F^n \mid Ax = 0 \},$$

is the solution set of the homogeneous linear equation $Ax = 0$. The kernel of f_A is often called the null space of A ,

$$\text{null}(A) \stackrel{\text{def}}{=} \ker(f_A).$$

- Since $V = F^n$ is finite dimensional and $\ker(f_A)$ is a linear subspace of V , it also has a finite dimension. It is called the nullity of A ,

$$\text{nullity}(A) \stackrel{\text{def}}{=} \dim(\ker(f_A)).$$

Injectivity

Theorem: Let $f : V \rightarrow W$ be a linear transformation. The following statements are equivalent.

1. The function f is injective.
2. The kernel of f is zero; that is $\ker(f) = \{0\}$.

Proof:

- If f is injective, 0 can only be mapped to 0. Thus, $\ker(f) = \{0\}$.
- Reversely, if $\ker(f) = \{0\}$, then f is injective, because

$$f(a) = f(b) \implies 0 = f(a) - f(b) = f(a-b) \implies a-b = 0 \implies a = b.$$

Matrices and Their Linear Transformations

Interpretation for Matrices:

- Let us revisit the map $f_A : x \rightarrow Ax$.
- Let $a_1, a_2, \dots, a_n \in F^m$ be the column vectors of $A = (a_1, \dots, a_n)$.
- We recall that the map $f_A : F^n \rightarrow F^m$ can be written as

$$f_A(x) = Ax = x_1a_1 + x_2a_2 + \dots + x_na_n$$

- The map f_A is surjective if and only if $V = \text{span}(a_1, \dots, a_n)$.
- The map f_A is injective if and only if the a_i s are linearly independent.
- The map f_A is bijective if and only if A is invertible.

Dimension Formula

Theorem:

- Let $f : V \rightarrow W$ be a linear transformation between finite dimensional vector spaces. Then

$$\dim(\ker(f)) + \dim(\operatorname{Im}(f)) = \dim(V).$$

Proof:

- Let $e_1, \dots, e_m$ be a basis of $\ker(f)$, with $m = \dim(\ker(f))$.
- And let $e_1, \dots, e_m, e_{m+1}, \dots, e_n$ be a basis of V , with $n = \dim(V)$.
- Now, $f(e_1), \dots, f(e_n)$ generate $\operatorname{Im}(f)$, but $f(e_1) = \dots = f(e_m) = 0$.
- Next, we show that $f(e_{m+1}), \dots, f(e_n)$ are linearly independent.

Dimension Formula

Proof (continued):

- Assume that $C_{m+1}f(e_{m+1}) + \dots + C_n f(e_n) = 0$.
- Then $C_{m+1}e_{m+1} + \dots + C_n e_n$ is in $\ker(f)$, and, consequently, must be a linear combination of $e_1, \dots, e_m$.
- But the e_i s are linearly independent $\implies C_{m+1} = \dots = C_n = 0$.
- Hence, the vectors $f(e_{m+1}), \dots, f(e_n)$ are linearly independent.
- It follows that

$$\dim(V) = n, \quad \dim(\ker(f)) = m, \quad \text{and} \quad \dim(\operatorname{Im}(f)) = n - m.$$

This completes the proof.

Dimension Formula for Matrices

Interpretation for Matrices:

- Recall that our map $f_A : F^n \rightarrow F^m$ with $f_A(x) = Ax$ satisfies

$$\text{rank}(A) = \dim(\text{Im}(f_A)),$$

$$\text{nullity}(A) = \dim(\ker(f_A)),$$

$$\text{and } n = \dim(F^n).$$

- Thus, the dimension formula for matrices with n columns is

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

Contents

- Introduction
- Linear Transformations
- Kernels and Images
- **Isomorphisms**
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- Summary

Composition of Linear Transformations

- If we have two linear transformations, $f : V \rightarrow W$ and $g : W \rightarrow U$, we can introduce their composition $h : V \rightarrow U$,

$$h = g \circ f$$

- The composition h is also a linear transformation, because

$$h(x + y) = g(f(x + y)) = g(f(x) + f(y))$$

$$= g(f(x)) + g(f(y)) = h(x) + h(y)$$

$$h(Cx) = g(f(Cx)) = g(Cf(x)) = Cg(f(x)) = Ch(x)$$

for all $x, y \in V$ and all $C \in F$.

Isomorphisms

Definition:

- The linear map $f : V \rightarrow W$ is called an isomorphism if it is bijective.

Observation:

- We can use isomorphisms for “transporting” properties from V to W :

$$V = \text{span}(a_1, \dots, a_n) \implies W = \text{span}(f(a_1), \dots, f(a_n))$$

$$a_1, \dots, a_n \text{ lin. independent} \implies f(a_1), \dots, f(a_n) \text{ lin. independent}$$

$$a_1, \dots, a_n \text{ is a basis of } V \implies f(a_1), \dots, f(a_n) \text{ is a basis of } W$$

- Thus, if an isomorphism $f : V \rightarrow W$ exists, then $\dim(V) = \dim(W)$.

Isomorphic Vector Spaces

Definition:

- The vector spaces V and W are called isomorphic, $V \cong W$, if there exists an isomorphism from V to W .

Remark:

- If $f : V \rightarrow W$ and $g : W \rightarrow U$ are isomorphisms, then so is $g \circ f$.
- Consequently, $V \cong W$ and $W \cong U$ imply $V \cong U$.
- It follows that “ $\cong$ ” is an equivalence relation; that is, we have

$$V \cong V$$

Reflexivity,

$$V \cong W \implies W \cong V$$

Symmetry,

$$\text{and } (V \cong W \text{ and } W \cong U) \implies V \cong U$$

Transitivity.

Consequence of Dimension Formula

Theorem: Let $f : V \rightarrow W$ be a linear transformation between finite dimensional vector spaces with $n = \dim(V) = \dim(W)$. Then the following statements are equivalent.

1. The map f is injective.
2. The map f is surjective.
3. The map f is an isomorphism.

Proof

- The statement is a consequence of the general dimension formula

$$\dim(\ker(f)) + \dim(\operatorname{Im}(f)) = \dim(V).$$

Isomorphisms and Basis

Observation:

- Let $a_1, a_2, \dots, a_n \in V$ be elements of a vector space. Then, the map

$$F^n \rightarrow V, \quad x \rightarrow x_1 a_1 + x_2 a_2 + \dots x_n a_n$$

is an isomorphism if and only if $a_1, a_2, \dots, a_n$ is a basis of V .

- Equivalently, we could also say that if $a_1, a_2, \dots, a_n$ is a basis of V , then $F^n \cong V$.

Very important consequence:

Every finite dimensional vector space is isomorphic to F^n .

Comments

Short Pause and Discussion:

- The focus of this lecture is on the case $F = \mathbb{R}$ (later also $F = \mathbb{C}$).
- But a finite dimensional vector space over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$.

Thus, one could ask:

Why do we need general vector spaces?

Wouldn't it be enough to study $\mathbb{R}^n$ as in Lecture 2?

Comments

Short Pause and Discussion:

- There are two answers to this question.
- **Answer 1:** Vector space arise naturally in zillions of application. But it is not always easy to see that they are isomorphic to $\mathbb{R}^n$. For example, if we think of the set of solutions of a linear equation systems, or the set of polynomials up to a given degree, we had to work hard to understand the connection.

Comments

Short Pause and Discussion:

- **Answer 2:** Many practical vector spaces are infinite dimensional.

This is not so much the focus of this lecture, but it will be needed in your future lectures, for example, on analysis and functional analysis, algebra, partial differential equations, and many others. For instance, the set of differentiable functions is an infinite dimensional vector space that has many practical applications.

It is important to get to used to vector spaces early on!

Contents

- Introduction
- Linear Transformations
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- Summary

Sums and Intersections

- Let $V_1, V_2, \dots, V_m$ be linear subspaces of a vector space V .
- The sum of $V_1, V_2, \dots, V_m$, defined by

$$V_1 + V_2 + \dots + V_m \stackrel{\text{def}}{=} \{v_1 + v_2 + \dots + v_m \mid v_1 \in V_1, \dots, v_m \in V_m\},$$

is a linear subspace of V and, consequently, a vector space.

- Similarly, the intersection of all vector spaces,

$$V_1 \cap V_2 \cap \dots \cap V_m$$

is closed under addition and scalar multiplication and, consequently, again a vector space.

Products

- The products of two vector spaces V and W , denoted by $V \times W$ can be equipped with the operations

$$(a, b) + (c, d) \stackrel{\text{def}}{=} (a + c, b + d)$$

$$C(a, b) \stackrel{\text{def}}{=} (Ca, Cb)$$

with $(a, b), (c, d) \in V \times W$.

- Together with this definition of the addition and multiplication with a scalar $C \in F$, the product $V \times W$ is a vector space.
- It should be clear that general products, $V_1 \times V_2 \times \dots \times V_m$, can be constructed analogously.

Dimension Formula for Products

Theorem

- If V and W are finite dimensional vector spaces, then

$$\dim(V \times W) = \dim(V) + \dim(W).$$

Proof

- Let $v_1, v_2, \dots, v_n$ be a basis of V and let $w_1, w_2, \dots, w_m$ be a basis of W . Then the $m + n$ pairs $(v_i, 0)$ and $(0, w_j)$, with $1 \leq i \leq n$ and $1 \leq j \leq m$, are a basis of $V \times W$. This yields the dimension formula for products.

Dimension Formula for Sums

Theorem

- If V and W are subspaces of a finite dimensional vector space U , then

$$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W).$$

Proof

- We introduce the surjective linear transformation

$$V \times W \rightarrow V + W, \quad (v, w) \rightarrow v + w.$$

The kernel of this map consists of all pairs $(a, -a)$ with $a \in V \cap W$.

It is isomorphic to $V \cap W$. Thus, the dimension formula yields

$$\dim(V \times W) = \dim(V + W) + \dim(V \cap W).$$

By using the dimension formula for products, our theorem follows.

Direct Sum

Remarks

- The general sum formula

$$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W)$$

can sometimes be simplified.

- For example, if we have $V \cap W = \{0\}$, we have

$$\dim(V + W) = \dim(V) + \dim(W),$$

which is, perhaps, a bit easier to remember.

- If the condition $V \cap W = \{0\}$ holds, the sum “ $V + W$ ” is called the direct sum of V and W . In this case, we have $V \times W \cong V + W$.

Contents

- Introduction
- Linear Transformations
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- **Change of Basis**
- Endomorphisms
- Summary

Linear Transformations and Matrices

- Recall our definition of the map $f_A : F^n \rightarrow F^m$ as $f_A(x) = Ax$.
- Here, we have used the matrix A to define a linear transformation.
- But we could also reverse this construction. Namely, if

$e_1, e_2, \dots, e_n$ denotes a basis of $V \cong F^n$,

$e'_1, e'_2, \dots, e'_m$ denotes a basis of $W \cong F^m$,

and $f : V \rightarrow W$ denotes a linear transformation, then there exists a unique $(m \times n)$ -matrix A with coefficients $a_{i,j}$ such that

$$f(e_i) = a_{1i}e'_1 + a_{2i}e'_2 + \dots + a_{mi}e'_m.$$

Linear Transformations and Matrices

- Similarly, we denote with

$e_1, e_2, \dots, e_n$ a basis of $V \cong F^n$,

$e'_1, e'_2, \dots, e'_m$ a basis of $W \cong F^m$, and with

$e''_1, e''_2, \dots, e''_l$ a basis of $U \cong F^l$.

- Consider linear maps $f : V \rightarrow W$, $g : W \rightarrow U$, and $h = g \circ f$.
- As above, there exist matrices A , B , and C , such that

$$f(e_i) = a_{1i}e'_1 + a_{2i}e'_2 + \dots + a_{mi}e'_m = \sum_{j=1}^m a_{ji}e'_j,$$

$$g(e'_j) = b_{1j}e''_1 + b_{2j}e''_2 + \dots + b_{lj}e''_l = \sum_{k=1}^l b_{kj}e''_k,$$

$$\text{and } h(e_i) = c_{1i}e''_1 + c_{2i}e''_2 + \dots + c_{li}e''_l = \sum_{k=1}^l c_{ki}e''_k.$$

Linear Transformations and Matrices

- Thus, we have

$$\begin{aligned}\sum_{k=1}^l c_{ki} e''_k &= h(e_i) = g(f(e_i)) = g\left(\sum_{j=1}^m a_{ji} e'_j\right) \\ &= \sum_{j=1}^m a_{ji} g(e'_j) = \sum_{j=1}^m a_{ji} \left(\sum_{k=1}^l b_{kj} e''_k\right) \\ &= \sum_{k=1}^l \left(\sum_{j=1}^m b_{kj} a_{ji}\right) e''_k\end{aligned}$$

- Since the e''_k s are linearly independent, the last equation implies

$$c_{ki} = \sum_{j=1}^m b_{kj} a_{ji} \quad \implies \quad C = BA.$$

What is a Matrix Product?

- In Lecture 2, we had introduced the matrix product $C = BA$ by defining it coefficient-wise as

$$c_{ki} \stackrel{\text{def}}{=} \sum_{j=1}^m b_{kj} a_{ji}.$$

But this definition emerged without motivation for its specific form.

- In the above analysis of $h = g \circ f$, however, the equation

$$c_{ki} = \sum_{j=1}^m b_{kj} a_{ji}$$

comes up naturally in the context of analyzing linear transformation.

Thus, better late than never:

This is the motivation for our definition of matrix products!

Change of Basis

- Let us assume that $e_1, e_2, \dots, e_n$ is a basis of the vector space V .
- We have just learned that any linear transformation $f : V \rightarrow V$ can be represented by a matrix A , such that

$$f(e_i) = \sum_{j=1}^n a_{ji} e_j.$$

- What happens if we choose another basis of V , say $e'_1, e'_2, \dots, e'_n$?
- The map f will of course be the same, but its representation,

$$f(e'_i) = \sum_{j=1}^n b_{ji} e'_j,$$

depends on the basis. In general, we have $A \neq B$.

Question: How can we convert A to B when changing our basis?

Transition Matrix

Definition

- Let $e_1, e_2, \dots, e_n$ and $e'_1, e'_2, \dots, e'_n$ be bases of a vector space V .

The transition matrix S from the first to the second basis is defined such that

$$e'_i = \sum_{j=1}^n s_{ji} e_j \quad (1 \leq i \leq n).$$

Remark:

- The transition matrix S is invertible: the inverse S^{-1} is the transition matrix from the second to the first basis.

Change of Basis

- Using the notation from previous slides, we collect three equations

$$f(e_i) = \sum_{j=1}^n a_{ji} e_j,$$

$$f(e'_i) = \sum_{j=1}^n b_{ji} e'_j,$$

$$\text{and} \quad e'_i = \sum_{j=1}^n s_{ji} e_j.$$

- Let us substitute the third into the second equation

$$\sum_{j=1}^n s_{ji} f(e_j) = \sum_{j=1}^n b_{ji} \left(\sum_{k=1}^n s_{kj} e_k \right).$$

Change of Basis

- Next, we substitute the first equation,

$$\sum_{j=1}^n s_{ji} \left(\sum_{k=1}^n a_{kj} e_k \right) = \sum_{j=1}^n b_{ji} \left(\sum_{k=1}^n s_{kj} e_k \right).$$

- Since the e_k s are linear independent, we must have

$$\sum_{j=1}^n a_{kj} s_{ji} = \sum_{j=1}^n s_{kj} b_{ji}.$$

- We can also write this equation by using matrix products,

$$AS = SB \quad \implies \quad B = S^{-1}AS.$$

Change of Basis

Summary of our derivation from the last two slides:

Theorem

- Let $f : V \rightarrow V$ be a linear transformation, let $e_1, e_2, \dots, e_n$ and $e'_1, e'_2, \dots, e'_n$ be bases of V , and let S be the transition matrix from the first to the second basis. If the matrix A represents f with respect to the first basis (from left and right) and if the matrix B represents f with respect to the second basis (from left and right), then we have

$$B = S^{-1}AS.$$

Change of Basis

Remark

- The change of basis formula for $f : V \rightarrow V$,

$$B = S^{-1}AS,$$

is sufficient for most practical applications.

- Nevertheless, in some cases one needs more general linear transformation $f : V \rightarrow W$, which map from a vector space V to a vector space W . In this case, we can change the bases of both V and W . The derivation of the corresponding change of basis formula is analogous to the above formula. We summarize it on the next slide.

Change of Basis

Theorem

- Let $f : V \rightarrow W$ be a linear transformation between different vector spaces V and W . Let S denote the transition matrix from the first to the second basis of V and let T denotes the transition matrix from the first to the second basis of W . The representation matrix A of f with respect to the first bases in V and W and the representation matrix B of f with respect to the second bases in V and W satisfy

$$B = S^{-1}AT.$$

Change of Basis

Example

- Let V denote the set of polynomials up to order 2.
- As discussed in Lecture 4, a basis for V are the monomials,

$$e_1(x) = 1, \quad e_2(x) = x, \quad e_3(x) = x^2$$

- Now, let us consider the linear transformation

$$f(p) = \frac{dp}{dx}$$

that maps a polynomial $p \in V$ to its first order derivative.

- Note that an application of f to the basis e_1, e_2, e_3 yields

$$f(e_1) = 0, \quad f(e_2) = e_1, \quad \text{and} \quad f(e_3) = 2e_2.$$

Change of Basis

Example (continued)

- Consequently, the linear transformation f , with respect to the monomial basis e_1, e_2, e_3 , is represented by the matrix

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

- The derivative of $p(x) = ax^2 + bx + c$, given by $\frac{d}{dx}p(x) = 2ax + b$, can be computed by evaluating the matrix vector product

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} b \\ 2a \\ 0 \end{pmatrix}.$$

Change of Basis

Example (continued)

- Next, let us consider the first three Legendre polynomials,

$$e'_1(x) = 1, \quad e'_2(x) = x, \quad e'_3(x) = \frac{1}{2}(3x^2 - 1).$$

It is not difficult to see that (e'_1, e'_2, e'_3) is also a basis of V .

- Let us collect the relations

$$e'_1 = e_1, \quad e'_2 = e_2, \quad \text{and} \quad e'_3 = \frac{3}{2}e_3 - \frac{1}{2}e_1.$$

- Thus, our transition matrix is given by

$$S = \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{pmatrix}.$$

Change of Basis

Example (continued)

- Next, we apply our change of basis formula to find the representation matrix B of $f = \frac{d}{dx}$ with respect to our new basis:

$$B = S^{-1}AS = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{pmatrix}$$
$$\Rightarrow B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Change of Basis

Example (continued)

- Summary: we can differentiate a polynomial in Legendre basis,

$$p = a \frac{3x^2 - 1}{2} + bx + c \quad \implies \quad \frac{dp}{dx} = 3ax + b,$$

by an application of the matrix B to its coefficients,

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} b \\ 3a \\ 0 \end{pmatrix}.$$

Contents

- Introduction
- Linear Transformations
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- **Endomorphisms**
- Summary

The Set of Linear Transformations

Definition

- Let V and W be vector spaces. Linear transformations $f : V \rightarrow W$ are also called homomorphisms. This motivates the notation

$$\text{Hom}(V, W) \stackrel{\text{def}}{=} \{ f : V \rightarrow W \mid f \text{ is linear} \}.$$

- Linear transformations $f : V \rightarrow V$ are also called endomorphisms. This motivates the notation

$$\text{End}(V) \stackrel{\text{def}}{=} \text{Hom}(V, V).$$

The Set of Linear Transformations

- The set $\text{Hom}(V, W)$ is itself a vector space. Here, we define the addition of the linear transformations $f, g \in \text{Hom}(V, W)$ and scalar multiplication with $C \in F$ by

$$(f + g)(x) \stackrel{\text{def}}{=} f(x) + g(x) \quad \text{and} \quad (Cf)(x) \stackrel{\text{def}}{=} Cf(x).$$

- Let V and W be finite dimensional, $n = \dim(V)$ and $m = \dim(W)$. Because, for given bases of V and W , any $f \in \text{Hom}(V, W)$ can uniquely be represented by a matrix $A \in F^{n \times m}$, we have

$$\text{Hom}(V, W) \cong F^{n \times m} \quad \text{and} \quad \dim(\text{Hom}(V, W)) = nm.$$

Endomorphisms

- If V is finite dimensional,

$$\text{End}(V) = \text{Hom}(V, V) \cong F^{n \times n}$$

is, of course, a finite dimensional vector space, too.

