Review:

Linear algebra and vector spaces

Keywords: vector, matrix, eigenvalue, eigenvector, diagonalization, linear transformation, vector space, metric, norm, inner product, complete space, continuous function, uniformly continuous function, convergent function, uniformly convergent function

Contents

1 Matrices and vectors

1.1 Real Vectors

- n-dimensional space $\mathbb{R}^n$
- elements $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are called n-vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T \quad \text{and } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- scalar product and norm:

$$\begin{aligned} \mathbf{x} \bullet \mathbf{y} &= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\ ||\mathbf{x}|| &= \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \end{aligned}$$

$$\boxed{\mathbf{x} \bullet \mathbf{y} = ||\mathbf{x}|| \cdot ||\mathbf{y}|| \cdot \cos \angle(\mathbf{x}, \mathbf{y})}$$

- $\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_k} \in \mathbb{R}^n$
  - If $a_1, a_2, \dots, a_k \in \mathbb{R}$, then $\mathbf{z} = a_1\mathbf{x_1} + a_2\mathbf{x_2} + \cdots + a_k\mathbf{x_k}$ is called a linear combination of $\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_k}$.
  - $\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_k}$ are called linearly dependent, if there exist $b_1, b_2, \dots, b_k \in \mathbb{R}$ such that $b_1\mathbf{x_1} + b_2\mathbf{x_2} + \cdots + b_k\mathbf{x_k} = \mathbf{0}$ and not all $b_j = 0$.
  - $\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_k}$ are called linearly independent, if a linear combination of the zero vector

    $$b_1\mathbf{x_1} + b_2\mathbf{x_2} + \cdots + b_k\mathbf{x_k} = \mathbf{0}$$

    is possible only with $b_1 = b_2 = \cdots = b_k = 0$.

1.2 Real Matrices

$\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix}, \dots, \mathbf{a}_m = \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} \quad \rightarrow \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

is called an $n \times m$ matrix.

Notation: $\mathbf{A} \in \mathbb{R}^{n \times m}$

- The <u>inverse matrix</u> $\mathbf{A}^{-1}$ of the $n \times n$ matrix $\mathbf{A} = (a_{ij})$ is defined by

$$\mathbf{A^{-1}} \cdot \mathbf{A} \;=\; \mathbf{A} \cdot \mathbf{A^{-1}} \;=\; \mathbf{I_n} \;=\; \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

- For the $n \times n$ matrix $\mathbf{A}$ let $\mathbf{A_{ij}}$ denote the $(n-1) \times (n-1)$ submatrix of $\mathbf{A}$ generated by cancelling the i-th row and the j-th column of $\mathbf{A}$. Then the <u>determinant</u> $\det \mathbf{A}$ is given (recursively) by

$$\boxed{\det(\mathbf{A}) = |\mathbf{A}| = a_{11} \det \mathbf{A}_{11} - a_{12} \det \mathbf{A}_{12} + \cdots + (-1)^{n+1} a_{1n} \det \mathbf{A}_{1n}}$$

- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$

Example 1.1

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

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

1.3 Linear transformations and matrices

Definition 1.1 *A <u>linear transformation</u> is a map $T : \mathbb{R}^m \to \mathbb{R}^n$ such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and all $\lambda, \mu \in \mathbb{R}$ we have:*

$$\boxed{T(\lambda \cdot \mathbf{x} + \mu \cdot \mathbf{y}) \;=\; \lambda \cdot T(\mathbf{x}) + \mu \cdot T(\mathbf{y})}$$

Each $n \times m$ matrix $\mathbf{A}$ defines a linear transformation by matrix multiplication

$$T_{\mathbf{A}}(\mathbf{x}) \;=\; \mathbf{A} \cdot \mathbf{x} \;=\; x_1 \mathbf{a}_1 + \cdots + x_m \mathbf{a}_m.$$

The image of the vector $\mathbf{x} \in \mathbb{R}^m$ is a linear combination of the column vectors of the matrix $\mathbf{A}$.