- For any given basis, we can represent any $f \in \text{End}(V)$ by a matrix A .
- Interestingly, the determinant of this matrix A is invariant with respect to the choice of the basis. Namely, if S is a basis transition matrix, and $B = S^{-1}AS$, we still have

$$\begin{aligned}\det(B) &= \det(S^{-1}AS) = \det(S^{-1})\det(A)\det(S) \\ &= \det(S)^{-1}\det(A)\det(S) = \det(A).\end{aligned}$$

Determinant of an Endomorphism

- As mentioned earlier, invariants are our friends!
- In our context, since $\det(A) = \det(S^{-1}AS)$, the determinant is such an invariant. Consequently, the following definition makes sense.

Definition

- Let V be a finite dimensional vector space. The determinant of an endomorphism $f \in \text{End}(V)$ is defined by

$$\det(f) \stackrel{\text{def}}{=} \det(A),$$

where A is a representation matrix of f (with respect to any basis).

Orientation Preserving Isomorphisms

Definition

- A bijective linear transformation $f \in \text{End}(V)$ (isomorphism) is called orientation preserving if $\det(f) > 0$.
- Otherwise, if $\det(f) < 0$, it is called orientation reversing.

Intuition

- If you look into a mirror, you can see an isomorphic image of yourself (in $V = \mathbb{R}^3$), but the corresponding isomorphism f between you and your mirror image is orientation reversing, $\det(f) < 0$ (since your front and back are interchanged).

Trace of an Endomorphism

- There is no reason to believe that the determinant is the only expression that is invariant with respect to change of the basis. For example, the trace of an $(n \times n)$ -matrix A ,

$$\operatorname{Tr}(A) \stackrel{\text{def}}{=} \sum_{i=1}^n a_{ii}$$

satisfies

$$\operatorname{Tr}(AS) = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} s_{ji} \right) = \sum_{j=1}^n \left(\sum_{i=1}^n s_{ji} a_{ij} \right) = \operatorname{Tr}(SA).$$

- Thus, the trace of an endomorphism, $\operatorname{Tr}(f) \stackrel{\text{def}}{=} \operatorname{Tr}(A)$, is also well-defined, because it is independent of the choice of the basis:

$$\operatorname{Tr}(S^{-1}AS) = \operatorname{Tr}(SS^{-1}A) = \operatorname{Tr}(A).$$

Similar Matrices

Definition

- Two square matrices A and B are called similar, if there exists an invertible matrix S with $B = S^{-1}AS$ or, equivalently, $A = SBS^{-1}$.

Remarks:

- If A and B are similar, then $\det(A) = \det(B)$ and $\text{Tr}(A) = \text{Tr}(B)$.
- Practical necessary and sufficient conditions for checking whether two given matrices are similar will be discussed a bit later, in Lecture 7.

Contents

- Introduction
- Linear Transformations
- Kernels and Images
- Isomorphisms
- Sums of Vector Spaces
- Change of Basis
- Endomorphisms
- **Summary**

Summary – Part I

- In this lecture we have learned about linear transformations, $f : V \rightarrow W$, from a vector space V to a vector space W .
- We have introduced the subspaces $\ker(f) \subseteq V$ and $\operatorname{Im}(f) \subseteq W$.
- Additionally, we have established the dimension formula

$$\dim(\ker(f)) + \dim(\operatorname{Im}(f)) = \dim(V).$$

- In particular, for the map $f_A : x \rightarrow Ax$, $A \in F^{m \times n}$, we have

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

Summary – Part II

- In this lecture, we have also learned about isomorphisms, $f : V \rightarrow W$, which are bijective linear transformations.
- We have used such isomorphisms to prove that every finite dimensional vector space V (over F) is isomorphic to F^n .
- Finally, we have discussed that every linear transformation f between finite dimensional vector spaces can be represented by a matrix A , but this matrix depends on the choice of the bases of V and W .
- For $f : V \rightarrow V$ and a given basis transition matrix S we can convert our matrix representation of f by using the change of basis formula

$$B = S^{-1}AS.$$

- The determinant and trace are invariant w.r.t. such change of basis.

Chapter 6

Euclidean Spaces

Euclidean Spaces

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- Cross Product
- Unitary Vector Spaces
- Summary

Contents

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- Cross Product
- Unitary Vector Spaces
- Summary

Objectives

In this lecture we will learn about the definition and basic properties of Euclidean spaces, $\mathbb{R}^n$. Topics include:

- standard scalar products (“dot product”) and other inner products,
- length of a vector, angles between vectors, Cauchy-Schwarz inequality,
- orthonormal basis, orthogonal complements, orthogonal projection,
- Gram-Schmidt algorithm,
- orthogonal matrices, difference between rotations and reflections,
- the cross product in $\mathbb{R}^3$ and its algebraic properties, and
- short introduction to unitary spaces, differences between $\mathbb{R}^n$ and $\mathbb{C}^n$.

Contents

- Introduction
- **Euclidean Spaces**
- Orthogonal Basis
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- Cross Product
- Unitary Vector Spaces
- Summary

Euclidean Vector Spaces

- This lecture is about the vector space $\mathbb{R}^n$.
- It can be equipped with the standard scalar product,

$$\langle a, b \rangle \stackrel{\text{def}}{=} a^\top b = a_1 b_1 + a_2 b_2 + \dots + a_n b_n.$$

In some books the standard scalar product is called dot product.

Therefore, one sometimes writes " $a \bullet b$ " instead of " $\langle a, b \rangle$ ".

- The map $(a, b) \rightarrow \langle a, b \rangle$ is linear in both variables, symmetric, $\langle a, b \rangle = \langle b, a \rangle$, and positive definite, $\langle a, a \rangle > 0$ for all $a \neq 0$.
- In the following, however, we also allow more general inner products

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

as long as they have the above three properties.

Euclidean Scalar Products

Definition: The map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is called an (Euclidean) inner product if it satisfies for all $a, b, c \in \mathbb{R}^n$ and all $C \in \mathbb{R}$,

1. $\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$ and $\langle Ca, b \rangle = C\langle a, b \rangle$, (linearity)
2. $\langle a, b \rangle = \langle a, b \rangle$ (symmetry), and
3. $\langle a, a \rangle > 0$ if $a \neq 0$ (positive definiteness).

If these conditions hold, the pair $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ is called an Euclidean space.

Remark

- Recall that finite dimensional spaces over $\mathbb{R}$ are isomorphic to $\mathbb{R}^n$.

Therefore, at least for this lecture, it will be sufficient to focus on $\mathbb{R}^n$.

Euclidean Scalar Products

- In order to get used to working with Euclidean scalar products, it is helpful to first check that

$$\begin{aligned}\langle a + b, a + b \rangle &= \langle a, a \rangle + \langle a, b \rangle + \langle b, a \rangle + \langle b, b \rangle \\ &= \langle a, a \rangle + 2\langle a, b \rangle + \langle b, b \rangle,\end{aligned}$$

$$\langle a - b, a - b \rangle = \langle a, a \rangle - 2\langle a, b \rangle + \langle b, b \rangle,$$

$$\text{and} \quad \langle a + b, a - b \rangle = \langle a, a \rangle - \langle b, b \rangle.$$

By adding the first two equations, we also find the so-called parallelogram equation

$$\langle a + b, a + b \rangle + \langle a - b, a - b \rangle = 2(\langle a, a \rangle + \langle b, b \rangle).$$

Length of a Vector

- The Euclidean norm of a vector $a \in \mathbb{R}^n$ is defined as

$$\|a\| \stackrel{\text{def}}{=} \sqrt{\langle a, a \rangle}.$$

- Note that the positive definiteness and linearity of the scalar product implies that we have $\|a\| = 0$ if and only if $a = 0$.
- For the important special case that we use the standard scalar product, the Euclidean norm can be written in the form

$$\|a\| = \sqrt{a^\top a} = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}.$$

- Consequently, having Pythagoras' theorem in mind, one could also interpret $\|a\|$ as the length of the vector a .

Cauchy-Schwarz Inequality

Theorem: Let $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ be a Euclidean space. Then for all $a, b \in \mathbb{R}^n$:

$$|\langle a, b \rangle| \leq \|a\| \cdot \|b\| \quad (\text{Cauchy-Schwarz Inequality}).$$

Proof: The inequality holds for $b = 0$. Thus, we assume $b \neq 0$:

$$\begin{aligned} \|a\|^2 &= \left\| \frac{\langle a, b \rangle}{\langle b, b \rangle} b + a - \frac{\langle a, b \rangle}{\langle b, b \rangle} b \right\|^2 = \left| \frac{\langle a, b \rangle}{\langle b, b \rangle} \right|^2 \|b\|^2 + \left\| a - \frac{\langle a, b \rangle}{\langle b, b \rangle} b \right\|^2 \\ \implies \|a\|^2 &\geq \frac{\langle a, b \rangle^2}{\langle b, b \rangle} \implies |\langle a, b \rangle| \leq \|a\| \cdot \|b\|. \end{aligned}$$

Remark: The above proof merely uses linearity, symmetry, and positive definiteness property of $\langle \cdot, \cdot \rangle$. The same proof also works in infinite dimensional vector spaces (not our focus).

Angle Between Two Vectors

- For two non-zero vectors $a \neq 0$ and $b \neq 0$, Cauchy-Schwarz implies

$$-1 \leq \frac{\langle a, b \rangle}{\|a\| \|b\|} \leq 1.$$

- Consequently, the equation

$$\cos(\theta) = \frac{\langle a, b \rangle}{\|a\| \|b\|}$$

has a unique solution $\theta \in \mathbb{R}$ with $0 \leq \theta \leq \pi$. The number θ is called the angle between the vectors $a, b \in \mathbb{R}^n \setminus \{0\}$.

- The above definition is general but rooted in studying the standard scalar product, which provides a geometric motivation.

Angle Between Two Vectors

Example

- Let us consider the standard scalar product $\langle a, b \rangle = a^\top b$.
- We recall that the unit vectors $e_1, e_2, \dots, e_n$ are a basis of $\mathbb{R}^n$.
- These unit vectors satisfy

$$\langle e_i, e_j \rangle = \delta_{i,j} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- Thus, the angle between e_i and e_j is $\frac{\pi}{2}$ if $i \neq j$.

Otherwise, if $i = j$, the angle between e_i and e_i is 0.

Triangle Inequality

- The Cauchy-Schwarz inequality implies that

$$\begin{aligned}\|a + b\|^2 &= \langle a + b, a + b \rangle = \|a\|^2 + 2\langle a, b \rangle + \|b\|^2 \\ &\leq \|a\|^2 + 2\|a\| \cdot \|b\| + \|b\|^2 = (\|a\| + \|b\|)^2.\end{aligned}$$

- By taking the square-root, we obtain the triangle inequality,

$$\|a + b\| \leq \|a\| + \|b\|, \quad a, b \in \mathbb{R}^n.$$

Contents

- Introduction
- Euclidean Spaces
- **Orthogonal Basis**
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- Cross Product
- Unitary Vector Spaces
- Summary

Orthonormal Basis Vectors

Throughout, we assume that $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ is a Euclidean space.

Definition:

- A basis $q_1, q_2, \dots, q_n \in \mathbb{R}^n$ is called orthonormal if the q_i s are pairwise orthogonal and have norm 1,

$$\langle q_i, q_j \rangle = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Remark:

- If a basis $q_1, q_2, \dots, q_n \in \mathbb{R}^n$ is only orthogonal, $\langle q_i, q_j \rangle = 0$ for $i \neq j$, we can make it orthonormal by scaling all basis vectors, $q_i \leftarrow \frac{q_i}{\|q_i\|}$.

Existence of an Orthonormal Basis

Theorem

- Let $a_1, \dots, a_m \in \mathbb{R}^n$ be m given vectors with $m < n$. Then there exists a vector $a_{m+1} \in \mathbb{R}^n$ with $a_{m+1} \neq 0$, such that

$$\langle a_{m+1}, a_i \rangle = 0 \quad (1 \leq i \leq m).$$

Proof

- We can find a non-zero vector $a_{m+1} = C_1 e_1 + \dots + C_n e_n$, because the following homogeneous system has more variables than equations:

$$0 = \langle a_{m+1}, a_1 \rangle = C_1 \langle e_1, a_1 \rangle + \dots + C_n \langle e_n, a_1 \rangle$$

$$\vdots$$

$$0 = \langle a_{m+1}, a_m \rangle = C_1 \langle e_1, a_m \rangle + \dots + C_n \langle e_n, a_m \rangle.$$

Existence of an Orthonormal Basis

Corollary

- Every (finite dim.) Euclidean space, $\mathbb{R}^n$, has an orthonormal basis.

Proof (Part I)

- The first part of this proof shows that it is possible to find n pairwise orthogonal vectors $q_1, \dots, q_n \in \mathbb{R}^n$, with $\|q_i\| = 1$. We start with any non-zero vector $q_1 \in \mathbb{R}^n$, which, after scaling, satisfies $\|q_1\| = 1$.
Next, let $q_1, q_2, \dots, q_m \in \mathbb{R}^m$ be pairwise orthogonal vectors, with $\|q_i\| = 1$. If $m < n$, the theorem from the previous slide can then be used as an induction step to prove the above claim.

Existence of an Orthonormal Basis

Proof (Part II)

- Next, let us assume that the coefficients $C_1, \dots, C_n \in \mathbb{R}$ are such that

$$C_1q_1 + C_2q_2 + \dots + C_nq_n = 0,$$

for the pairwise orthogonal vectors $q_1, \dots, q_n$, with $\|q_i\| = 1$. Then,

$$0 = \langle q_i, C_1q_1 + C_2q_2 + \dots + C_nq_n \rangle = C_i \langle q_i, q_i \rangle = C_i$$

for all $i \in \{1, \dots, n\}$. Consequently, the vectors $q_1, \dots, q_n$ are an orthonormal basis.

Orthogonal Complement

Definition

- Let $V \subseteq \mathbb{R}^n$ be a linear subspace of $\mathbb{R}^n$. The set

$$V^\perp = \{ x \in \mathbb{R}^n \mid \langle x, z \rangle = 0 \text{ for all } z \in V \}$$

is called the orthogonal complement of V .

Properties

- The set $V^\perp$ is a linear subspace of $\mathbb{R}^n$. This is because if $x, y \in V^\perp$ and $C \in \mathbb{R}$, then the following equations hold for all $z \in V$:

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle = 0 \quad \text{and} \quad \langle Cx, z \rangle = C\langle x, z \rangle = 0.$$

- If $x \in V \cap V^\perp$, then $\langle x, x \rangle = 0$. Thus, we have $V \cap V^\perp = \{0\}$.

Orthogonal Complement

Theorem

- Let V be a linear subspace of $\mathbb{R}^n$ and let $V^\perp$ be its orthogonal complement. Then we have $\dim(V) + \dim(V^\perp) = n$ and

$$V + V^\perp = \{ v + w \mid v \in V, w \in V^\perp \} = \mathbb{R}^n.$$

Proof

- The set $V + V^\perp$ is a linear subspace of $\mathbb{R}^n$. It has an orthogonal basis. If we had $\dim(V + V^\perp) < n$, we could apply the result from Slide 6-15 to this basis, to construct a non-zero vector $x \in \mathbb{R}^n$ that is orthogonal to all vectors in $V + V^\perp$. But then we would have $x \in V^\perp$, which is a contradiction. Thus, we have $\dim(V + V^\perp) = n$. Since $V \cap V^\perp = \{0\}$, the statement of the theorem follows.

Orthogonal Projection

Definition

- Let V be a linear subspace of $\mathbb{R}^n$ and $q_1, q_2, \dots, q_s$ an orthonormal basis of V . The linear transformation $\pi : \mathbb{R}^n \rightarrow V$,

$$\pi(a) = \langle a, q_1 \rangle q_1 + \langle a, q_2 \rangle q_2 + \dots + \langle a, q_s \rangle q_s$$

is called an orthogonal projection onto the subspace V .

Basic Observations

- The vectors $v = \pi(a)$ and $w = a - \pi(a)$ are orthogonal to each other, as they satisfy $v \in V$, $w \in V^\perp$, and $v + w = a$.
- We have $\pi(a) = a$ for all $a \in V$.
- Similarly, we have $\pi(a) = 0$ for all $a \in V^\perp$.

Orthogonal Projection

Theorem

1. Let V be a linear subspace of $\mathbb{R}^n$ and let $\pi : \mathbb{R}^n \rightarrow V$ denote an orthogonal projection onto V . Then we have

$$\|a - \pi(a)\| \leq \|a - b\|$$

for all $a \in \mathbb{R}^n$ and all $b \in V$.

2. We have $\|a - \pi(a)\| = \|a - b\|$ if and only if $b = \pi(a)$.

Remark: The above theorem establishes the fact that $\pi(a)$ is the unique minimizer of the the optimization problem

$$\min_{x \in V} \|a - x\|; \quad \text{that is,} \quad \|a - \pi(a)\| < \|a - x\| \quad \text{for all} \quad x \in V.$$

Orthogonal Projection

Proof

- Since V is a linear subspace, $b, \pi(a) \in V$, we have $b - \pi(a) \in V$.
- Moreover, we recall that $a - \pi(a) \in V^\perp$. Thus, we have

$$\langle a - \pi(a), b - \pi(a) \rangle = 0,$$

which implies

$$\begin{aligned}\|a - b\|^2 &= \|a - \pi(a) - b + \pi(a)\|^2 \\&= \|a - \pi(a)\|^2 - 2 \underbrace{\langle a - \pi(a), b - \pi(a) \rangle}_{=0} + \|b - \pi(a)\|^2 \\&= \|a - \pi(a)\|^2 + \|b - \pi(a)\|^2.\end{aligned}$$

- Both statements of the theorem are an immediate consequence.

Isometric Isomorphism

Definition

- Two Euclidean vector spaces $(V, \langle \cdot, \cdot \rangle)$ and $(W, [\cdot, \cdot])$ are called isometrically isomorphic if there exists an vector space isomorphism $f : V \rightarrow W$ (see Lecture 5) such that for all $a, b \in V$:

$$\langle a, b \rangle = [f(a), f(b)].$$

Theorem

- Every finite dimensional Euclidean space $(V, \langle \cdot, \cdot \rangle)$ is isometrically isomorphic to $\mathbb{R}^n$, equipped with the standard scalar product.

Isometric Isomorphism

Proof:

- Let $q_1, q_2, \dots, q_n$ be an orthonormal basis of V and let $e_1, e_2, \dots, e_n$ be the unit vectors in $\mathbb{R}^n$. We already know from Lecture 5 that there exists an isomorphism $f : V \rightarrow \mathbb{R}^n$ that satisfies

$$f(q_i) = e_i \quad (1 \leq i \leq n).$$

Now, we have $\langle q_i, q_j \rangle = \delta_{i,j} = e_i^\top e_j = f(q_i)^\top f(q_j)$. By using the linearity of the scalar product and f , it follows that f is isometric.

Contents

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- **Gram-Schmidt Algorithm**
- Orthogonal Matrices
- Cross Product
- Unitary Vector Spaces
- Summary

Construction of an Orthonormal Basis

- The construction of an orthonormal basis is important in the context of many practical applications.
- A typical problem is that we have m given vectors $a_1, \dots, a_m \in \mathbb{R}^n$. But, a-priori, we know nothing about these vectors.
- Now, our goal is to find an orthonormal basis, $q_1, q_2, \dots, q_s \in \mathbb{R}^n$, of the vector space $V \subseteq \mathbb{R}^n$ that is spanned by the a_i s, such that

$$V = \text{span}(a_1, a_2, \dots, a_m) = \text{span}(q_1, q_2, \dots, q_s).$$

Here, s denotes the (unknown) dimension of V .

Construction of an Orthonormal Basis

- We may assume that the a_i s are all non-zero—otherwise, we delete them from our list of given vectors.
- Given this assumption, we know that $V \cong \mathbb{R}^s$ for some s with $1 \leq s \leq m$. And, consequently, the above theorem ensures that V has an orthonormal basis $q_1, q_2, \dots, q_s \in \mathbb{R}^n$ with

$$V = \text{span}(a_1, a_2, \dots, a_m) = \text{span}(q_1, q_2, \dots, q_s).$$

Question: How can we systematically construct such an orthonormal basis?

Gram-Schmidt Algorithm

- **Input:** Vectors $a_1, a_2, \dots, a_m \in \mathbb{R}^m$. Initialize $k = 0$.
- **Loop:** For $i = 1, \dots, m$:
 1. For $j = 1, \dots, k$, set $r_{ji} = \langle q_j, a_i \rangle$.
 2. Set $p_{k+1} = a_i - r_{1i}q_1 - \dots - r_{ki}q_k$. (ORTHOGONALIZATION)
 3. If $p_{k+1} \neq 0$ (DEPENDENCY TEST)
 - Set $r_{k+1,i} = \|p_{k+1}\|$.
 - Set $q_{k+1} = \frac{p_{k+1}}{r_{k+1,i}}$. (NORMALIZATION)
 - Update the dimension counter $k \leftarrow k + 1$.

End

End

- **Output:**
 - The dimension $s = k$ of the vector space $V = \text{span}(a_1, \dots, a_m)$.
 - An orthonormal basis $(q_1, q_2, \dots, q_s)$ of V .

Gram-Schmidt Algorithm

Properties:

- The second step inside the for-loop (for the current i and k) ensures

$$a_i = r_{1i}q_1 + \dots + r_{ki}q_k + p_{k+1}.$$

- Thus, there are only two possibilities. If $p_{k+1} = 0$, then

$$a_i = r_{1i}q_1 + \dots + r_{ki}q_k \in \text{span}(q_1, \dots, q_k).$$

Or we have $p_{k+1} \neq 0$, but then Step 3 ensures that

$$a_i = r_{1i}q_1 + \dots + r_{ki}q_k + r_{k+1,i}q_{k+1} \in \text{span}(q_1, \dots, q_{k+1}).$$

- Thus, upon termination, the span of the outputs vectors $q_1, q_2, \dots, q_s$ coincides with the span of the input vectors $a_1, a_2, \dots, a_m$.

Gram-Schmidt Algorithm

Properties:

- As we may assume $a_1 \neq 0$, the first iteration initially sets $q_1 = \frac{a_1}{\|a_1\|}$, which is a vector with norm 1.
- Now, assume that $q_1, q_2, \dots, q_k$ are already pairwise orthogonal and satisfy $\|q_i\| = 1$. Then, the auxiliary vector p_{k+1} in Step 2 satisfies

$$\begin{aligned}\langle q_j, p_{k+1} \rangle &= \langle q_j, a_i \rangle - r_{1i} \langle q_j, q_1 \rangle - \dots - r_{ki} \langle q_j, q_k \rangle \\ &= \langle q_j, a_i \rangle - r_{ji} = 0\end{aligned}$$

for all $j \in \{1, \dots, k\}$. Thus, if $p_{k+1} \neq 0$, the vector q_{k+1} will be orthogonal on all $q_1, q_2, \dots, q_k$.

- By induction over k , it follows that the algorithm outputs an orthonormal basis $q_1, q_2, \dots, q_s$ of the span of $a_1, a_2, \dots, a_m$.

Gram-Schmidt Algorithm

Summary:

- Gram-Schmidt's algorithm constructs an orthonormal basis $(q_1, q_2, \dots, q_s)$ of the vector space $V = \text{span}(a_1, \dots, a_m)$.

Remark:

- Gram-Schmidt's algorithm can be used to find an orthonormal basis of any (finite dimensional) Euclidean space.
- Namely, if the input vectors $a_1, a_2, \dots, a_n \in \mathbb{R}^n$ are an arbitrary basis of $\mathbb{R}^n$, then the output vectors $a_1, a_2, \dots, a_n \in \mathbb{R}^n$ are an orthonormal basis of $\mathbb{R}^n$, with respect to the given scalar product.

Gram-Schmidt Algorithm

Example 1

- Let us first consider the standard scalar product $\langle a, b \rangle = a^\top b$.
- Our example assumes that the given vectors are

$$a_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad a_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \text{and} \quad a_3 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

- Let us apply the Gram-Schmidt algorithm to compute an orthonormal basis of the vector space

$$\text{span}(a_1, a_2, a_3) \subseteq \mathbb{R}^2.$$

Gram-Schmidt Algorithm

Example 1 (continued)

- **First Iteration; Step 1:** skipped, since $k = 0$.
- **First Iteration; Step 2:** set

$$p_1 = a_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- **First Iteration; Step 3:** set

$$r_{11} = \|a_1\| = \sqrt{2},$$

$$q_1 = \frac{a_1}{\|a_1\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},$$

$$\text{and } k \leftarrow 1.$$

Gram-Schmidt Algorithm

Example 1 (continued)

- **Second Iteration; Step 1:** Since $k = 1$, we compute

$$r_{12} = \langle q_1, a_2 \rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}^T \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\sqrt{2}$$

- **Second Iteration; Step 2:** set

$$p_2 = a_2 - r_{12}q_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - 2\sqrt{2} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 0.$$

- **Second Iteration; Step 3:** since $p_2 = 0$, we have a dependency and, consequently, we skip this step.

Gram-Schmidt Algorithm

Example 1 (continued)

- **Third Iteration; Step 1:** We still have $k = 1$. Thus, we compute

$$r_{13} = \langle q_1, a_3 \rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}^T \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{2}}$$

- **Third Iteration; Step 2:** set

$$p_2 = a_3 - r_{13}q_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} - \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} -\frac{3}{2} \\ \frac{3}{2} \end{pmatrix}.$$

- **Third Iteration; Step 3:** set

$$r_{23} = \|p_2\| = \frac{3}{\sqrt{2}}, \quad q_2 = \frac{p_2}{r_{23}} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad k \leftarrow 2,$$

End.

Gram-Schmidt Algorithm

Example 1 (summary)

- For our example, the algorithm returns the dimension of the span of the input vectors, $s = k = 2$, and the orthonormal basis vectors

$$q_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{and} \quad q_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Note that $\|q_1\| = \|q_2\| = 1$ and $\langle q_1, q_2 \rangle = 0$, as required.

- The coefficients of the original vectors with respect to the orthonormal basis are given by the r_{ji} s,

$$\begin{aligned} a_1 &= r_{11}q_1 = \sqrt{2}q_1, & a_2 &= 2\sqrt{2}q_1, \\ \text{and} \quad a_3 &= r_{13}q_1 + r_{23}q_2 = \frac{1}{\sqrt{2}}q_1 + \frac{3}{\sqrt{2}}q_2. \end{aligned}$$

Contents

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- Gram-Schmidt Algorithm
- **Orthogonal Matrices**
- Cross Product
- Unitary Vector Spaces
- Summary

Euclidean Space with Standard Scalar Product

- The Gram-Schmidt method is a very general algorithm. As discussed on the previous slides, it can be used in any Euclidean space.
- We will come back to this general case later (see Lecture 8 on quadratic forms).
- Regarding the considerations in this section, however, it will be sufficient to focus on the standard scalar product. This is because any finite dimensional Euclidean vector space is isometrically isomorphic to $\mathbb{R}^n$, equipped with the standard scalar product.

On the following slides, we assume that $\langle a, b \rangle = a^\top b$ is the standard scalar product.

Gram-Schmidt Method for Standard Scalar Products

- If $\langle a, b \rangle = a^\top b$ is the standard Euclidean scalar product, the output matrix $Q = (q_1, q_2, \dots, q_s) \in \mathbb{R}^{n \times s}$ satisfies

$$Q^\top Q = \begin{pmatrix} \langle q_1, q_1 \rangle & \dots & \langle q_1, q_s \rangle \\ \vdots & & \vdots \\ \langle q_s, q_1 \rangle & \dots & \langle q_s, q_s \rangle \end{pmatrix} = E \in \mathbb{R}^{s \times s}.$$

- Moreover, the input matrix $A = (a_1, a_2, \dots, a_m)$ satisfies

$$QR = A \quad \text{with} \quad R \stackrel{\text{def}}{=} Q^\top A \in \mathbb{R}^{s \times m}.$$

- And the matrix $R = Q^\top A$ is an upper triangular matrix. Its non-zero coefficients correspond to the r_{ij} s in the above algorithm.

Orthogonal Projection in Matrix Form

- Since $Q = (q_1, q_2, \dots, q_s)$ is an orthonormal basis of the vector space $V = \text{span}(a_1, a_2, \dots, a_m)$, the orthogonal projection π onto V can be written in the form:

$$\pi(a) = \langle a, q_1 \rangle q_1 + \langle a, q_2 \rangle q_2 + \dots + \langle a, q_s \rangle q_s = QQ^T a, \quad (a \in \mathbb{R}^n).$$

- Consequently, the matrix $P = QQ^T$ is a representation of the linear transformation π . It is called a projection matrix. It satisfies

$$P^T = (QQ^T)^T = QQ^T = P$$

$$\text{and} \quad P^2 = QQ^T QQ^T = QEQ^T = QQ^T = P.$$

Orthogonal Matrices

- For the special case that the input vectors $a_1, \dots, a_n \in \mathbb{R}^n$ are a basis of $\mathbb{R}^n$, the Gram-Schmidt algorithm terminates with $s = n$.
- In this case, the output vectors $q_1, \dots, q_n$ are an orthonormal basis of $\mathbb{R}^n$; that is, $Q = (q_1, \dots, q_n)$ is invertible. In particular, if we use the standard scalar product $\langle a, b \rangle = a^\top b$, the matrix Q satisfies

$$Q^\top Q = E \quad \implies \quad Q^{-1} = Q^\top.$$

Definition

- A square matrix Q is called orthogonal if $Q^\top Q = E$.

Orthogonal Matrices

Properties

- We recall from Lecture 2 that the left inverse of a full-rank matrix is also a right inverse and vice versa. Hence, orthogonal matrices satisfy

$$Q^{\top}Q = E \quad \text{and} \quad QQ^{\top} = E.$$

- It follows that if Q is an orthogonal matrix, then $Q^{\top}$ is also an orthogonal matrix.
- If Q and P are both orthogonal matrices, then we have

$$(QP)^{\top}(QP) = P^{\top}Q^{\top}QP = P^{\top}EP = P^{\top}P = E.$$

Thus, the product of orthogonal matrices is orthogonal.

Orthogonal Group

Properties

- Because orthogonal matrices Q are invertible, $Q \in \text{GL}(n, \mathbb{R})$, and because the product of two orthogonal matrices is orthogonal, the set of (real) orthogonal matrices $\text{O}(n, \mathbb{R}) \subseteq \text{GL}(n, \mathbb{R})$ is a subgroup of $\text{GL}(n, \mathbb{R})$ and, hence, itself a group (see also Lecture 4).

- If Q is orthogonal, we have $E = Q^T Q$. Consequently, we also have

$$1 = \det(E) = \det(Q^T Q) = \det(Q)^2 \implies \det(Q) \in \{\pm 1\}.$$

- An orthogonal matrix Q with $\det(Q) = 1$ is called a rotation matrix. The set of rotation matrices, $\text{SO}(n, \mathbb{R})$, is called the special orthogonal group. It is a subgroup of $\text{O}(n, \mathbb{R})$.

Orthogonal Matrices

Example 1

- The unit matrix E is an orthogonal matrix, because $E^T E = EE = E$.

Example 2

- The matrix

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is an orthogonal matrix, because

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E.$$

Orthogonal Matrices

Example 3

- Let $w \in \mathbb{R}^n$ be any vector with length one, $w^\top w = 1$. Then, the reflection matrix

$$Q = E - 2ww^\top$$

is an orthogonal matrix, because it satisfies

$$\begin{aligned} Q^\top Q &= (E - 2ww^\top)(E - 2ww^\top) \\ &= E - 2Eww^\top - 2ww^\top E + 4ww^\top ww^\top \\ &= E - 4ww^\top + 4w \underbrace{w^\top w}_{=1} w^\top = E. \end{aligned}$$

- Note that Q “reflects” the vector w , as $Qw = w - 2w = -w$.
- If $v \in \mathbb{R}^n$ is orthogonal to w , such that $w^\top v = 0$, then $Qv = v$.

Orthogonal Matrices

Example 4

- The (2×2) -rotation matrix

$$Q = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

is for all $\alpha \in \mathbb{R}$ an orthogonal matrix, because

$$\begin{aligned} & \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}^T \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\alpha)^2 + \sin(\alpha)^2 & 0 \\ 0 & \sin(\alpha)^2 + \cos(\alpha)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

- Given a vector $v \in \mathbb{R}^2$, Qv rotates v counterclock-wise by α .

Orthogonal Matrices

Example 5

- In the three dimensional space R^3 , different types of rotations are possible. For example, the matrix

$$Q_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

can be used to “rotate” around the x -axis, with angle $\alpha \in \mathbb{R}$. Similar to the previous example, Q_x is an orthogonal matrix, $Q_x^T Q_x = E$.

Orthogonal Matrices

Example 5 (continued)

- Similarly, we could rotate around the y - or z -axis,

$$Q_y = \begin{pmatrix} \cos(\beta) & 0 & -\sin(\beta) \\ 0 & 1 & 0 \\ \sin(\beta) & 0 & \cos(\beta) \end{pmatrix}$$

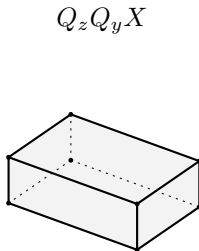
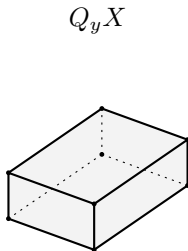
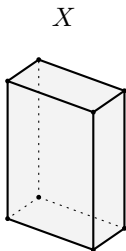
$$Q_z = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- Note that first rotating around the y -axis then around the z -axis differs from rotating in opposite order, $Q_y Q_z \neq Q_z Q_y$.

Visualization

Example 5 (continued)

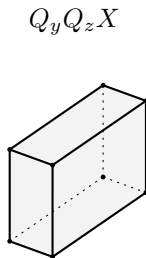
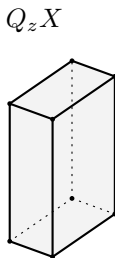
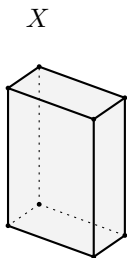
- We set $\beta = \gamma = \frac{\pi}{2}$ to define Q_y and Q_z .
- Consider a cuboid $X \subseteq \mathbb{R}^3$. First rotating around the y -axis and then around the z -axis yields:



Visualization

Example 5 (continued)

- The other way around, first rotating around the z -axis then around the y -axis yields:



Note that $Q_z Q_y X \neq Q_y Q_z X$.

Contents

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- **Cross Product**
- Unitary Vector Spaces
- Summary

Cross Product

- The three-dimensional Euclidean space, $\mathbb{R}^3$, equipped with the standard scalar product, $\langle a, b \rangle = a^\top b$, plays a special role in certain applied sciences, for example, physics.
- Therefore, some definitions in linear algebra are tailored to $\mathbb{R}^3$.

Definition:

- For two vectors $a, b \in \mathbb{R}^3$, we define their cross product as

$$a \times b \stackrel{\text{def}}{=} \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} \in \mathbb{R}^3.$$

Cross Product

Properties

- The cross product is linear in its first argument. This means that for all $a, b, c \in \mathbb{R}^3$ and all $C \in \mathbb{R}$, we have

$$(a + b) \times c = (a \times c) + (b \times c) \quad \text{and} \quad (Ca) \times b = C(a \times b).$$

- The cross product is skew-symmetric, $a \times b = -b \times a$.
- The skew-symmetry property implies that $a \times a = 0$.
- Skew-symmetry and linearity in the first argument imply linearity in the second argument. Consequently, the cross product is bilinear.
And, in particular, we have $a \times 0 = 0$ and $0 \times b = 0$.

Cross Product

Properties

- In addition to being bilinear and skew-symmetric the cross product and the standard scalar product $\langle a, b \rangle = a^\top b$ satisfy for all $a, b \in \mathbb{R}^3$,

$$\langle a, a \times b \rangle = 0 \quad \text{and} \quad \|a \times b\|^2 + \langle a, b \rangle^2 = \|a\|^2 \|b\|^2.$$

- Since $\langle a, a \times b \rangle = 0$ and skew-symmetry imply $\langle b, a \times b \rangle = 0$, the vector $a \times b$ is orthogonal to a and b .
- Finally, if a and b are linearly independent, the linear transformation

$$x \rightarrow Ax \quad \text{with} \quad A = \begin{pmatrix} a & b & a \times b \end{pmatrix} \in \mathbb{R}^{3 \times 3}$$

is orientation preserving; that is, $\det(A) > 0$ (= "right-hand rule").

The above properties together uniquely characterize the cross-product. (Proof: Exercise.)

Rotations

- The above algebraic characterization of the cross-product implies that we have

$$(Qa) \times (Qb) = Q(a \times b)$$

for all rotation matrices $Q \in \text{SO}(3, \mathbb{R})$, recalling that such matrices satisfy $QQ^T = Q^TQ = E$ and $\det(Q) = 1$.