1.4 Complex matrices and vectors

Sometimes it is helpful to allow complex matrices and vectors (matrices whose elements are complex numbers). A complex matrix can be viewed as a combination of two real matrices:

$$\begin{aligned}\mathbf{A} &= \begin{pmatrix} a_{11}+ib_{11} & a_{12}+ib_{12} & \dots & a_{1m}+ib_{1m} \\ a_{21}+ib_{21} & a_{22}+ib_{22} & \dots & a_{2m}+ib_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}+ib_{n1} & a_{n2}+ib_{n2} & \dots & a_{nm}+ib_{nm} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} + i \cdot \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{pmatrix}\end{aligned}$$

1.5 Matrix calculus

1a.	$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$	1b.	$\mathbf{AB}\neq\mathbf{BA}$
2a.	$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$	2b.	$(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})$
3a.	$\mathbf{A}+\mathbf{0}=\mathbf{A}$	3b.	$\mathbf{AI}=\mathbf{IA}=\mathbf{A}$, ($\mathbf{A}$ square)

4. $\mathbf{AB}=\mathbf{0} \quad \not\Rightarrow \quad \mathbf{A}=\mathbf{0}$ or $\mathbf{B}=\mathbf{0}$
5. $\mathbf{AB}=\mathbf{AC} \quad \not\Rightarrow \quad \mathbf{B}=\mathbf{C}$

6. $\lambda(\mathbf{A}+\mathbf{B}) = \lambda\mathbf{A}+\lambda\mathbf{B} \quad \lambda\in\mathbb{R}$
7. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$
8. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$

9. $(\mathbf{A^{-1}})^{-1} = \mathbf{A}$
10. $(\mathbf{AB})^{-1} = \mathbf{B^{-1}A^{-1}}$
11. $(\mathbf{A^T})^{\mathbf{T}} = \mathbf{A}$
12. $(\mathbf{A}+\mathbf{B})^{\mathbf{T}} = \mathbf{A^T}+\mathbf{B^T}$
13. $(\mathbf{AB})^{\mathbf{T}} = \mathbf{B^TA^T}$
14. $(\mathbf{A^{-1}})^{\mathbf{T}} = (\mathbf{A^T})^{-1}$

For $\mathbf{A}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad-bc\neq 0$ is $\mathbf{A}^{-1}=\dfrac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

All these definitions and results can be generalized to vectors and matrices with complex entries.

2 Eigenvalues and eigenvectors

2.1 Definition and determination

Definition 2.1 *If* $\mathbf{A}$ *is a real (or complex)* $n \times n$ *matrix, then a (complex) number* λ *is an* eigenvalue *of* $\mathbf{A}$ *if there is a nonzero (complex) vector* $\mathbf{x} \in \mathbb{C}^{\mathbf{n}}$ *such that*

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

Then $\mathbf{x}$ *is an* eigenvector *of* $\mathbf{A}$ *(associated with* λ*).*

Remark: If $\mathbf{x}$ is an eigenvector associated with the eigenvalue λ, then so is $\alpha\mathbf{x}$ for every real number $\alpha \neq 0$.

$$\mathbf{A}\,(\alpha\,\mathbf{x}) \;=\; \alpha\,\mathbf{A}\,\mathbf{x} \;=\; \alpha\,(\lambda\,\mathbf{x}) \;=\; \lambda\,(\alpha\,\mathbf{x})$$

How to find eigenvalues? The equation can be written as

$$\begin{aligned} \mathbf{A}\,\mathbf{x} &= \lambda\,\mathbf{x} \\ \Leftrightarrow \quad \mathbf{A}\,\mathbf{x} - \lambda\,\mathbf{I}\,\mathbf{x} &= \mathbf{0} \\ \Leftrightarrow \quad (\mathbf{A} - \lambda\,\mathbf{I})\,\mathbf{x} &= \mathbf{0} \end{aligned}$$

This is a homogeneous linear system of equations. It has a solution $\mathbf{x} \neq \mathbf{0}$ if and only if the matrix $(\mathbf{A} - \lambda\,\mathbf{I})$ is singular which means that it has determinant equal to 0.

$$\boxed{(\mathbf{A} - \lambda\,\mathbf{I}) \text{ singular} \quad \Leftrightarrow \quad \underbrace{\det(\mathbf{A} - \lambda\,\mathbf{I})}_{p_A(\lambda)} = 0}$$

$p_A(\lambda) = 0$ is called the characteristic equation of $\mathbf{A}$. The function $p_A(\lambda)$ is a polynomial of degree n in λ, called the characteristic polynomial of $\mathbf{A}$.