- Note that the above equation does not hold for general orthogonal matrices Q . For example, the unit vectors $e_1, e_2, e_3 \in \mathbb{R}^3$ satisfy

$$e_1 \times e_2 = e_3 \quad \text{but} \quad (-e_1) \times e_2 = -e_3.$$

Consequently, for the reflection matrix $R = (-e_1, e_2, e_3)$, we have

$$(Re_1) \times (Re_2) \neq R(e_1 \times e_2) \quad \text{but} \quad R^T R = E.$$

Cross Product

Application Examples (not part of this course)

- Torque = (rel.) position vector $\times$ force vector.
- Angular momentum = (rel.) position vector $\times$ linear momentum.
- Magnetic force = charge times velocity $\times$ magnetic field.
- Vorticity = $\nabla \times$ velocity field.

There are zillions of other applications of the cross product, but the basic physical intuition is that cross products can be used to describe rotational phenomena and electromagnetic interactions in $\mathbb{R}^3$.

Contents

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- Cross Product
- **Unitary Vector Spaces**
- Summary

Hermitian Scalar Products

- The definitions from the previous section can be extended for vector spaces over the field $F = \mathbb{C}$.
- Instead of the Euclidean scalar product, one then introduces an Hermitian scalar product.
- The standard Hermitian scalar product for the vector space $V = \mathbb{C}^n$,

$$\langle a, b \rangle = a^T \bar{b} = \sum_{k=1}^n a_k \bar{b}_k,$$

is linear in the first argument, antisymmetric $\langle a, b \rangle = \overline{\langle b, a \rangle}$, and positive definite, $\langle a, a \rangle > 0$ if $a \neq 0$ (since $\langle a, a \rangle = \overline{\langle a, a \rangle}$, $\langle a, a \rangle \in \mathbb{R}$).

- Similar to Euclidean space, we'll again allow more general inner products as long as they have the above three properties.

Unitary Vector Space

Definition: Let V be a vector space of the complex numbers. The map $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$ is called an Hermitian inner product if it satisfies for all $a, b, c \in V$ and all $C \in \mathbb{C}$,

1. $\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$ and $\langle Ca, b \rangle = C\langle a, b \rangle$, (linearity)
2. $\langle a, b \rangle = \overline{\langle b, a \rangle}$ (antisymmetry), and
3. $\langle a, a \rangle > 0$ if $a \neq 0$ (positive definiteness).

If these conditions hold, the pair $(V, \langle \cdot, \cdot \rangle)$ is called a unitary space.

Caution: The Hermitian inner product is not linear in the second argument. Instead, we have

$$\langle a, Cb \rangle = \overline{\langle Cb, a \rangle} = \overline{C \langle b, a \rangle} = \overline{C} \overline{\langle b, a \rangle} = \overline{C} \langle a, b \rangle.$$

Unitary Vector Space

Properties. Unitary spaces are very similar to Euclidean spaces:

- Two vectors $a, b \in V$ are called orthogonal if $\langle a, b \rangle = 0$.
- The norm/length of a vector is again defined by $\|a\| \stackrel{\text{def}}{=} \sqrt{\langle a, a \rangle}$.
- The Cauchy-Schwarz and triangle inequality hold in unitary spaces as

$$|\langle a, b \rangle| \leq \|a\| \|b\| \quad \text{and} \quad \|a + b\| \leq \|a\| + \|b\|.$$

Here, we recall our definition $|z| \stackrel{\text{def}}{=} \sqrt{z\bar{z}}$ for $z \in \mathbb{C}$, see Lecture 4.

- Every finite dimensional unitary vector space has an orthonormal basis and is isometrically isomorphic to $\mathbb{C}^n$, equipped with the standard Hermitian inner product.

The proofs are all analogous to the Euclidean space!

Unitary Vector Space

- As unitary spaces are very similar to Euclidean spaces, we keep this section extremely short.
- For example, it should be clear that the Gram-Schmidt algorithm also works in unitary spaces.
- The definition of the orthogonal complement $W^\perp$ of a subspace W of a unitary vector space V is defined—in analogy to the Euclidean space—as the set of vectors are that orthogonal to W . We also have

$$W + W^\perp = V \quad \text{and} \quad W \cap W^\perp = \{0\}.$$

Unitary Matrices

- For the case that $\langle a, b \rangle = a^\top \bar{b}$ denotes the standard Hermitian scalar product and $u_1, u_2, \dots, u_n$ an orthonormal basis of $\mathbb{C}^n$, the matrix

$$U = (u_1, u_2, \dots, u_n)$$

satisfies

$$\overline{U}^\top U = \begin{pmatrix} \langle u_1, u_1 \rangle & \dots & \langle u_n, u_1 \rangle \\ \vdots & & \vdots \\ \langle u_1, u_n \rangle & \dots & \langle u_n, u_n \rangle \end{pmatrix} = E \in \mathbb{C}^{n \times n}.$$

- Note that this construction is analogous to the construction of orthogonal matrices in Euclidean spaces.

Unitary Matrices

Definition

- A square matrix $U \in \mathbb{C}^{n \times n}$ is called unitary if

$$\overline{U}^T U = E.$$

Properties

- The columns of a unitary matrix are an orthonormal basis of $\mathbb{C}^n$.
- We have $U^{-1} = \overline{U}^T$.
- Consequently, we also have $\overline{U}^T U = U \overline{U}^T = E$.
- If a unitary matrix is real, $\overline{U} = U$, it is orthogonal, as $U^T U = E$.

Hermitian Matrices

Definition

- A square matrix $H \in \mathbb{C}^{n \times n}$ is called Hermitian if $H = \overline{H}^\top$.

Properties

- If $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$ are the real and imaginary part of a Hermitian matrix $H = A + \mathbf{i}B$, it follows that

$$\overline{H}^\top = A^\top - \mathbf{i}B^\top = A + \mathbf{i}B = H.$$

Thus, A must be a symmetric matrix, $A = A^\top$, and B must be a skew-symmetric matrix, $B = -B^\top$.

- The diagonal coefficients of a Hermitian matrix are real.

Contents

- Introduction
- Euclidean Spaces
- Orthogonal Basis
- Gram-Schmidt Algorithm
- Orthogonal Matrices
- Cross Product
- Unitary Vector Spaces
- **Summary**

Summary

In this lecture, we have learned about:

- basic definition of the Euclidean space $\mathbb{R}^n$,
- proof of the Cauchy-Schwarz inequality and the triangle inequality,
- construction of an orthonormal basis (Gram-Schmidt algorithm)
- properties and representation of orthogonal projections,
- general orthogonal matrices versus rotation matrices,
- algebraic characterization of the cross product in $\mathbb{R}^3$, and
- the unitary space $\mathbb{C}^n$ including the Hermitian scalar product as well as the basic definition of unitary and Hermitian matrices.

Chapter 7

Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

- Introduction
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- Generalized Eigenspaces
- Jordan Normal Form
- Matrix Powers
- Summary
- Appendix: Fundamental Theorem of Algebra

Contents

- **Introduction**
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- Generalized Eigenspaces
- Jordan Normal Form
- Matrix Powers
- Summary
- Appendix: Fundamental Theorem of Algebra

Objectives

In this lecture we will learn about the definition and basic properties of eigenvalues and eigenvectors. Topics include:

- eigenvalues, eigenvectors, and eigenspaces,
- characteristic polynomials and the fundamental theorem of algebra,
- difference between geometric multiplicity and algebraic multiplicity,
- characterization of diagonalizable matrices,
- generalized eigenspaces and Cayley Hamilton theorem,
- co-prime polynomials and direct sums of generalized eigenspaces,
- writing matrices as sums of diagonalizable and nilpotent matrices,
- Jordan normal form and the computation of matrix powers.

Contents

- Introduction
- **Eigenvalues and Eigenvectors**
- Characteristic Polynomials
- Generalized Eigenspaces
- Jordan Normal Form
- Matrix Powers
- Summary
- Appendix: Fundamental Theorem of Algebra

Representation of Endomorphisms

- We recall from Lecture 5 that an endomorphism $f : V \rightarrow V$ is a linear transformation from a vector space V to itself.
- As before, we focus on the case that V is finite dimensional.
- Consequently, as soon as we choose a basis of V , f can be represented by a matrix A .

Question: Can we choose the basis of V such that the representation matrix A of f has a simple structure?

For example, can we choose the basis such that A is diagonal?

Representation of Endomorphisms

- As discussed in Lecture 5, if the matrix A is a representation of the linear transformation $f : V \rightarrow V$ with respect to any basis, we can change the basis by computing a new representation matrix

$$B = S^{-1}AS.$$

Here, S denotes the basis transition matrix.

- Thus, the question from the previous slide can equivalently be formulated as follows:

Given a square matrix A , can we construct an invertible matrix S , such that the matrix $B = S^{-1}AS$ has a simple structure? If possible, we would like to achieve that B is a diagonal matrix.

Eigenvectors and Eigenvalues

Motivation:

- Let us assume that the matrix equation

$$B = S^{-1}AS \iff AS = SB$$

is satisfied for a square matrix A and a diagonal matrix B .

- If this equation holds, we can denote the diagonal coefficients of B by $\lambda_1, \lambda_2, \dots, \lambda_n \in F$ (for example, $F = \mathbb{R}$ or $F = \mathbb{C}$) and write it as

$$Av_i = \lambda_i v_i \quad (1 \leq i \leq n),$$

where v_i denotes the i -th column vector of S . This equation motivates the definition of eigenvectors (see next slide).

Eigenvectors and Eigenvalues

Definition

- Let $A \in F^{n \times n}$ be a given matrix. A non-zero vector, $v \neq 0$, is called an eigenvector of A if there exists a constant $\lambda \in F$ with

$$Av = \lambda v.$$

The scalar λ is then called an eigenvalue of A .

Remark

- The above definition can also be used for endomorphisms $f : V \rightarrow V$. In this case, a non-zero vector $v \in V$ is called an eigenvector of f , with eigenvalue $\lambda \in F$, if $f(v) = \lambda v$.

Eigenvectors and Eigenvalues

More Remarks

- Recall that the dimension of the null space of a square matrix A is larger than or equal to 1 if and only if there exists a vector $v \neq 0$ with $Av = 0$. That is, the dimension of the null space of A is larger than or equal to 1 if and only if 0 is an eigenvalue of A .
- By negating the previous statement, we can alternatively formulate it as: A square matrix A is invertible if and only if all of its eigenvalues are nonzero.

Eigenvectors and Eigenvalues

Theorem

- If $v_1, v_2, \dots, v_m$ are eigenvectors of $A \in F^{n \times n}$ with pairwise different eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m \in F$, then they are linearly independent.

Proof

- Our proof is indirect. Thus, let us assume that the eigenvectors $v_1, v_2, \dots, v_m$ are linear dependent.
- Since we can always leave away some vectors, we may further assume that the vectors $v_1, v_2, \dots, v_{m-1}$ are linearly independent (otherwise, we make m smaller and use an induction argument).

Eigenvectors and Eigenvalues

Proof (continued)

- The above constructions imply that

$$C_1 v_1 + C_2 v_2 + \dots + C_m v_m = 0.$$

for constants $C_1, C_2, \dots, C_m$ that are not all zero.

- Consequently, we also have

$$\begin{aligned} 0 &= A(C_1 v_1 + C_2 v_2 + \dots + C_m v_m) \\ &= C_1 \lambda_1 v_1 + C_2 \lambda_2 v_2 + \dots + C_m \lambda_m v_m, \end{aligned}$$

where we have used the relations $Av_i = \lambda_i v_i$ for $1 \leq i \leq m$.

Eigenvectors and Eigenvalues

Proof (continued)

- By multiplying the first equation with λ_m and subtracting it from the second equation we further find that

$$\begin{aligned} 0 &= C_1(\lambda_1 - \lambda_m)v_1 + \dots + C_m(\lambda_m - \lambda_m)v_m \\ &= C_1(\lambda_1 - \lambda_m)v_1 + \dots + C_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1}. \end{aligned}$$

- Since the λ_i s are pairwise different and since the vectors $v_1, \dots, v_{m-1}$ are linearly independent, it follows that $C_1 = C_2 = \dots = C_{m-1} = 0$, and, by substituting this into the previous equation, also $C_m = 0$.
But this is a contradiction, completing our proof.

Eigenvectors and Eigenvalues

Consequences

- The above theorem implies that if $v_1, v_2, \dots, v_n$ are n eigenvectors of the $(n \times n)$ matrix A with pairwise different eigenvalues $\lambda_1, \dots, \lambda_n$, then

$$B = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad \text{and} \quad S = (v_1, v_2, \dots, v_n)$$

satisfy $AS = SB$ and $B = S^{-1}AS$, since S is invertible.

But now there are at least two remaining questions:

How can we find the eigenvectors and eigenvalues of A ?

Are there n pairwise different eigenvalues? (and what if not?)

Contents

- Introduction
- Eigenvalues and Eigenvectors
- **Characteristic Polynomials**
- Generalized Eigenspaces
- Jordan Normal Form
- Matrix Powers
- Summary
- Appendix: Fundamental Theorem of Algebra

Properties of Eigenvalues

The equation $Av = \lambda v$ can, alternatively, be written as

$$0 = Av - \lambda v = (A - \lambda E)v.$$

Consequently, the following statements are equivalent:

1. The scalar $\lambda \in F$ is an eigenvalue of the $(n \times n)$ -matrix A .
2. There exists a vector v with $(A - \lambda E)v = 0$.
3. The dimension of the null-space of A satisfies $\text{nullity}(A) \geq 1$.
4. The matrix $A - \lambda E$ is not invertible.
5. We have $\text{rank}(A) \leq n - 1$.
6. We have $\det(A - \lambda E) = 0$.

Characteristic Polynomial

- The last condition from the previous slide, $\det(A - \lambda E) = 0$, can be regarded as a polynomial equation in λ . This observation turns out to be useful for answering our remaining questions about diagonalization.

Definition

- Let $A \in F^{n \times n}$ be a given matrix, with $F = \mathbb{R}$ or $F = \mathbb{C}$. The function $p : F \rightarrow F$, given by

$$p(\lambda) = \det(A - \lambda E) \quad (\lambda \in F),$$

is called the characteristic polynomial of A .

Characteristic Polynomial

Example

- Let us consider the matrix

$$A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \quad \Rightarrow \quad A - \lambda E = \begin{pmatrix} 4 - \lambda & 2 \\ 1 & 3 - \lambda \end{pmatrix}.$$

- The characteristic polynomial is given by

$$p(\lambda) = \det(A - \lambda E) = (4 - \lambda)(3 - \lambda) - 2 = \lambda^2 - 7\lambda + 10.$$

- The equation $\lambda^2 - 7\lambda + 10 = 0$ has the solutions $\lambda_1 = 2$ and $\lambda_2 = 5$.

Thus, the matrix A has the eigenvalues 2 and 5.

Characteristic Polynomial

Example (continued)

- Let us first analyze the matrix

$$A - \lambda_1 E = \begin{pmatrix} 4 - 2 & 2 \\ 1 & 3 - 2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}.$$

Since $\lambda_1 = 2$ is an eigenvalue, the null space of this matrix has at least dimension 1. We can see this directly, as the first and second row are linearly dependent. More precisely, we have

$$\text{null}(A - \lambda_1 E) = \text{span}(v_1) \quad \text{with} \quad v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus, v_1 is an eigenvector of A , with eigenvalue λ_1 .

Characteristic Polynomial

Example (continued)

- Similarly, we analyze the matrix

$$A - \lambda_2 E = \begin{pmatrix} 4 - 5 & 2 \\ 1 & 3 - 5 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}.$$

A basis of the null-space of $A - \lambda_2 E$ is given by

$$\text{null}(A - \lambda_2 E) = \text{span}(v_2) \quad \text{with} \quad v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Thus, v_2 is also an eigenvector of A , with eigenvalue λ_2 .

Characteristic Polynomial

Example (summary)

- In summary, the matrix

$$S = \begin{pmatrix} v_1 & v_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$$

is an invertible basis transition matrix, $\det(S) = 3$, and

$$S^{-1} = \frac{1}{3} \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}.$$

- The final diagonalization of A is given by $B = S^{-1}AS$; that is,

$$B = \frac{1}{3} \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}.$$

Polynomials

Remarks

- The example from the previous slides is relatively “simple”.
- In general, there arises the question: what do we know about the characteristic polynomial

$$p(\lambda) = \det(A - \lambda E) ?$$

- Unfortunately, this question is not easy to answer.
- In order to make things simpler, we first remark that we will only focus on the cases $\lambda \in \mathbb{R}$ and $\lambda \in \mathbb{C}$. For more general fields F (especially finite ones), the considerations on the following slides are not directly applicable.

Real versus Complex Vector Space

Example

- Let us first assume that $F = \mathbb{R}$ and consider to matrix

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

- The characteristic polynomial of this matrix,

$$p(\lambda) = \det(A - \lambda E) = \lambda^2 + 1,$$

has no roots in $\mathbb{R}$. Thus, the matrix A has no real valued eigenvalues.

- If we set $F = \mathbb{C}$, however, $\lambda_1 = \mathbf{i}$ and $\lambda_2 = -\mathbf{i}$ are eigenvalues of A .

Real versus Complex Vector Space

Example (continued)

- For the case $F = \mathbb{C}$, we have $\lambda_1 = \mathbf{i}$ and

$$A - \lambda_1 E = \begin{pmatrix} -\mathbf{i} & 1 \\ -1 & -\mathbf{i} \end{pmatrix}.$$

- A basis of the null-space of this matrix is given by

$$\text{null}(A - \lambda_1 E) = \text{span}(v_1) \quad \text{with} \quad v_1 = \begin{pmatrix} -\mathbf{i} \\ 1 \end{pmatrix}.$$

- Consequently, v_1 is an eigenvector of A with eigenvalue $\lambda_1 = \mathbf{i}$.

Real versus Complex Vector Space

Example (continued)

- Similarly, for $\lambda_2 = -\mathbf{i}$, we find

$$\text{null}(A - \lambda_2 E) = \text{span}(v_2) \quad \text{with} \quad v_2 = \begin{pmatrix} \mathbf{i} \\ 1 \end{pmatrix}.$$

- In summary, we can construct an invertible basis transition matrix

$$S = (v_1 \ v_2) = \begin{pmatrix} -\mathbf{i} & \mathbf{i} \\ 1 & 1 \end{pmatrix}, \quad S^{-1} = \frac{1}{2} \begin{pmatrix} \mathbf{i} & 1 \\ -\mathbf{i} & 1 \end{pmatrix}.$$

The matrix $B = S^{-1}AS$ is diagonal, as

$$B = \frac{1}{2} \begin{pmatrix} \mathbf{i} & 1 \\ -\mathbf{i} & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} -\mathbf{i} & \mathbf{i} \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{i} & 0 \\ 0 & -\mathbf{i} \end{pmatrix}.$$

Algebraically Closed Fields

Definition

- A field F is called algebraically closed if every polynomial p of degree $n \geq 1$ can be written in the form

$$p(z) = C(z - z_1)(z - z_2) \dots (z - z_n),$$

for constants $C, z_1, z_2, \dots, z_n \in F$.

Remarks

- The z_i s don't have to be pairwise different.
- If z is a root of p , there may be more than one i with $z = z_i$.
- The number of i s with $z = z_i$ is called the algebraic multiplicity of z .

Fundamental Theorem of Algebra

Theorem

- The field $\mathbb{C}$ is algebraically closed.