Determination of the eigenvalues and eigenvectors

1. The polynomial equation $p_A(\lambda) = 0$ has always n complex solutions (counted with multiplicity) and may have no real solutions. If $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ are the pairwise distinct solutions (the eigenvalues of $\mathbf{A}$) with the multiplicities $k_1, \ldots, k_r$ then the characteristic polynomial can be written as

   $$p_A(\lambda) \;=\; (\lambda_1 - \lambda)^{k_1}\,(\lambda_2 - \lambda)^{k_2} \ldots (\lambda_r - \lambda)^{k_r}.$$

   The multiplicity k_i of the zero λ_i is called algebraic multiplicity of the eigenvalue λ_i. Generally, the determination of the (exact) zeros is impossible for $n \geq 5$ and we have to use numerical methods.

2. For each eigenvalue λ_i $(1 \leq i \leq r)$ we compute the so called eigenspace for λ_i

   $$V(\lambda_i) \;=\; \{\, \mathbf{x} \in \mathbb{C}^n \mid (\mathbf{A} - \lambda_i \mathbf{I})\,\mathbf{x} = \mathbf{0} \,\}.$$

   The dimension of the vector space $V(\lambda_i)$ is called the geometric multiplicity of the eigenvalue λ_i.

Definition 2.2 *The* spectral radius *of a quadratic matrix* A *is the real number*

$$\rho(A) \;:=\; \max\{|\lambda_1|, \ldots, |\lambda_r|\,\}.$$

2.2 *Generalized Eigenvectors*

To solve some interesting problems we have to generalize the notion of eigenvectors.

Definition 2.3 *A vector $\mathbf{x} \in \mathbb{C}^n$ is called* <u>*generalized eigenvector*</u> *of degree $l \in \mathbb{N}$ associated to the eigenvalue λ of $\mathbf{A}$, if*

$$(\mathbf{A} - \lambda \mathbf{I})^l \, \mathbf{x} \; = \; \mathbf{0} \quad \text{and} \quad (\mathbf{A} - \lambda \mathbf{I})^{l-1} \, \mathbf{x} \; \neq \; \mathbf{0}.$$

Of course, an eigenvector is a generalized eigenvector of degree 1.

Example 2.1 *The matrix*

$$\mathbf{A} \;=\; \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

has the eigenvalue 1 of (algebraic) multiplicity 3 with $\dim V(1) = 1$ (geometric multiplicity). We have:

$$\begin{array}{lll} (\mathbf{A} - \mathbf{I})\, \mathbf{e}_1 \;=\; \mathbf{0} & (\mathbf{A} - \mathbf{I})\, \mathbf{e}_2 \;=\; \mathbf{e}_1 & (\mathbf{A} - \mathbf{I})^2\, \mathbf{e}_2 \;=\; \mathbf{0} \\ (\mathbf{A} - \mathbf{I})\, \mathbf{e}_3 \;=\; \mathbf{e}_1 + \mathbf{e}_2 & (\mathbf{A} - \mathbf{I})^2\, \mathbf{e}_3 \;=\; \mathbf{e}_1 & (\mathbf{A} - \mathbf{I})^3\, \mathbf{e}_3 \;=\; \mathbf{0} \end{array}$$

This means, that $\mathbf{e}_1$ is an eigenvector, $\mathbf{e}_2$ a generalized eigenvector of degree 2 and $\mathbf{e}_3$ a generalized eigenvector of degree 3.

Theorem 2.1 *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a complex (or real) matrix with*

$$p_A(\lambda) \;=\; (\lambda_1 - \lambda)^{k_1} \, (\lambda_2 - \lambda)^{k_2} \dots (\lambda_r - \lambda)^{k_r}.$$

- *Let λ be an eigenvalue of $\mathbf{A}$ of (algebraic) multiplicity l. Then there exist l linearly independent generalized eigenvectors (of degree $\leq l$). This means:*

$$\dim\{\, \mathbf{x} \in \mathbb{C}^n \mid (\mathbf{A} - \lambda \mathbf{I})^l \, \mathbf{x} \;=\; \mathbf{0} \,\} \;=\; l.$$

- *Generalized eigenvectors associated to pairwise different eigenvalues of $\mathbf{A}$ are linearly independent.*

- *There exists a basis $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$ of $\mathbb{C}^n$ consisting of generalized eigenvectors of $\mathbf{A}$. If $\mathbf{P}$ is the matrix with this basis as the columns, then*

$$\mathbf{P}^{-1} \, \mathbf{A} \, \mathbf{P} \;=\; \begin{pmatrix} \boxed{\mathbf{A}_1} & & & \mathbf{0} \\ & \boxed{\mathbf{A}_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \boxed{\mathbf{A}_r} \end{pmatrix}$$

with $\mathbf{A}_i \in \mathbb{C}^{k_i \times k_i}$ for all $i = 1, 2, \dots, r$.