Remarks

- The above theorem is known as the fundamental theorem of algebra.
- Elegant proofs are usually discussed in courses on complex analysis.
- Algebraic proofs rely on very basic results from real analysis, but are generally a bit more tricky, see our appendix slides.
- An important consequence of this theorem is that every $(n \times n)$ -matrix A has n eigenvalues in $\mathbb{C}$, counted with multiplicity.

Fundamental Theorem of Algebra

- If $x, y \in \mathbb{R}$ are real numbers, then $z = x + \mathbf{i}y$ satisfies

$$\begin{aligned}\overline{z^2} &= \overline{(x + \mathbf{i}y)^2} = \overline{(x^2 - y^2) + \mathbf{i}(2xy)} \\ &= (x^2 - y^2) + \mathbf{i}(-2xy) = (x - \mathbf{i}y)^2 = \overline{z}^2.\end{aligned}$$

- A recursive application of this formula yields $\overline{z^k} = \overline{z}^k$.
- Consequently, if all the coefficients of a polynomial p are real,

$$p(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_n$$

with $a_0, a_1, a_2, \dots, a_n \in \mathbb{R}$ and $z \in \mathbb{C}$ is a root, $p(z) = 0$, then

$0 = \overline{p(z)} = p(\overline{z})$. Thus, if z is a root of p , then so is $\overline{z}$.

Fundamental Theorem of Algebra

Consequence: Any polynomials p of degree n and real-valued coefficients $a_0, a_1, \dots, a_n \in \mathbb{R}$, can be written in the form

$$p(z) = C \cdot \left(\prod_{i=1}^k (z - z_i)(z - \overline{z_i}) \right) \cdot \left(\prod_{i=1}^{n-2k} (z - x_i) \right).$$

- Here, $z_1, z_2, \dots, z_k \in \mathbb{C} \setminus \mathbb{R}$ denote the complex roots of p .
- Similarly, $x_1, \dots, x_{n-2k} \in \mathbb{R}$ denote the real roots of p , and the real-valued constant $C = a_0 \neq 0$ is the leading coefficient of p .
- The factors $(z - z_i)(z - \overline{z_i}) = (z - r_i)^2 + s_i^2$ can be interpreted as quadratic factors of p , where $r_i = \operatorname{Re}(z_i)$ and $s_i = \operatorname{Im}(z_i)$.

Eigenspaces

Definition

- The eigenspace of $A \in F^{n \times n}$ and $\lambda \in F$ is defined as

$$V(A, \lambda) \stackrel{\text{def}}{=} \text{null}(A - \lambda E) = \{ v \in V \mid Av = \lambda v \}.$$

- Its dimension, $\dim(V(A, \lambda))$, is called the geometric multiplicity of λ .

Remarks

- The above definition can be used for endomorphisms $f : V \rightarrow V$, too.
The eigenspace is then the set of $v \in V$ with $f(v) = \lambda v$.
- We have $V(A, \lambda) = \{0\}$ if and only if λ is not an eigenvalue of A .

Eigenspaces

- From the fundamental theorem of algebra, we know that for $F = \mathbb{C}$ the characteristic polynomial of $A \in \mathbb{C}^{n \times n}$ can be written as

$$\det(A - \lambda E) = (-1)^n (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_k)^{m_k},$$

where $\lambda_1, \lambda_2, \dots, \lambda_k$ are pairwise different eigenvalues of A .

- Here, m_i denotes the algebraic multiplicity of λ_i . We must have

$$m_1 + m_2 + \dots + m_k = n.$$

- We already know that if $k = n$ and $m_1 = m_2 = \dots = m_n = 1$, then the eigenvectors $v_1, v_2, \dots, v_n$ with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$ are linearly independent and, hence, a basis of $V \cong \mathbb{C}^n$.

Characterization of Diagonalizable Matrices

Theorem: Let $V \cong \mathbb{C}^n$ be a finite dimensional complex vector space and $A \in \mathbb{C}^{n \times n}$ a given matrix. The following statements are equivalent.

1. The matrix A is diagonalizable; that is, there exists an invertible basis transition matrix $S \in \mathbb{C}^{n \times n}$ such that $B = S^{-1}AS$ is diagonal.
2. The algebraic multiplicity of all eigenvalues coincides with their geometric multiplicity, $\dim(V(A, \lambda_i)) = m_i$ for $1 \leq i \leq k$, and

$$m_1 + m_2 + \dots + m_k = n.$$

3. The vector space V has a basis that consists of eigenvectors of A .
4. We have $V = V(A, \lambda_1) + V(A, \lambda_2) + \dots + V(A, \lambda_k)$ (direct sum).

Characterization of Diagonalizable Matrices

Proof:

- We already know that Statements 1 and 3 are equivalent. Similarly, the equivalence of Statements 3 and 4 follows directly from our definition of eigenspaces.
- Next, let $v_1, v_2, \dots, v_k$ denote a basis of $V(A, \lambda_i)$ and let $v_1, \dots, v_k, v_{k+1}, \dots, v_n$ a basis of V . The representation matrix of the map $x \rightarrow Ax$ with respect to this basis has the form

$$\begin{pmatrix} \lambda_i E_k & * \\ 0 & B \end{pmatrix},$$

where E_k denotes the $(k \times k)$ -unit matrix. Thus, the characteristic polynomial of A is $(x - \lambda_i)^k$ times the characteristic polynomial of B .

Characterization of Diagonalizable Matrices

Proof (continued):

- Consequently, if the dimension k_i of $V(A, \lambda_i)$ is smaller than or equal to the algebraic multiplicity m_i of λ_i .
- But if we had $k_i < m_i$ for some i , we cannot simultaneously have

$$m_1 + m_2 + \dots + m_n = n \quad \text{and} \quad k_1 + k_2 + \dots + k_n = n.$$

- We additionally recall that $V(A, \lambda_i) \cap V(A, \lambda_j) = \{0\}$ for $\lambda_i \neq \lambda_j$, since eigenvectors with pairwise different eigenvalues are linearly independent. Hence, $V(A, \lambda_1) + \dots + V(A, \lambda_k)$ is a direct sum.
- In summary, it follows that Statements 2 and 4 are equivalent. This completes the proof of the theorem.

Non-Diagonalizable Matrices

- The characterization of diagonalizable matrices of the above theorem is certainly helpful—but it also has several problems.
- First of all, as much as the fundamental theorem of algebra asserts the existence of eigenvalues, finding and analyzing the roots of higher order polynomials in practice is a very different story.
- In Lecture 8, we'll see that there are classes of practical matrices, for example, symmetric matrices, which are much easier to diagonalize.
- But an even more important problem of the above theorem is that its conditions may fail to hold. In fact, there are matrices that are simply not diagonalizable; see the example on the next slide.

Example for a Non-Diagonalizable Matrix

Example

- Let us consider the matrix

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

- The characteristic polynomial of this matrix is given by

$$\det(A - \lambda E) = \lambda^2.$$

- Consequently, $\lambda = 0$ is the only eigenvalue.
- But a diagonalizable matrix whose eigenvalues are all 0 must be 0.
Consequently, the above matrix A is not diagonalizable.
- The root λ has algebraic multiplicity 2 but geometric multiplicity 1.

Contents

- Introduction
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- **Generalized Eigenspaces**
- Jordan Normal Form
- Matrix Powers
- Summary
- Appendix: Fundamental Theorem of Algebra

Invariant Subspaces

Definition

- Let $f : V \rightarrow V$ be an endomorphism. A subspace W of V is called invariant if

$$f(W) \subseteq W \quad \text{with} \quad f(W) \stackrel{\text{def}}{=} \{ f(w) \mid w \in W \}.$$

- The endomorphism $g : W \rightarrow W$, given by $g(a) = f(a)$ for all $a \in W$, is called the restriction of f on W .

Example

- The eigenspace, $V(f, \lambda) = \{a \mid f(a) = \lambda a\}$, is an invariant subspace for any $\lambda \in F$. The restriction of f on $V(f, \lambda)$ is diagonalizable.

Invariant Subspaces

Properties

- Let $f : V \rightarrow V$ be an endomorphism and W be an invariant subspace.
- If we start with a basis $v_1, v_2, \dots, v_m$ of W and augment it, such that $v_1, v_2, \dots, v_n$ is a basis of V , then the representation matrix A of f with respect to this basis has the form

$$A = \begin{pmatrix} B & * \\ 0 & C \end{pmatrix}.$$

- Note that the characteristic polynomials p, q , and r of the matrices A, B , and C satisfy $p = qr$. Consequently, q and r are divisors of p .

Short Interlude: Co-Prime Numbers

- Can we find numbers $a, b \in \mathbb{Z}$ such that $7a + 9b = 1$?
- We use Euclid's method (recursive division with remainder),

$$9 = 1 * 7 + 2, \quad 7 = 3 * 2 + 1 \quad (\text{or } 7 = 4 * 2 - 1),$$

to figure out that 7 and 9 are co-prime and read backwards:

$$1 = 7 - 3 * 2 = 7 - 3 * (9 - 7) = 4 * 7 - 3 * 9.$$

Thus, $a = 4$ and $b = -3$ is a solution (or $a = -5$ and $b = 4$).

- Obviously, we can do the same for any pair of co-prime numbers. A formal proof uses induction over the recursive division with remainder.

Co-Prime Polynomials

- In analogy to integers, we say that a polynomial q is a divisor of a polynomial p if there exists a polynomial r with $p = qr$.
- Moreover, two polynomials p and q are called co-prime if they have—apart from the constant polynomials—no common divisor.
- And again in analogy to integers: if p and q are co-prime polynomials, then there exist polynomials a and b such that

$$ap + bq = 1.$$

- We can construct a and b by running Euclid's method; that is, recursive polynomial division with remainder.

Co-Prime Polynomials

Example

- The polynomials

$$p(x) = 6x^4 + 2x^3 + 5x^2 + 3x + 2 \quad \text{and} \quad q(x) = 2x^2 + 1$$

are co-prime, because Euclid's algorithm yields,

$$\begin{aligned} p(x) &= (3x^2 + x + 1)q(x) + (2x + 1), \\ q(x) &= \left(x - \frac{1}{2}\right)(2x + 1) + \frac{3}{2}. \end{aligned}$$

- Reversely,

$$1 = \underbrace{\left(\frac{1}{3} - \frac{2}{3}x\right)}_{=a(x)} p(x) + \underbrace{\left(2x^3 - \frac{1}{3}x^2 + \frac{1}{3}x + \frac{1}{3}\right)}_{=b(x)} q(x).$$

Polynomials of Endomorphisms

- Let $p(x) = a_0x^0 + a_1x^1 + a_2x^2 + \dots a_mx^m$ be a polynomial.
- We are only interested in the case that $a_0, a_1, \dots, a_m \in \mathbb{C}$.
- In the following, we allow to formally substitute an endomorphism $f : V \rightarrow V$ for the variable x . This means that we write

$$p(f) = a_0f^0 + a_1f^1 + a_2f^2 + \dots a_mf^m.$$

Here, $f_0 = \text{id}_V$ is the identity map on V and $f^k = f \circ f \circ \dots \circ f$ denotes k times the composition of f with itself, $k \geq 1$.

- Similarly, if A is a square matrix, we define $A^0 = E$ and write

$$p(A) = a_0A^0 + a_1A^1 + a_2A^2 + \dots a_mA^m.$$

Cayley-Hamilton Theorem

Theorem

- Let A be a square matrix, and let $p(\lambda) = \det(A - \lambda E)$ be the characteristic polynomial of A . Then, we have $p(A) = 0$.

Proof

- Let us set $C(\lambda) = A - \lambda E$. Next, let $D(\lambda)$ denote the complementary matrix of $C(\lambda)$, such that

$$D(\lambda)C(\lambda) = p(\lambda)E.$$

Recall that this equation is called Cramer's rule (see Lecture 3) for the $(n \times n)$ -matrix $C(\lambda)$.

Cayley-Hamilton Theorem

Proof (continued)

- The coefficients of $D(\lambda)$ are determinants of submatrices of $C(\lambda)$.

Thus, there are matrices $D_0, D_1, \dots, D_{n-1}$ such that

$$D(\lambda) = D_0 + D_1\lambda + D_2\lambda^2 + \dots + D_{n-1}\lambda^{n-1}.$$

- Now, we have on the one hand

$$\begin{aligned} D(\lambda)C(\lambda) &= (D_0 + D_1\lambda + D_2\lambda^2 + \dots + D_{n-1}\lambda^{n-1})(A - \lambda E) \\ &= D_0A + \sum_{i=1}^{n-1} (D_iA - D_{i-1})\lambda^i - D_{n-1}\lambda^n \end{aligned}$$

and, on the other hand, there are coefficients $a_0, a_1, \dots, a_{n-1} \in \mathbb{C}$ with

$$p(\lambda)E = a_0E + a_1E\lambda + a_2E\lambda^2 + \dots + a_{n-1}E\lambda^{n-1} + E\lambda^n.$$

Cayley-Hamilton Theorem

Proof (continued)

- Thus, by comparing coefficients, we find

$$D_0A = a_0E, \quad D_iA - D_{i-1}A = a_iE, \quad \text{and} \quad -D_{n-1}A = E$$

for all $1 \leq i \leq n-1$.

- Now, it remains to verify that

$$\begin{aligned} p(A) &= a_0E + a_1A + a_2A^2 + \dots + a_{n-1}A^{n-1} + A^n \\ &= D_0A + (D_1A - D_0)A + \dots + (D_{n-1}A - D_{n-2})A^{n-1} + A^n \\ &= 0. \end{aligned}$$

This completes the proof.

Subspace Splitting Theorem

Theorem

- Let A be an $(n \times n)$ matrix and p a polynomial with $p(A) = 0$, for example, the characteristic polynomial of A . If $p = qr$ is a product of two co-prime polynomials q and r , then their associated null-spaces

$$V \stackrel{\text{def}}{=} \text{null}(q(A)) \quad \text{and} \quad W \stackrel{\text{def}}{=} \text{null}(r(A))$$

satisfy $V + W = \mathbb{C}^n$ and $V \cap W = \{0\}$. Moreover, the subspaces V and W are invariant for $x \rightarrow Ax$; that is $AV \subseteq V$ and $AW \subseteq W$.

Subspace Splitting Theorem

Proof

- Since q and r are co-prime, there are polynomials a and b such that

$$aq + br = 1.$$

- For any $x \in \mathbb{C}^n$ the vectors $v = b(A)r(A)x$ and $w = a(A)q(A)x$ satisfy $x = v + w$.
- Moreover, we have $v \in V = \mathbf{null}(q(A))$, because

$$q(A)v = q(A)r(A)b(A)x = p(A)b(A)x = 0.$$

And, analogously, we have $w \in W = \mathbf{null}(r(A))$.

- Consequently, we have $V + W = \mathbb{C}^n$.

Subspace Splitting Theorem

Proof

- Additionally, we still have to show that $V \cap W = \{0\}$.
- But for $x \in V \cap W$, we have

$$x = a(A)q(A)x + b(A)r(A)x = 0.$$

- Finally, $v \in V$ implies $q(A)Av = Aq(A)v = 0$. Thus, $AV \subseteq V$.
- Similarly, we have $AW \subseteq W$.
- This concludes our proof.

Subspace Splitting Theorem

Consequences

- We know from the fundamental theorem of algebra that the characteristic polynomial of a square matrix A has the form

$$p(\lambda) = \det(A - \lambda E) = (-1)^n (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_k)^{m_k}$$

for pairwise different roots $\lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{C}$ and algebraic multiplicities $m_1, m_2, \dots, m_k$ that satisfy

$$m_1 + m_2 + \dots + m_k = n.$$

- If $k \geq 2$, the polynomials $p_i(\lambda) = (\lambda - \lambda_i)^{m_i}$ for $1 \leq i \leq k$ are all co-prime. Thus, the above theorem can be applied recursively.

Subspace Splitting Theorem

Consequences

- In summary, the subspace splitting theorem implies that

$$V_{m_1}(A, \lambda_1) + V_{m_2}(A, \lambda_2) + \dots + V_{m_k}(A, \lambda_k) = \mathbb{C}^n$$

is a direct sum, where the notation

$$V_r(A, \lambda) \stackrel{\text{def}}{=} \text{null}((A - \lambda E)^r)$$

is used to denote the generalized eigenspaces of order r .

- Note that for $r = 1$, the definition of generalized eigenspaces coincides with our previous definition of eigenspaces.

Subspace Splitting Theorem

Example

- The matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{with} \quad \det(A - \lambda E) = (\lambda - 1)^2$$

is not diagonalizable, because the eigenvalue $\lambda_1 = 1$ has algebraic multiplicity 2 but geometric multiplicity 1.

- The unit vectors $e_1 = (1 \ 0)^T$ and $e_2 = (0 \ 1)^T$ satisfy

$$(A - E)e_2 = e_1 \quad \text{and} \quad (A - E)e_1 = 0.$$

Thus, we have $V_1(A, 1) = \text{span}(e_1)$, but $V_2(A, 1) = \mathbb{C}^2$.

Contents

- Introduction
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- Generalized Eigenspaces
- **Jordan Normal Form**
- Matrix Powers
- Summary
- Appendix: Fundamental Theorem of Algebra

Diagonalizable and Nilpotent Matrices

Theorem: Let $A \in \mathbb{C}^{n \times n}$ be given. Then there exist unique matrices L and N such that

1. we have $A = L + N$,
2. the matrices L and N commute, $LN = NL$,
3. the matrix L is diagonalizable, and
4. the matrix N is nilpotent, $N^n = 0$.

Remark: If A is diagonalizable, the uniqueness statement implies that we must have $L = A$ and $N = 0$.

Diagonalizable and Nilpotent Matrices

Proof:

- Recall that $\mathbb{C}^n$ is a direct sum of the generalized eigenspaces of A ,

$$V_{m_1}(A, \lambda_1) + V_{m_2}(A, \lambda_2) + \dots + V_{m_k}(A, \lambda_k) = \mathbb{C}^n.$$

- The goal of the first part of our proof is to show that A is the sum of a diagonalizable matrix L and a nilpotent matrix N that have the desired properties.
- The main idea is to define L such that $Lx = \lambda_i x$ for all $x \in V_{m_i}(A, \lambda_i)$ for $1 \leq i \leq k$. We aim to show that L is diagonalizable, $N = A - L$ nilpotent, and $LN = NL$.

Diagonalizable and Nilpotent Matrices

Proof (continued):

- Since $Lx = \lambda_i x$ for all $x \in V_{m_i}(A, \lambda_i)$, it follows that

$$V(L, \lambda_i) = \text{null}(L - \lambda_i E) = V_{m_i}(A, \lambda_i).$$

- Since $\mathbb{C}^n$ is the direct sum of the generalized eigenspaces of A , there exists a basis of $\mathbb{C}^n$ that consists of eigenvectors of L . Thus, L is a diagonalizable matrix.
- Next, we recall that the generalized eigenspaces of A are invariant; that is, $x \in V_{m_i}(A, \lambda_i)$ implies $Ax \in V_{m_i}(A, \lambda_i)$. It follows that

$$x \in V_{m_i}(A, \lambda_i) \implies LAx = \lambda_i(Ax) = A(\lambda_i x) = ALx.$$

Thus, since A and L commute on all subspaces, we have $LA = AL$.

Diagonalizable and Nilpotent Matrices

Proof (continued):

- The equations $LA = AL$ and $N = A - L$ imply that

$$NL = (A - L)L = AL - L^2 = LA - L^2 = L(A - L) = LN.$$

In other words, L and N commute as desired.

- Finally, $x \in V_{m_i}(A, \lambda_i)$ implies

$$Nx = (A - L)x = (A - \lambda_i E)x \implies N^n x = (A - \lambda_i E)^n x = 0,$$

since $n \geq m_i$. By repeating the same argument on all generalized eigenspaces, we find that $N^n = 0$ as desired.

- This concludes the first part of our proof, establishing existence.

Diagonalizable and Nilpotent Matrices

Proof (continued):

- The second part of this proof is about uniqueness.
- Let us assume that $L + N = A$, $LN = NL$, and $N^n = 0$. Then, N and A commute, too, and the binomial formula yields

$$\begin{aligned}(L - \lambda_i E)^{2n} &= (A - \lambda_i E - N)^{2n} \\&= \sum_{j=0}^{2n} \binom{2n}{j} (-N)^j (A - \lambda_i E)^{2n-j} \\&= \sum_{j=0}^{n-1} \binom{2n}{j} (-N)^j (A - \lambda_i E)^{2n-j},\end{aligned}$$

where the last equation follows by using that $N^n = 0$.

Diagonalizable and Nilpotent Matrices

Proof (continued):

- Consequently, any $x \in V_{m_i}(A, \lambda_i) = \text{null}((A - \lambda_i E)^{m_i})$ satisfies

$$(L - \lambda_i E)^{2n} x = \sum_{j=0}^{n-1} \binom{2n}{j} (-N)^j (A - \lambda_i E)^{2n-j} x = 0,$$

since $m_i \geq n$.

- But, since L , if diagonalizable, the above relation implies that $Lx = \lambda_i x$ for all $x \in V_{m_i}(A, \lambda_i)$. As the direct sum of the generalized eigenspaces is equal to $\mathbb{C}^n$, there is only one such matrix L with this property. Hence, L and $N = L - A$ are unique.
- This completes our proof.

Diagonalizable and Nilpotent Matrices

Discussion

- The above theorem shows that we can write any matrix A in the form

$$A = L + N = S^{-1}(D + M)S$$

for a diagonal matrix D , an invertible basis transition matrix S that diagonalizes the matrix L , and $M = SNS^{-1}$.

- Regarding uniqueness statements, however, we need to be careful. Namely, as much as L and N are unique, the choice of S is in general not unique. For example, if $A = E + N$ is the sum of the unit matrix E and a nilpotent matrix N , we have $A = S^{-1}(E + M)S$ with $M = SNS^{-1}$, where S can be any invertible matrix.

Diagonalizable and Nilpotent Matrices

Discussion (continued)

- Now, there arises the following question: can we choose S such that

$$A = S^{-1}(D + M)S$$

such that D is diagonal and M is a nilpotent matrix that has a suitable, “simple” structure?

- In order to give a systematic answer to this question, we introduce generalized eigenvectors. Specifically, a vector $v \in \mathbb{C}^n$ is called a generalized eigenvector of order r if

$$(A - \lambda_i E)^r v = 0 \quad \text{and} \quad (A - \lambda_i E)^{r-1} v \neq 0.$$

Generalized Eigenvectors

- If we start with a generalized eigenvector v_r with maximum order r , the backward recursion

$$v_{k-1} = (A - \lambda_i E)v_k \quad 2 \leq k \leq r$$

generates r linearly independent vectors $v_1, v_2, \dots, v_r$.

- The first vector v_1 is then an eigenvector, $Av_1 = \lambda_i v_1$.
- Now, let $f(x) = Ax$ be the endomorphism associated to A and let g denote its restriction onto the subspace $\text{span}(v_1, v_2, \dots, v_r)$.

Jordan Blocks

- The function g from the previous slide satisfies

$$g(v_1) = \lambda_i v_1 \quad \text{and} \quad g(v_k) = v_{k-1} + \lambda_i v_k \quad 2 \leq k \leq r.$$

- Thus, the representation matrix of g with respect to the basis $(v_1, v_2, \dots, v_r)$ is a Jordan block matrix of the form

$$\begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_i & 1 \\ 0 & \dots & 0 & 0 & \lambda_i \end{pmatrix}$$

Jordan Normal Form

- In words, a Jordan block is a matrix whose diagonal elements are all equal to a constant λ_i , the entries above the diagonal are all 1s, and all other entries are equal to 0.
- The (1×1) matrix (λ_i) is also a Jordan block.

Theorem

- Let $A \in \mathbb{C}^{n \times n}$ be a given matrix. Then there exists an invertible matrix S such that $J = S^{-1}AS$ is a Jordan normal form; that is,

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix}$$

is a block-diagonal matrix, with Jordan blocks on the diagonal.

Jordan Normal Form

Proof

- We have already shown that the representation of the restriction of $f(x) = Ax$ to the subspace $\text{span}(v_1, v_2, \dots, v_r)$ is a Jordan block with respect to the basis vectors $v_1, v_2, \dots, v_r$.
- We can then simply pick the next generalized eigenvector in $\mathbb{C}^n$ with maximum order among all generalized eigenvectors that are not in $\text{span}(v_1, v_2, \dots, v_r)$, run another backward recursion and get a second Jordan block.
- Due to the above subspace splitting theorem, this procedure can be repeated until a complete basis of $\mathbb{C}^n$ is found. The representation matrix of f with respect to this basis is then, by construction, in Jordan normal form, completing the proof.

Jordan Normal Form

Example

- Let us consider the matrix

$$A = \begin{pmatrix} -2 & 1 & 0 & 0 & 0 \\ -4 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 6 & -3 & 1 & -1 & 2 \end{pmatrix}.$$

- The first towards writing this matrix in Jordan normal form is work out its characteristic polynomial $p(\lambda) = \det(A - \lambda E)$.

Jordan Normal Form

Example (continued)

- We find

$$\begin{aligned} p(\lambda) &= \det \begin{pmatrix} -2-\lambda & 1 \\ -4 & 2-\lambda \end{pmatrix} \cdot \det \begin{pmatrix} -\lambda & 1 & 0 \\ 0 & 1-\lambda & 1 \\ 1 & -1 & 2-\lambda \end{pmatrix} \\ &= \lambda^2 \cdot \left(-\lambda \cdot \det \begin{pmatrix} 1-\lambda & 1 \\ -1 & 2-\lambda \end{pmatrix} + \det \begin{pmatrix} 1 & 0 \\ 1-\lambda & 1 \end{pmatrix} \right) \\ &= \lambda^2 \cdot (-\lambda(1-\lambda)(2-\lambda) - \lambda + 1) \\ &= \lambda^2 \cdot (-\lambda^2 + 3\lambda^2 - 3\lambda + 1) = -\lambda^2(\lambda - 1)^3. \end{aligned}$$

- Thus, the eigenvalues of A are $\lambda_1 = 0$ and $\lambda_2 = 1$.

Jordan Normal Form

Example (continued)

- Let us first work out the generalized eigenspace of the eigenvalue $\lambda_1 = 0$, which has algebraic multiplicity 2. We have

$$(A - \lambda_1 E)^2 = A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 6 & -3 & 1 & 0 & 3 \\ 12 & -6 & 2 & -2 & 3 \end{pmatrix}.$$

- The vectors $v_1 = (1, 2, 0, 0, 0)^T$ and $v_2 = (0, 1, 3, 0, 0)^T$ span the null space of A^2 . They are generalized eigenvectors. They satisfy

$$Av_2 = v_1 \quad \text{and} \quad Av_1 = 0.$$

Jordan Normal Form

Example (continued)

- Similarly, we work out the generalized eigenspace of $\lambda_2 = 1$,

$$(A - E)^3 = \begin{pmatrix} -7 & 3 & 0 & 0 & 0 \\ -12 & 5 & 0 & 0 & 0 \\ 6 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

- The vector $v_5 = (0, 0, 0, 0, 1)^T$ is in the null-space of $(A - E)^3$ and, hence, a generalized eigenvector. It has order 3, because the vectors $v_4 = (A - E)v_5$ and $v_3 = (A - E)v_4$ are non-zero,

$$v_4 = (0, 0, 0, 1, 1)^T \quad \text{and} \quad v_3 = (0, 0, 1, 1, 0)^T.$$

Jordan Normal Form

Example (continued)

- In summary, our basis transition matrix is given by

$$S = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

- It can be used to write A in Jordan normal form,

$$S^{-1}AS = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}, \quad J_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Contents

- Introduction
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- Generalized Eigenspaces
- Jordan Normal Form
- **Matrix Powers**
- Summary
- Appendix: Fundamental Theorem of Algebra

Matrix Powers

- Jordan normal forms can be used to compute the k -th power of a given matrix $A \in \mathbb{C}^{n \times n}$,

$$A^k = \underbrace{A \cdot A \cdot \dots \cdot A}_{k \text{ times}}$$

- Let $A = SJS^{-1}$ be a Jordan normal of A , with $J = D + N$ being the sum of a diagonal matrix D and a nilpotent matrix N , such that $DN = ND$. Then, we have

$$\begin{aligned} A^k &= SJS^{-1}SJS^{-1} \dots SJS^{-1} = SJ^kS^{-1} = S(D+N)^kS^{-1} \\ &= S \left(\sum_{i=0}^k \binom{k}{i} N^i D^{k-i} \right) S^{-1}. \end{aligned}$$

Matrix Powers

- The formula

$$A^k = S \left(\sum_{i=0}^k \binom{k}{i} N^i D^{k-i} \right) S^{-1}$$

is useful, because the powers of the diagonal matrix D are

$$D^k = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}^k = \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix}.$$

Similarly, since N is nilpotent, we have $N^k = 0$ for all $k \geq n$. Thus, the number of non-zero summands is at most n .

Matrix Powers

Example 1

- Let us compute an explicit expression for

$$\begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}^k$$

- We first need to work out the characteristic polynomial,

$$p(\lambda) = \det \begin{pmatrix} 1 - \lambda & 1 \\ -1 & 3 - \lambda \end{pmatrix} = 4 - 4\lambda + \lambda^2 = (\lambda - 2)^2.$$

- Thus, our only eigenvalue is given by $\lambda_1 = 2$.

Matrix Powers

Example 1 (continued)

- Next, we work out the generalized eigenspace. Since

$$A - 2E = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad (A - 2E)^2 = 0,$$

the geometric multiplicity of $\lambda_2 = 2$ is 1. We can start with the generalized eigenvector $v_2 = e_2$ and compute the eigenvector

$$v_1 = (A - 2E)v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- Thus, our basis transition matrix is given by

$$S = \begin{pmatrix} v_1 & v_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Matrix Powers

Example 1 (continued)

- Next, we can use the Jordan normal form to compute

$$\begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}^k = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^k \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{-1}$$

where the powers of the Jordan block are

$$\left(\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right)^k = \begin{pmatrix} 2^k & 0 \\ 0 & 2^k \end{pmatrix} + k \begin{pmatrix} 0 & 2^{k-1} \\ 0 & 0 \end{pmatrix}.$$

- Consequently,

$$\begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}^k = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^k & k2^{k-1} \\ 0 & 2^k \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{-1}.$$

Matrix Powers

Example 1 (continued)

- Our result can be further simplified finding

$$\begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}^k = \begin{pmatrix} 2^k - k2^{k-1} & k2^{k-1} \\ -k2^{k-1} & 2^k + k2^{k-1} \end{pmatrix}.$$

- This is the explicit expression for the coefficients of the matrix power that we were looking for.
- In this particular example, our matrix is invertible, which implies that the above formula is correct for all integers k , as long as we define $A^k = (A^{-1})^{-k}$ whenever k is negative (and $A^0 = E$).

Matrix Powers

Example 2

- Let us consider the Fibonacci sequence,

$$1, 1, 2, 3, 5, 8, 13, \dots$$

It starts with $F_1 = 1$ and $F_2 = 1$ and evaluates $F_{k+2} = F_k + F_{k+1}$ recursively, for $k \geq 1$.

- Alternatively, we could set $x_k = (F_{k+1}, F_{k+2})^T$ and write our recursion in matrix-vector form

$$x_{k+1} = Ax_k \quad \text{with} \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Matrix Powers

Example 2 (continued)

- The solution to the above recursion is given by

$$x_k = A^k x_0 \quad \text{with} \quad x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- The characteristic polynomial of A is given by

$$\det(A - \lambda E) = \det \begin{pmatrix} -\lambda & 1 \\ 1 & 1 - \lambda \end{pmatrix} = \lambda^2 - \lambda - 1.$$

Thus, the eigenvalues are $\lambda_1 = \frac{1-\sqrt{5}}{2}$ and $\lambda_2 = \frac{1+\sqrt{5}}{2}$.

Matrix Powers

Example 2 (continued)

- Next, we compute associated eigenvectors

$$v_1 = \begin{pmatrix} 1 + \sqrt{5} \\ -2 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 2 \\ 1 + \sqrt{5} \end{pmatrix}$$

- Consequently, a diagonalization of A is given by

$$A = \begin{pmatrix} 1 + \sqrt{5} & 2 \\ -2 & 1 + \sqrt{5} \end{pmatrix} \begin{pmatrix} \frac{1 - \sqrt{5}}{2} & 0 \\ 0 & \frac{1 + \sqrt{5}}{2} \end{pmatrix} \begin{pmatrix} 1 + \sqrt{5} & 2 \\ -2 & 1 + \sqrt{5} \end{pmatrix}^{-1}.$$

Matrix Powers

Example 2 (continued)

- It remains to work out the explicit expression

$$x_k = A^k x_0 = S \begin{pmatrix} \left(\frac{1-\sqrt{5}}{2}\right)^k & 0 \\ 0 & \left(\frac{1+\sqrt{5}}{2}\right)^k \end{pmatrix} S^{-1} x_0.$$

- In order to further this expression, we compute

$$\begin{pmatrix} 1 + \sqrt{5} & 2 \\ -2 & 1 + \sqrt{5} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{10 + 2\sqrt{5}} \begin{pmatrix} \sqrt{5} - 1 \\ \sqrt{5} + 3 \end{pmatrix}.$$

Matrix Powers

Example 2 (continued)

- It follows that

$$\begin{aligned}x_k &= \frac{1}{10 + 2\sqrt{5}} \begin{pmatrix} 4 \left(\frac{1-\sqrt{5}}{2} \right)^k + (6 + 2\sqrt{5}) \left(\frac{1+\sqrt{5}}{2} \right)^k \\ (2 - 2\sqrt{5}) \left(\frac{1-\sqrt{5}}{2} \right)^k + (8 + 4\sqrt{5}) \left(\frac{1+\sqrt{5}}{2} \right)^k \end{pmatrix} \\&= \frac{1}{\sqrt{5}} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2} \right)^{k+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \\ \left(\frac{1+\sqrt{5}}{2} \right)^{k+2} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+2} \end{pmatrix}.\end{aligned}$$

- The corresponding expression for F_k is called Binet's formula,

$$F_k = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^k - \left(\frac{1-\sqrt{5}}{2} \right)^k \right).$$

Contents

- Introduction
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- Generalized Eigenspaces
- Jordan Normal Form
- Matrix Powers
- **Summary**
- Appendix: Fundamental Theorem of Algebra

Summary

In this lecture we have learned how to diagonalize a square matrix A by

1. working out the characteristic polynomial $p(\lambda) = \det(A - \lambda E)$,
2. locating the eigenvalues by determining the roots of p ,
3. determining their algebraic and geometric multiplicities, and
4. computing bases of their eigenspaces.

If the direct sum of the eigenspaces is equal to $\mathbb{C}^n$, the matrix A is diagonalizable. Otherwise, the direct sum of the generalized eigenspaces is equal to $\mathbb{C}^n$ and we can write A in Jordan normal form,

$$A = S^{-1}JS \quad \text{and} \quad J = D + N \quad \text{with} \quad \begin{cases} D : \text{diagonal,} \\ N : \text{nilpotent,} \\ DN = ND. \end{cases}$$

Contents

- Introduction
- Eigenvalues and Eigenvectors
- Characteristic Polynomials
- Generalized Eigenspaces
- Jordan Normal Form
- Matrix Powers
- Summary
- **Appendix: Fundamental Theorem of Algebra**

Fundamental Theorem of Algebra

- As mentioned above, the fundamental theorem of algebra is usually proven in courses on complex analysis.
- Nevertheless, in principle, it is possible to understand this theorem by applying the linear algebra methods from our course.
- Additionally, we will use that odd polynomials with real coefficients have at least one real root. For the proof of this statement we will need to use a result from analysis, but arguably a very basic one, namely, the intermediate value theorem.
- Note that the following slides are optional. They are not formally part of this course and, of course, we won't ask about this in exams.

Fundamental Theorem of Algebra

- Our goal is to show that any polynomial p with degree $n \geq 1$ and coefficients in $\mathbb{C}$ can be written in the form

$$p(x) = (x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n)$$

for some roots $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{C}$. Here, we assume without loss of generality that the leading coefficient is 1.

- For $n = 1$, the statement holds.
- Additionally, we recall our result from Lecture 4, where we have shown this statement for $n = 2$. This follows directly from the explicit solution formula for quadratic equations and the fact that every complex number has a square-root in $\mathbb{C}$ (see Lecture 4).

Fundamental Theorem of Algebra

- As mentioned above, we need an argument from Analysis to show that polynomials with odd degree n ,

$$p(x) = x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n,$$

and real-valued coefficients $a_1, a_2, \dots, a_n \in \mathbb{R}$, have at least one real root. Namely, if n is odd, we can find a sufficiently large $r \in \mathbb{R}$ such that $p(-r) < 0$ and $p(r) > 0$. For example, $r = 1 + |a_1| + \dots + |a_n|$. But then the intermediate value theorem implies that there is an $x \in [-r, r]$ with $p(x) = 0$, because polynomials are continuous.

Fundamental Theorem of Algebra

Lemma 1

- Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$ be real matrices. If $AB = BA$ and if n is odd, then A and B have a common eigenvector.

Proof

- The degree of the characteristic polynomial of A is odd. Thus, we know that A has at least one real eigenvalue $\lambda \in \mathbb{R}$. Let

$$V = \text{null}(A - \lambda E)$$

be its eigenspace. Note that $Ax = \lambda x$ implies $BAx = ABx = Bx$, since $AB = BA$. Thus, we have $BV \subseteq V$.

Fundamental Theorem of Algebra

Proof (continued)

- If $x \in V$, then we have $Ax = \lambda x$, which implies

$$A(Bx) = (AB)x = (BA)x = B(Ax) = B(\lambda x) = \lambda(Bx),$$

since $AB = BA$. Thus, B and V satisfy $BV \subseteq V$.

- Similarly, the image set $W = \{(A - \lambda E)x \mid x \in \mathbb{R}^n\}$ satisfies:

$$BW = \{(BA - \lambda B)x \mid x \in \mathbb{R}^n\} = \{(A - \lambda E)Bx \mid x \in \mathbb{R}^n\} \subseteq W.$$

- Since n is odd, B has a real eigenvector v . If $V = \mathbb{R}^n$, v is also an eigenvector of A . Thus, we may assume $V \neq \mathbb{R}^n$.