Example 2.2 *Let* $n = 2$ *and* $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

1. *Characteristic polynomial:*

$$\begin{aligned} p_A(\lambda) &= \det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} \\ &= \lambda^2 - \underbrace{(a+d)}_{=:tr(A)}\lambda + \underbrace{ad-bc}_{=\det(A)} = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \end{aligned}$$

   with $\lambda_{1,2} = \dfrac{a+d}{2} \pm \sqrt{\dfrac{(a+d)^2}{4} - \det(A)}$.

2. *For each* λ_i $(i = 1, 2)$ *we solve the linear system*

$$\begin{pmatrix} a-\lambda_i & b \\ c & d-\lambda_i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

If $n = 2$, *we have four different cases:*

1. $\lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_1 \neq \lambda_2$
   Example: $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

2. $\lambda = \lambda_1 = \lambda_2 \in \mathbb{R}$ *with* $\dim V(\lambda) = 2$
   Example: $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

3. $\lambda = \lambda_1 = \lambda_2 \in \mathbb{R}$ *with* $\dim V(\lambda) = 1$
   Example: $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$

4. $\lambda_2 = \overline{\lambda_1} \in \mathbb{C} - \mathbb{R}$
   Example: $\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$ *with* $\phi \neq k\pi$

3 Diagonalization

Let $\mathbf{A}$ and $\mathbf{P}$ be $n \times n$ matrices with $\mathbf{P}$ invertible. Then $\mathbf{A}$ and $\mathbf{P^{-1}AP}$ have the same eigenvalues (because they have the same characteristic polynomial).

Definition 3.1 *An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if there is an invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{D}$ such that*

$$\mathbf{P^{-1}AP} \quad = \quad \mathbf{D}.$$

Two natural questions:

1. Which square matrices are diagonalizable?
2. If $\mathbf{A}$ is diagonalizable, how do we find the matrix $\mathbf{P}$?

Theorem 3.1 *An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if it has a set of n linearly independent eigenvectors $\mathbf{p_1}, \ldots, \mathbf{p_n}$. In this case,*

$$\mathbf{P^{-1}AP} \quad = \quad diag(\lambda_1, \ldots, \lambda_n),$$

where $\mathbf{P}$ is the matrix with $\mathbf{p_1}, \ldots, \mathbf{p_n}$ as its columns, and $\lambda_1, \ldots, \lambda_n$ are the corresponding eigenvalues.

Many of the matrices encountered in economics are (real) symmetric and for these matrices we have the following important result.

Theorem 3.2 (Spectral Theorem for symmetric matrices) *If the real $n \times n$ matrix $\mathbf{A}$ is symmetric ($\mathbf{A} = \mathbf{A^T}$), then:*

1. *All n eigenvalues $\lambda_1, \ldots, \lambda_n$ are real.*
2. *Eigenvectors that correspond to different eigenvalues are orthogonal.*
3. *There exists an orthogonal and real matrix $\mathbf{P}$ ($\mathbf{P^{-1}} = \mathbf{P^T}$) such that*

$$\mathbf{P^{-1}AP} \quad = \quad diag(\lambda_1, \ldots, \lambda_n).$$

The columns $\mathbf{p_1}, \ldots, \mathbf{p_n}$ of the matrix $\mathbf{P}$ are eigenvectors of unit length corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$.