Fundamental Theorem of Algebra

Proof (continued)

- Since we may assume $V \neq \mathbb{R}^n$, we have $\dim(V) < n$, $\dim(W) < n$, and $\dim(V) + \dim(W) = n$.
- Since n is odd, either $\dim(V)$ or $\dim(W)$ is odd.
- But then we may assume—by induction—that A and B already have a common eigenvector in W or in V . (To be precise, we should add that the induction start, $n = 1$, is trivial.)
- As $BV \subseteq V$ and $BW \subseteq W$, this completes our proof.

Fundamental Theorem of Algebra

Lemma 2

- Let $A \in \mathbb{C}^{n \times n}$ be a complex matrices. If n is odd, then A has an eigenvector in $\mathbb{C}$.

Proof

- Let $V = \{H \in \mathbb{C}^{n \times n} \mid H = \overline{H}^T\}$ denote the set of Hermitian matrices $H = H_1 + \mathbf{i}H_2$ that satisfy

$$H = H_1 + \mathbf{i}H_2 = H_1^T - \mathbf{i}H_2^T = \overline{H}^T.$$

- Here, $H_1 = H_1^T$ and $H_2 = -H_2^T$ are real matrices. Thus, we can interpret V as a vector space over $\mathbb{R}$. We then have $\dim_{\mathbb{R}}(V) = n^2$.

Fundamental Theorem of Algebra

Proof (continued)

- Let us introduce two endomorphisms $f : V \rightarrow V$ and $g : V \rightarrow V$:

$$f(H) = \frac{AH + H\bar{A}^\top}{2} \quad \text{and} \quad g(H) = \frac{AH - H\bar{A}^\top}{2i}.$$

- It is not difficult to check that $f \circ g = g \circ f$. Namely, we have

$$f(g(H)) = \frac{A^2H - H(\bar{A}^\top)^2}{4i} = g(f(H)).$$

- Since $\dim_{\mathbb{R}}(V) = n^2$ is odd, we can use Lemma 1 to conclude that f and g have a common eigenvector, say $f(X) = \lambda X$ and $g(X) = \mu X$,

$$f(X) + \mathbf{i}g(X) = AX = (\lambda + \mathbf{i}\mu)X.$$

This completes the proof, since any non-zero column of X is an eigenvector of A , with eigenvalue $\lambda + \mathbf{i}\mu$.

Fundamental Theorem of Algebra

Lemma 3

- Every matrix $A \in \mathbb{C}^{n \times n}$ has an eigenvector in $\mathbb{C}$.

Proof

- The statement holds for $n = m$, where $m \geq 1$ is odd (Lemma 2).
- We may assume that the statement also holds if $n = 2^k m$, where $m \geq 1$ is odd and $k \geq 0$. This is our induction assumption, $k \in \mathbb{N}$.
- Let $V = \{S \in \mathbb{C}^{n \times n} \mid S = -S^\top\}$ denote the set of skew-symmetric matrices with complex entries. If $n = 2^{k+1}m$, the dimension of V is

$$\dim_{\mathbb{C}}(V) = \frac{n(n-1)}{2} = 2^k(2^{k+1} - 1).$$

Here, the key point is that the number $2^{k+1} - 1$ is odd.

Fundamental Theorem of Algebra

Proof (continued)

- Let the endomorphism $f : V \rightarrow V$ and $g : V \rightarrow V$ be given by

$$f(S) = \frac{AS - SA^\top}{2} \quad \text{and} \quad g(S) = ASA^\top.$$

- We have $f \circ g = g \circ f$ as

$$f(g(S)) = \frac{A^2SA^\top - AS(A^\top)^2}{2} = g(f(S)).$$

- Since $\dim_{\mathbb{C}}(V) = 2^k(2^{k+1} - 1)$, we can use our induction assumption to conclude that f and g have a common eigenvector X , such that $f(X) = \lambda X$ and $g(X) = \mu X$. This yields

$$\mu X = g(X) = \mu AXA^\top = A(AX - f(X)) = (A^2 - \lambda A)X.$$

Fundamental Theorem of Algebra

Proof (continued)

- We write the above equation in the form

$$0 = (A^2 - \lambda A - \mu E)X = (A - \alpha E)(A - \beta E)X.$$

Here, α and β are the complex roots of the quadratic equation

$$x^2 - \lambda x - \mu = 0.$$

- Now, there are only two possibilities: if $(A - \beta E)X = 0$, every non-zero column of X is an eigenvector of A with eigenvalue β . Or, if $(A - \beta E)X \neq 0$, every of its non-zero columns is an eigenvector of A with eigenvalue α . In any case, A has an eigenvector, completing our induction step. And this completes our proof.

Fundamental Theorem of Algebra

- Since every matrix $A \in \mathbb{C}^{n \times n}$ has an eigenvector, the matrix

$$A = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_n \\ 1 & 0 & \dots & 0 & -a_{n-1} \\ 0 & 1 & \dots & 0 & -a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_1 \end{pmatrix},$$

with $a_1, a_2, \dots, a_n \in \mathbb{C}$ also has an eigenvector. Thus, since the (negative) characteristic polynomial of this matrix is given by

$$p(\lambda) = \det(\lambda E - A) = \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \dots + a_n,$$

it follows that every polynomial of degree n has at least one root in $\mathbb{C}$.

Fundamental Theorem of Algebra

- As discussed above, every polynomial of the form

$$p(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_n$$

has at least one root $\alpha_1 \in \mathbb{C}$.

- Thus, we can use binomial formula to write the polynomial

$q(x) = p(x + \alpha_1)$ in the form

$$\begin{aligned} q(x) &= (x + \alpha_1)^n + a_1(x + \alpha_1)^{n-1} + a_2(x + \alpha_1)^{n-2} + \dots + a_n \\ &= x^n + b_1x^{n-1} + b_2x^{n-2} + \dots + b_{n-1}x + b_n \end{aligned}$$

for some coefficients $b_1, b_2, \dots, b_n \in \mathbb{C}$. Since $q(0) = 0$, we must have $b_n = 0$.

Fundamental Theorem of Algebra

- Now we have

$$\begin{aligned}p(x) &= x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_n \\&= q(x - \alpha_1) \\&= (x - \alpha_1) \left((x - \alpha_1)^{n-1} + b_1(x - \alpha_1)^{n-2} + \dots + b_{n-1} \right) \\&= (x - \alpha_1)r(x)\end{aligned}$$

for some polynomial r with degree $n - 1$. By induction on n , we may assume that r is already a product of $n - 1$ linear factors. Thus,

$$p(x) = (x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n) \quad \text{with } \alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{C}.$$

This completes our proof of the fundamental theorem of algebra.

Chapter 8

Quadratic Forms

Quadratic Forms

- Introduction
- Quadratic Forms
- Spectral Theorems
- Summary

Contents

- Introduction
- Quadratic Forms
- Spectral Theorems
- Summary

Objectives

In this lecture we will learn about the definition and basic properties of bilinear forms. Topics include:

- symmetric bilinear forms and their quadratic forms,
- Gram matrix and change of basis,
- spectral theorems for Hermitian and symmetric matrices,
- positive definite matrices, and
- Sylvester's index theorem

Contents

- Introduction
- Quadratic Forms
- Spectral Theorems
- Summary

Symmetric Bilinear Forms

Definition

- Let V be a vector space over the field F . A function $B : V \times V \rightarrow F$ is called a symmetric bilinear form if

$$B(x + y, z) = B(x, z) + B(y, z),$$

$$B(\alpha x, y) = \alpha B(x, y),$$

$$\text{and} \quad B(x, y) = B(y, x)$$

for all $x, y, z \in V$ and all $\alpha \in F$.

Remark

- As through the whole course, we assume that $F = \mathbb{R}$ or $F = \mathbb{C}$.

Symmetric Bilinear Forms

- We recall from Lecture 6 that Euclidean inner products, $B(x, y) = \langle x, y \rangle$, are bilinear forms.
- In contrast to inner products, however, bilinear forms don't have to be positive definite.
- In analogy to the parallelogram equation for inner products, we have

$$2B(x, y) = B(x + y, x + y) - B(x, x) - B(y, y).$$

- Thus, if we know the function $q(x) = \frac{1}{2}B(x, x)$, we can recover B as

$$B(x, y) = q(x + y) - q(x) - q(y).$$

Quadratic Forms

Definition

- A function $q : V \rightarrow \mathbb{R}$ is called a quadratic form if

$$B(x, y) = q(x + y) - q(x) - q(y)$$

is a symmetric bilinear form.

Remark

- Working with bilinear forms or quadratic forms is essentially the same, as the above relations establish a bijective correspondence between them. It depends on the context which one is more convenient to work with.

Quadratic Forms

Example 1

- One of the most basic quadratic forms in $V = \mathbb{F}^n$ is given by

$$q(x) = \frac{1}{2} (x_1^2 + x_2^2 + \dots + x_n^2).$$

Its bilinear form is the standard Euclidean scalar product,

$$B(x, y) = x_1y_1 + x_2y_2 + \dots + x_ny_n.$$

Quadratic Forms

Example 2

- Another important quadratic form is given by

$$q(x) = \frac{1}{2} (x_1^2 + x_2^2 + \dots + x_r^2 - x_{r+1}^2 - \dots - x_n^2)$$

for $1 \leq r < n$. The corresponding bilinear form

$$B(x, y) = x_1 y_1 + \dots + x_r y_r - x_{r+1} y_{r+1} - \dots - x_n y_n$$

is not positive definite and, consequently, not an inner product.

Quadratic Forms

Example 3

- In general, there may be mixed terms. For example, the function

$$q(x) = \frac{1}{2} (x_1^2 + 2x_1x_2 + x_2^2 + 4x_1x_3 + 6x_3^2)$$

is a quadratic form. The corresponding bilinear form is given by

$$B(x, y) = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2 + 2x_1y_3 + 2x_3y_1 + 6x_3y_3.$$

- Without further analysis, it might be difficult to tell whether bilinear forms are positive definite. In this example, we can perhaps still see that the vector $x = [1, -1, 0]^T$ satisfies $B(x, x) = 0$. Thus, B is not positive definite. But, in general, we'll have to be more systematic.

Gram Matrix

- Let $V \cong F^n$ be a finite dimensional vector space and $B : V \times V \rightarrow \mathbb{R}$ a bilinear form.
- If $v_1, v_2, \dots, v_n$ is a basis of V , we can construct the matrix

$$A = \begin{pmatrix} B(v_1, v_1) & B(v_1, v_2) & \dots & B(v_1, v_n) \\ B(v_2, v_1) & B(v_2, v_2) & \dots & B(v_2, v_n) \\ \dots & \dots & \ddots & \dots \\ B(v_n, v_1) & B(v_n, v_2) & \dots & B(v_n, v_n) \end{pmatrix}$$

- The matrix A is called the Gram matrix of the bilinear form B .
- If B is symmetric, then A is symmetric, $A = A^T$.

Gram Matrix

- The Gram matrix A , with coefficients $a_{ij} = A_{ji}$ can be used to express B as

$$\begin{aligned} B\left(\sum_i x_i v_i, \sum_j y_j v_j\right) &= \sum_{i,j} x_i B(v_i, v_j) y_j \\ &= \sum_{i,j} x_i a_{ij} y_j = x^\top A y. \end{aligned}$$

Here and on the next slide, all sums run from 1 to $n = \dim(V)$.

- If $V = F^n$, we can use the unit vectors, $v_i = e_i$, as basis. In this case, we can write the above equation in the simpler form

$$B(x, y) = x^\top A y \quad \text{and} \quad q(x) = x^\top A x.$$

Change of Basis

- Let $v_1, v_2, \dots, v_n$ and $w_1, w_2, \dots, w_n$ be two different bases of V .
- Next, let S be the associated basis transition matrix, such that

$$w_i = \sum_j s_{ji} v_j.$$

- Consequently, with $a_{ij} = B(v_i, v_j)$, we work out:

$$\begin{aligned} B\left(\sum_i x_i w_i, \sum_j y_j w_j\right) &= \sum_{i,j} B(w_i, w_j) x_i y_j \\ &= \sum_{i,j} B\left(\sum_k s_{ki} v_k, \sum_l s_{lj} v_l\right) x_i y_j \\ &= \sum_{i,j} \left(\sum_{k,l} s_{ki} a_{kl} s_{lj}\right) x_i y_j \\ &= x^\top (S^\top A S) y. \end{aligned}$$

Change of Basis

Summary:

- If A is the Gram matrix of B with respect to the first basis, the Gram matrix of B with respect to the second basis is given by

$$S^T A S,$$

where S denotes the basis transition matrix.

Remark

- The change of basis formula for the Gram matrix of bilinear forms $A \rightarrow S^T A S$ should not be mixed up with the change of basis formula for the representation matrix of linear transitions, $A \rightarrow S^{-1} A S$.

Bilinear and Quadratic Forms

Summary

- If we use the unit vectors as basis of $V = F^n$, there exists for every quadratic form $q : V \rightarrow \mathbb{R}$ a unique symmetric matrix A such that

$$q(x) = \frac{1}{2}x^{\top}Ax.$$

- Namely, A is the Gram matrix of its associated bilinear form

$$B(x, y) = q(x + y) - q(x) - q(y) = x^{\top}Ay.$$

- The goal of the next section is to discuss how we can find an invertible basis transition matrix S such that the new Gram matrix $S^{\top}AS$ has a simple structure, ideally, we want $S^{\top}AS$ to be diagonal.

Contents

- Introduction
- Quadratic Forms
- Spectral Theorems
- Summary

Self-Adjoint Linear Operators

- As we had discussed in Lecture 7, not all endomorphisms/square matrices are diagonalizable.
- In Lecture 7, we have solved this problem by introducing Jordan normal forms, which—at least over the field $\mathbb{C}$ —always exist.
- A similar general analysis over $\mathbb{R}$ is, however, rather hopeless, because many real-valued matrices have no eigenvalues in $\mathbb{R}$.
- Fortunately, the situation turns out to be much easier if we study symmetric matrices (or, alternatively, Hermitian matrices).
- Nevertheless, in order to discuss the real-valued case, it is helpful to work in a unitary space for a while, introducing so-called self-adjoint linear transformations (also called self-adjoint linear operators).

Self-Adjoint Linear Operators

Definition

- Let $(V, \langle \cdot, \cdot \rangle)$ be an Euclidean space ($F = \mathbb{R}$) or a unitary space ($F = \mathbb{C}$). A linear transformation $f : V \rightarrow V$ is called self-adjoint if

$$\langle f(x), y \rangle = \langle x, f(y) \rangle \quad \text{for all } x, y \in V.$$

Remark

- If $V = \mathbb{C}^n$ and $\langle a, b \rangle = a^T \bar{b}$ and if $A \in \mathbb{C}^n$ is a representation of $f(x) = Ax$, with respect to some basis of V , we have

$$\langle f(x), y \rangle = x^T A^T y \quad \text{and} \quad \langle x, f(y) \rangle = x^T \bar{A} y.$$

Thus, $f(x) = Ax$ is self-adjoint if and only if A is Hermitian, $A^T = \bar{A}$.

Spectral Theorem

Theorem

- Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional unitary space over $\mathbb{C}$ and $f : V \rightarrow V$ a linear transformation. Then there exists an orthonormal basis $v_1, v_2, \dots, v_n \in \mathbb{C}$ consisting of eigenvectors, such that

$$f(v_i) = \lambda_i v_i \quad 1 \leq i \leq n.$$

The eigenvalues $\lambda_i \in \mathbb{R}$ are all real.

Remark

- By substituting the standard scalar product, the above theorem implies: there exists for every Hermitian matrix H a unitary matrix U such that $U^{-1}HU = \overline{U}^T HU$ is a real-valued diagonal matrix.

Spectral Theorem

Proof

- The huge advantage of working with unitary vector spaces is that we know that there exists an eigenvalue $\lambda \in \mathbb{C}$ and an eigenvector $v \in V$,

$$f(v) = \lambda v.$$

- Since f is self-adjoint, we have

$$\lambda \langle v, v \rangle = \langle f(v), v \rangle = \langle v, f(v) \rangle = \bar{\lambda} \langle v, v \rangle.$$

Thus, we have $\lambda = \bar{\lambda}$ and, consequently, λ is real.

Spectral Theorem

Proof

- Let W be the orthogonal complement of $\text{span}(v) = \mathbb{C}v$ in V .
- Then any $w \in W$ satisfies

$$\langle f(w), v \rangle = \langle w, f(v) \rangle = \lambda \langle w, v \rangle = 0.$$

Thus, we have $f(W) \subseteq W$. In other words, W is an invariant subspace of V . The restriction of f on W is again self-adjoint. Since $\dim(W) = \dim(V) - 1$, we can set $V \leftarrow W$ and proceed recursively to construct an orthonormal basis of V consisting of eigenvectors.

This completes our proof.

Spectral Theorem

- Since the eigenvalues of Hermitian matrices are real, the eigenvalues of any real-valued symmetric matrix $A \in \mathbb{R}^{n \times n}$ are also real.
- Due to the above spectral theorem, we know that there exists a unitary matrix $U \in \mathbb{C}^{n \times n}$ with $AU = DU$.
- Now, we can write $U = U_1 + \mathbf{i}U_2$, where $U_1 \in \mathbb{R}^{n \times n}$ is the real part and $U_2 \in \mathbb{R}^{n \times n}$ the imaginary part of U . Since U_1 and U_2 cannot both be zero, there is at least one column vector, either of U_1 or of U_2 , which is not zero. Thus, A has at least one real eigenvector.
- Consequently, we can repeat the induction argument of the above proof, but this time over the field $F = \mathbb{R}$.

Spectral Theorem

Theorem

- Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix, $A = A^\top$. Then there exist an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ such that

$$D = Q^\top A Q = Q^{-1} A Q$$

is a diagonal matrix. Its diagonal entries are the eigenvalues of A .

Remark

- The above theorem implies that there exists for every bilinear form $B : V \times V \rightarrow \mathbb{R}$ over a finite dimensional vector space V an orthonormal basis such that the Gram matrix of B is diagonal with respect to this basis.

Positive Definite Matrices

- Let A be a symmetric matrix, $A = A^\top$, and let D be the diagonal matrix consisting of its eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$, such that $D = Q^\top A Q$ for an orthogonal matrix Q .
- Let us set $x = Qy$ or, equivalently, $y = Q^\top x$. The quadratic form $q(x) = \frac{1}{2}x^\top A x$ satisfies

$$\begin{aligned} q(y) &= \frac{1}{2}y^\top A y = (Qx)^\top A (Qx) \\ &= x^\top Q^\top A Q x = x^\top D x \\ &= \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2. \end{aligned}$$

- Clearly, q is positive definite if and only if the eigenvalues $\lambda_1, \dots, \lambda_n$ of the symmetric matrix A are strictly positive.

Positive Definite Matrices

Definition

- A matrix $P \in \mathbb{R}^{n \times n}$ is called positive definite if its associated quadratic form $q(x) = \frac{1}{2}x^\top Px$ is positive definite; that is

$$q(x) = \frac{1}{2}x^\top Px > 0 \quad \text{for all } x \neq 0.$$

Remark

- The bilinear form of the quadratic form q is given by

$$B(x, y) = q(x + y) - q(x) - q(y) = x^\top Ax \quad \text{with} \quad A = \frac{P + P^\top}{2}.$$

Thus, P is positive definite if and only if all eigenvalues of its associated Gram matrix, $A = \frac{1}{2}(P + P^\top)$, are strictly positive.

Positive Definite Matrices

More Remarks

- If P is already symmetric, we have $A = \frac{1}{2}(P + P^T) = P$. Thus, symmetric matrices are positive definite if and only if all of their eigenvalues are strictly positive.
- If the quadratic form of a matrix $P \in \mathbb{R}^{n \times n}$ satisfies

$$q(x) = \frac{1}{2}x^T Px \geq 0 \quad \text{for all } x \in \mathbb{R}^n$$

the matrix P is called positive semi-definite. This is the case if and only if all eigenvalues of $\frac{1}{2}(P + P^T)$ are non-negative.

Positive Definite Matrices

More Remarks

- A symmetric and positive semi-definite matrix $P \in \mathbb{R}^{n \times n}$ can be written in the form

$$P = QDQ^{\top} \quad \text{with} \quad D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$ are the non-negative eigenvalues of P .

- Next, let $D^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n})$ denote the diagonal matrix that is obtained by taking the square roots of the diagonal elements of D . Then, the matrix

$$P^{\frac{1}{2}} \stackrel{\text{def}}{=} QD^{\frac{1}{2}}Q^{\top}$$

is called the symmetric square root of P . It satisfies $P^{\frac{1}{2}}P^{\frac{1}{2}} = P$.

Sylvester's Index Theorem

Theorem

- Let $q(x) = x^T Px$ be a non-degenerate quadratic form; that is, $P \in \mathbb{R}^{n \times n}$ is symmetric and invertible. Then there exists an invertible basis transformation matrix $S \in \mathbb{R}^{n \times n}$ and a unique integer, $n - k$, called the index of q , such that

$$q(Sx) = x_1^2 + x_2^2 + \dots + x_k^2 - x_{k+1}^2 - \dots - x_n^2.$$

Proof

- We can diagonalize P as $P = Q^T D Q$. As P is invertible all diagonal entries of D are non-zero.

Sylvester's Index Theorem

Proof (continued)

- Since permutation matrices are orthogonal, we may also assume that the diagonal entries of D are sorted such that $d_{ii} = \lambda_i > 0$ for $1 \leq i \leq k$ and $d_{ii} = \lambda_i < 0$ for $k + 1 \leq i \leq n$.
- We define the diagonal matrix $|D|^{-\frac{1}{2}} = \text{diag}(|\lambda_1|^{-\frac{1}{2}}, \dots, |\lambda_n|^{-\frac{1}{2}})$.
- Next, if we set $S = Q^T |D|^{-\frac{1}{2}}$, then

$$S^T P S = |D|^{-\frac{1}{2}} D |D|^{-\frac{1}{2}} = \text{diag}(1, 1, \dots, 1, -1, \dots, -1).$$

is a diagonal matrix whose first k diagonal entries are 1 and all other diagonal entries are -1 . Now, $q(Sx)$ has the desired form.

Sylvester's Index Theorem

Proof (continued)

- In order to additionally show that the index, $n - k$, is unique, we denote the column vectors of S by $v_1, v_2, \dots, v_n$.
- Next, we introduce the subspaces

$$V = \text{span}(v_1, v_2, \dots, v_k) \quad \text{and} \quad W = \text{span}(v_{k+1}, \dots, v_n).$$

- Now, q is positive definite on V and negative definite on W .
- Next, suppose we have a second basis, such that q is positive definite on V' and negative definite on W' .
- Then we must have $V \cap W' = \{0\}$.

Sylvester's Index Theorem

Proof (continued)

- Thus, the dimension formula for the sum of vector spaces (see Lecture 5) implies that

$$\dim(V) + \dim(W') \leq n.$$

- Because $\dim(V) + \dim(W) = n$, it follows that $\dim(W') \leq \dim(W)$.
- By exchanging the roles of W and W' , we find $\dim(W) \leq \dim(W')$.
- Thus, we have $\dim(W') = \dim(W)$ (and $\dim(V) = \dim(V')$).
- Consequently, the index $\dim(W) = n - k$ is unique.

Index of Critical Points

Remark

- Sylvester's index theorem is simple but important.
- For example, let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a multivariate polynomial of order 2,

$$f(x) = x^{\mathsf{T}}Px + r^{\mathsf{T}}x + s,$$

with coefficients $P \in \mathbb{R}^{n \times n}$, $P = P^{\mathsf{T}}$, $r \in \mathbb{R}^n$ and $s \in \mathbb{R}$.

- If P is invertible, we can set $c = -\frac{1}{2}P^{-1}r$. The point c is-called a non-degenerate critical point of f . It satisfies

$$f(x) = (x - c)^{\mathsf{T}}P(x - c) + \underbrace{(s - r^{\mathsf{T}}P^{-1}r)}_{\text{const.}}$$

Index of Critical Points

Remark (continued)

- Writing f in the form

$$f(x) = (x - c)^T P(x - c) + \text{constant offset}$$

is practical, since we can easily analyze f at c :

- If the index of $q(x) = x^T P x$ is 0, then f has a unique minimizer at c ; that is $f(c) < f(x)$ for all $x \neq c$.
- Similarly, if q has index n , f has a unique maximizer at c .
- In all other cases, c is a saddle point with index $n - k$.
- Applications of the index theorem in analysis use that smooth functions can locally be approximated by multivariate polynomials of degree 2 (not discussed in this course).

Contents

- Introduction
- Quadratic Forms
- Spectral Theorems
- **Summary**

Summary

In this lecture we have learned about:

- converting symmetric bilinear forms to quadratic forms and vice versa,
- computing the Gram matrix of a symmetric bilinear form,
- the change of basis formula for quadratic forms: $B = S^T A S$,
- the spectral theorems for Hermitian matrices, $H = U D \bar{U}^T = \bar{H}^T$,
- its consequences for real symmetric matrices, $P = Q D Q^T = P^T$, and
- Sylvester's index theorem and its interpretation.

Chapter 9

Singular Value Decomposition

Singular Value Decomposition

- Introduction
- Singular Value Decomposition
- Pseudoinverse
- Least-Squares Problems
- Summary

Contents

- Introduction
- Singular Value Decomposition
- Pseudoinverse
- Least-Squares Problems
- Summary

Objectives

In this lecture we will learn about the definition and basic properties of singular value decomposition. Topics include:

- singular value decomposition,
- Penrose's theorem and the pseudoinverse
- least square problems.

Contents

- Introduction
- Singular Value Decomposition
- Pseudoinverse
- Least-Squares Problems
- Summary

Singular Values

- Let $A \in \mathbb{R}^{m \times n}$ be a given matrix. The matrix $A^T A$ is symmetric and, consequently, there exists an orthogonal matrix V such that

$$A^T A = V D V^T,$$

where $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ is a real diagonal matrix (see Lecture 8).

- If v_i denotes the i -th column of V , we have $A^T A v_i = \lambda_i v_i$. Thus,

$$0 \leq \|A v_i\|^2 = v_i^T A^T A v_i = v_i^T (\lambda_i v_i) = \lambda_i \cdot v_i^T v_i = \lambda_i.$$

Hence, the eigenvalues λ_i of $A^T A$ are non-negative, $\lambda_i \geq 0$.

Singular Values

Definition

- Let $A \in \mathbb{R}^{m \times n}$ be a given matrix. And let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the non-negative eigenvalues of $(n \times n)$ -matrix $A^T A$. The square roots

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \quad \dots \quad \sigma_n = \sqrt{\lambda_n}$$

are called the singular values of the matrix A .

Remark

- We recall that permutation matrices P_σ are orthogonal, $P_\sigma^{-1} = P_\sigma^T$. Thus, we may assume—without loss of generality—that the eigenvalues of $A^T A$ and then also the singular values are sorted,

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

Singular Values

Notation

- The rank $r = \text{rank}(A^T A)$ of the matrix $A^T A \in \mathbb{R}^{n \times n}$ satisfies $r \leq n$.
Since we sort all eigenvalues, the last $n - r$ eigenvalues of $A^T A$ are 0,

$$\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0.$$

- We collect the associated singular values in a block matrix of the form

$$\Sigma \stackrel{\text{def}}{=} \left(\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & & 0 \\ 0 & & & 0 \end{array} \right) \in \mathbb{R}^{m \times n}.$$

The zero-block on the lower right has dimension $(m - r) \times (n - r)$.

Eigenvectors

Observations

- Let $v_1, v_2, \dots, v_r$ denote first r columns of V , where $r = \text{rank}(A^T A)$.
- Note that we have

$$(Av_i)^T(Av_j) = v_i^T A^T A v_j = \lambda_j v_i^T v_j = \begin{cases} \lambda_i & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

- We have $\|Av_i\| = \sigma_i > 0$ for $1 \leq i \leq r$ and $\|Av_i\| = 0$ for $i > r$.
- Thus, we can define the orthonormal vectors

$$u_1 \stackrel{\text{def}}{=} \frac{Av_1}{\|Av_1\|}, \quad u_2 \stackrel{\text{def}}{=} \frac{Av_2}{\|Av_2\|}, \quad \dots \quad u_r \stackrel{\text{def}}{=} \frac{Av_r}{\|Av_r\|}.$$

Orthonormal Basis

- The orthonormal vectors $u_1, \dots, u_r$ can be augmented to an orthonormal basis,

$$U = (u_1, \dots, u_r, u_{r+1}, \dots, u_n),$$

such that U is orthogonal matrix, $U^T U = E$.

- Now, it follows that

$$\begin{aligned} U\Sigma &= (\sigma_1 u_1, \dots, \sigma_r u_r, 0, \dots, 0) \\ &= (Av_1, \dots, Av_r, 0, \dots, 0) \\ &= (Av_1, \dots, Av_r, Av_{r+1}, \dots, Av_n) = AV. \end{aligned}$$

- By multiplying with V^T from the right, we find $A = U\Sigma V^T$.

Singular Value Decomposition

Summary

- There exists for every matrix $A \in \mathbb{R}^{m \times n}$ a non-negative matrix $\Sigma \in \mathbb{R}^{m \times n}$ with $\Sigma_{ij} = 0$ for $i \neq j$ and orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that

$$A = U\Sigma V^T.$$

This equation is called the singular-value-decomposition of A .

Singular Value Decomposition

Consequences

- The equation $A = U\Sigma V^T$ implies that we have

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

and

$$A A^T = U \Sigma V^T V \Sigma^T U^T = U \Sigma \Sigma^T U^T.$$

- Consequently, the non-zero eigenvalues of the $(n \times n)$ -matrix $A^T A$ and the $(m \times m)$ -matrix $A A^T$ coincide.

Contents

- Introduction
- Singular Value Decomposition
- Pseudoinverse
- Least-Squares Problems
- Summary

Pseudoinverse

- Let us introduce the shorthands

$$\Sigma_r = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \in \mathbb{R}^{r \times r} \quad \text{and} \quad \Sigma = \left(\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array} \right) \in \mathbb{R}^{m \times n}$$

to describe the block structure of the $(m \times n)$ matrix Σ .

- We assume $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, such that Σ_r is invertible.
- In the following, we additionally define the matrix

$$\Sigma^\dagger \stackrel{\text{def}}{=} \left(\begin{array}{c|c} \Sigma_r^{-1} & 0 \\ \hline 0 & 0 \end{array} \right) \in \mathbb{R}^{n \times m}.$$

It is called the pseudoinverse of the matrix Σ .

Pseudoinverse

Theorem: There exists a unique matrix $C \in \mathbb{R}^{n \times m}$ such that

1. $\Sigma C \Sigma = \Sigma$,
2. $C \Sigma C = C$,
3. $(\Sigma C)^\top = \Sigma C$, and
4. $(C \Sigma)^\top = C \Sigma$.

Namely, the matrix C is given by $C = \Sigma^\dagger$.

Proof

- The matrix $C = \Sigma^\dagger$ satisfies all four requirements. Thus, our focus is on verifying the uniqueness statement.

Pseudoinverse

Proof (continued)

- We write the matrix C and the matrix Σ in block form

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}.$$

- The equation $\Sigma C \Sigma = \Sigma$ implies directly that $C_{11} = \Sigma_r^{-1}$.
- Next, we work out the matrix products

$$C \Sigma = \begin{pmatrix} E & 0 \\ C_{21} \Sigma_r & 0 \end{pmatrix} \quad \text{and} \quad \Sigma C = \begin{pmatrix} E & \Sigma_r C_{12} \\ 0 & 0 \end{pmatrix}.$$

Since, both matrices must be symmetric and since Σ_r is invertible, it follows that $C_{12} = 0$ and $C_{21} = 0$.

Pseudoinverse

Proof (continued)

- After substituting $C_{11} = \Sigma_r^{-1}$ as well as $C_{12} = 0$ and $C_{21} = 0$, it follows that

$$C\Sigma C = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

- But then the equation $C\Sigma C = C$ implies that we must have $C_{22} = 0$. Thus, $C = \Sigma^\dagger$ is the only matrix that satisfies all four conditions of the theorem. This completes our proof.

Penrose's Algebraic Characterization of the Pseudoinverse

Theorem (Penrose 1955)

- Let $A \in \mathbb{R}^{m \times n}$ be a given matrix. Then there exists a unique matrix $B \in \mathbb{R}^{n \times m}$ such that
 - $ABA = A$,
 - $BAB = B$,
 - $(AB)^\top = AB$, and
 - $(BA)^\top = BA$.
- If $A = U\Sigma V^\top$ is a singular value decomposition of A , then this matrix is given by $B = V\Sigma^\dagger U^\top$. It is called the pseudoinverse of A .

Penrose's Algebraic Characterization of the Pseudoinverse

Proof

- Let $A = U\Sigma V^T$ be the singular-value decomposition of A .
- Let us set $C = V^T B U$. Since V and U are orthogonal:
 1. $ABA = A$ if and only if $\Sigma C \Sigma = \Sigma$,
 2. $BAB = B$ if and only if $C \Sigma C = C$,
 3. $(AB)^T = AB$ if and only if $(\Sigma C)^T = \Sigma C$, and
 4. $(BA)^T = BA$ if and only if $(C \Sigma)^T = C \Sigma$.
- Thus, this theorem reduces to our previous theorem.
- It follows that $B = V C U^T = V \Sigma^\dagger U^T$. This completes the proof.

Pseudoinverse

Notation

- The pseudoinverse of a matrix $A \in \mathbb{R}^{m \times n}$ is denoted by $A^\dagger \in \mathbb{R}^{n \times m}$.

Special Cases

- If $m = n = \text{rank}(A)$, the matrix A is invertible and, consequently,

$$AA^\dagger A = A \quad \implies \quad A^\dagger = A^{-1}.$$

- If $m > n = \text{rank}(A)$, the matrix $A^\top A$ is invertible and we have,

$$AA^\dagger A = A \implies A^\top AA^\dagger A = A^\top A \implies A^\dagger A = E.$$

Thus, in this case $A^\dagger = (A^\top A)^{-1} A^\top$ is the unique left inverse of A .

Pseudoinverse

Special Cases (continued)

- Similarly, if $n > m = \text{rank}(A)$, the matrix AA^T is invertible and

$$AA^\dagger A = A \implies AA^\dagger AA^T = AA^T \implies AA^\dagger = E.$$

Thus, in this case $A^\dagger = A^T(AA^T)^{-1}$ is the unique right inverse of A .

Pseudoinverse

Other Properties

- Using the same notation as above, the $(m \times n)$ -matrix Σ satisfies

$$\Sigma^\dagger = (\Sigma^\top \Sigma)^\dagger \Sigma^\top = \Sigma^\top (\Sigma \Sigma^\top)^\dagger$$

$$\text{as well as} \quad \Sigma^\top \Sigma \Sigma^\dagger = \Sigma^\top .$$

- Since any $(m \times n)$ -matrix A can be written in the form $A = U \Sigma V^\top$, we also have

$$A^\dagger = (A^\top A)^\dagger A^\top = A^\top (A A^\top)^\dagger$$

$$\text{as well as} \quad A^\top A A^\dagger = A^\top .$$

Contents

- Introduction
- Singular Value Decomposition
- Pseudoinverse
- **Least-Squares Problems**
- Summary

Least Squares Problem

This section is about least-squares problems of the form

$$\min_x \|Ax - b\|^2.$$

- Here, $A \in \mathbb{R}^{m \times n}$ denotes a given matrix and $b \in \mathbb{R}^m$ a given vector.
- We use the standard Euclidean norm, $\|a\|^2 = a^\top a$.
- We say that $x^* \in \mathbb{R}^n$ is a solution of the least-squares problem if

$$\|Ax^* - b\|^2 \leq \|Ax - b\|^2 \quad \text{for all } x \in \mathbb{R}^n.$$

- From a practical perspective, least-squares problems are among the most important optimization problems all around.

Least Squares Problem

- If the solution set X of the linear equation system $Ax = b$ is non-empty, then the set of solutions $x^\star$ with

$$\|Ax^\star - b\|^2 \leq \|Ax - b\|^2 \quad \text{for all } x \in \mathbb{R}^n$$

is equal to X . We have already discussed this case in Lecture 1.

- More generally, if $x^\star$ is a solution of the least-squares problem,

$$r^\star \stackrel{\text{def}}{=} \|Ax^\star - b\| \leq \|Ax - b\|$$

is called the optimal value of the least-squares problem.

Solution Set of Least-Square Problems

Theorem

- Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ be given. Then, $x^\star \stackrel{\text{def}}{=} A^\dagger b$ is a solution to the least squares problem; that is,

$$\|r^\star\|^2 = \|(AA^\dagger - E)b\|^2 = \|Ax^\star - b\|^2 \leq \|Ax - b\|^2$$

for all $x \in \mathbb{R}^n$. Finally, the set of all solution is given by

$$X = \{x^\star\} + \text{null}(A) = \{A^\dagger b + y \mid Ay = 0\}.$$

Remark

- The above theorem implies that the least-squares problem always has at least one solution, namely, $x^\star = A^\dagger b$; and $\dim(X) = \text{nullity}(A)$.

Solution Set of Least-Squares Problems

Proof

- Let us set $x^* = A^\dagger b$ and $y = x - x^*$.
- We recall that $A^\top A A^\dagger = A^\top$. Thus, we have

$$(Ay)^\top (Ax^* - b) = y^\top A^\top (AA^\dagger b - b) = y^\top A^\top AA^\dagger b - y^\top A^\top b = 0.$$

- Consequently, it follows that

$$\|Ax - b\|^2 = \|Ax^* - b + Ay\|^2 = \|Ax^* - b\|^2 + \|Ay\|^2.$$

The statement of the theorem follows from the fact that $\|Ay\|^2 \geq 0$, where equality holds if and only if $y \in \text{null}(A)$.

Contents

- Introduction
- Singular Value Decomposition
- Pseudoinverse
- Least-Squares Problems
- **Summary**

Summary

In this lecture, we have learned about

- singular value decomposition of $(n \times m)$ -matrices, $A = U\Sigma V^T$,
- Penrose's algebraic characterization of the pseudoinverse including the formula, $A^\dagger = V\Sigma^\dagger U^T$, and
- the solution set of least squares problems,

$$\min_x \|Ax - b\|^2,$$

which is given by $X = \{A^\dagger b\} + \text{null}(A)$.