Example 3.1 *The matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}$ has the eigenvalues and eigenvectors*

$$\lambda_1 = 2 \qquad \mathbf{p}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\lambda_2 = 3 \qquad \mathbf{p}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Hence $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, $\mathbf{P^{-1}} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ and:

$$\mathbf{P^{-1}AP} \quad = \quad \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad = \quad \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

4 Vector spaces

Definition 4.1 *A* <u>*(real) vector space*</u> *is a set V together with two operations*

$$\begin{aligned} + : V \times V &\to V \text{ (vector addition)} \\ \cdot : \mathbb{R} \times V &\to V \text{ (scalar multiplication)} \end{aligned}$$

such that for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$:

- $u + (v + w) = (u + v) + w$
- $u + v = v + u$
- *There exists an element $0 \in V$, called the* <u>*zero element*</u>*, such that $v + 0 = 0 + v = v$ for all $v \in V$.*
- *For every $v \in V$ there exists an element $-v \in V$, called the* <u>*additive inverse of v*</u>*, such that $v + (-v) = 0$.*
- $\alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) \cdot v$
- $1 \cdot v = v$ *($1 \in \mathbb{R}$)*
- $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$
- $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$

Example 4.1 *The space $\mathbb{R}^n$ with the well-known (componentwise) vector addition and scalar multiplication is a real vector space.*

Example 4.2 *The space $\mathbb{C}^n$ with the well-known (componentwise) vector addition and scalar multiplication is a real vector space.*

Example 4.3 *The set $\mathcal{F}$ of functions $\mathbb{R}^n \to \mathbb{R}$ can be given the structure of a (real) vector space, where the operations are defined pointwise. For any $f, g : \mathbb{R}^n \to \mathbb{R}$ and any $\alpha \in \mathbb{R}$ define:*

$$\begin{aligned} + : \mathcal{F} \times \mathcal{F} \to \mathcal{F} \qquad & (f + g)(x) = f(x) + g(x) \\ \cdot : \mathbb{R} \times \mathcal{F} \to \mathcal{F} \qquad & (\alpha \cdot f)(x) = \alpha \cdot f(x) \end{aligned}$$

5 Metric spaces

Definition 5.1 *Let X be an arbitrary set (for instance a real vector space). A metric d on X is a function*

$$d : X \times X \quad \to \quad \mathbb{R}$$

such that for all $x, y, z \in X$:

1. $d(x,y) = 0 \Longleftrightarrow x = y$
2. $d(x,y) = d(y,x)$
3. $d(x,z) \leq d(x,y) + d(y,z)$ *(triangle inequality)*

The pair (X,d) is called a metric space.
Let $x \in X$ and $r > 0$. Then

$$B_r(x) \quad := \quad \{ \, y \in X \mid d(x,y) < r \, \} \ \subset \ X$$

is called the open ball with center x and diameter r.

Example 5.1 *For each set X*

$$d(x,y) := \begin{cases} 0 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases}$$

is a metric, the so-called discrete metric.

Example 5.2 $X = \mathbb{R}^n$ *and* $d(x,y) = ||x-y|| := \sqrt{(x_1-y_1)^2 + \cdots + (x_n-y_n)^2}$

Example 5.3 $X = \mathbb{R}^n$ *and*

$$d(x,y) := \begin{cases} ||x|| + ||y|| & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

Definition 5.2 *Let (X,d) be a metric space. A sequence $x_1, x_2, \ldots$ in X is called a Cauchy sequence if for every real number $\epsilon > 0$ there is an integer $N = N(\epsilon)$ such that for all $n, m > N$ we have*

$$d(x_n, x_m) \ < \ \epsilon.$$

(X,d) is called a complete metric space if every Cauchy sequence in X has a limit that is also an element in X.

$$\lim_{n\to\infty} x_n \ = \ x \in X$$

Example 5.4 *The set of rational numbers $\mathbb{Q}$ with the metric $d(x,y) = |x-y|$ is not complete. Consider for instance the Cauchy sequence*

$$x_1 = 1 \text{ and } x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n} \in \mathbb{Q}.$$

The limit is $\sqrt{2} \notin \mathbb{Q}$.

Example 5.5 *The set $\mathbb{R}$ with the metric $d(x,y) = |x-y|$ is complete.*

6 Normed vector spaces

6.1 Definition and examples

Definition 6.1 *Given a real vector space V. A norm on V is a real valued function*

$$|| \,.\, || : V \;\to\; \mathbb{R}$$

such that for all $v, w \in V$ and all $\lambda \in \mathbb{R}$:

1. $||v|| \geq 0$ *and* $||v|| = 0 \iff v = 0$
2. $||\lambda \cdot v|| = |\lambda| \cdot ||v||$
3. $||v + w|| \leq ||v|| + ||w||$ *(triangle inequality)*

The pair $(V, || \,.\, ||)$ is called a normed (vector) space.

The norm $|| \,.\, ||$ induces always a metric on V by $d(v, w) := ||v - w||$. Hence, a normed vector space is always a metric (vector) space (with this induced metric).

Definition 6.2 *A Banach space is a complete normed vector space.*

Example 6.1 $V = \mathbb{R}^n$

- $||\mathbf{v}||_2 \;:=\; \sqrt{v_1^2 + \cdots + v_n^2} \;=\; \left(\sum_{k=1}^{n} |v_i|^2 \right)^{1/2}$ *(Euclidean norm)*
- $||\mathbf{v}||_1 \;:=\; \sum_{k=1}^{n} |v_i|$ *(Manhattan norm)*
- $||\mathbf{v}||_p \;:=\; \left(\sum_{k=1}^{n} |v_i|^p \right)^{1/p}$ *(p-norm with $p \geq 1$ a real number)*
- $||\mathbf{v}||_\infty \;:=\; \max_i |v_i|$ *(maximum norm)*

Example 6.2 *Let $|| \,.\, ||$ be a norm on $\mathbb{R}^n$. The space $V = \mathbb{R}^{n \times n}$ of $n \times n$-matrices with coefficients in $\mathbb{R}$ is a real vector space. The map*

$$|| \,.\, || : \mathbb{R}^{n \times n} \;\to\; \mathbb{R} \qquad \textit{defined by} \qquad ||A|| := \sup_{\mathbf{v} \neq \mathbf{0}} \frac{||A\mathbf{v}||}{||\mathbf{v}||}$$

is called the (by the vector norm $|| \,.\, ||$) induced matrix norm. It is possible to prove, but not trivial, that

$$\begin{aligned}
||A||_\infty &= \max_{i=1,\ldots,n} \sum_{j=1}^{n} |a_{ij}| \\
||A||_1 &= \max_{j=1,\ldots,n} \sum_{i=1}^{n} |a_{ij}| \\
||A||_2 &= \sqrt{\rho(A^T A)}
\end{aligned}$$

The following result shows that the induced matrix norm is actually a norm on the vector space $\mathbb{R}^{n\times n}$.

Lemma 6.1 *Let $||\,.\,||$ be a norm on $\mathbb{R}^n$ resp. the induced matrix norm on $\mathbb{R}^{n\times n}$. Then*

1. $||A\mathbf{v}|| \;\leq\; ||A||\cdot||\mathbf{v}||$ *for all $A \in \mathbb{R}^{n\times n}$ and $\mathbf{v} \in \mathbb{R}^n$.*
2. *For all $A \in \mathbb{R}^{n\times n}$ there is a $\mathbf{v}_A \in \mathbb{R}^n$ such that $||\mathbf{v}_A|| = 1$ and $||A\mathbf{v}_A|| = ||A||$.*
3. *The induced matrix norm is a norm on $\mathbb{R}^{n\times n}$.*
   - (a) $||A|| \geq 0$ *for all $A \in \mathbb{R}^{n\times n}$*
   - (b) $||\lambda\cdot A|| = |\lambda|\cdot||A||$ *for all $A \in \mathbb{R}^{n\times n}$ and $\lambda \in \mathbb{R}$*
   - (c) $||A + B|| \leq ||A|| + ||B||$ *for all $A, B \in \mathbb{R}^{n\times n}$*

Continuous functions

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuous if small changes in the independent variables cause only small changes in the function values. The precise $\epsilon - \delta$- definition is as follows:

Definition 6.3 *Let $\mathbf{f} : D(\subset \mathbb{R}^n) \to \mathbb{R}^n$ be a function, $\mathbf{a} \in D$ and $||\,.\,||$ a norm on $\mathbb{R}^n$. Then $\mathbf{f}$ is continuous at $\mathbf{a}$ if for every $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that:*

$$\text{For all } \mathbf{x} \in D \text{ with } ||\mathbf{a} - \mathbf{x}|| < \delta \;\Rightarrow\; ||\mathbf{f}(\mathbf{a}) - \mathbf{f}(\mathbf{x})|| < \epsilon.$$

$\mathbf{f}$ is called continuous on D if $\mathbf{f}$ is continuous at every point $\mathbf{a} \in D$.

The following property is much stronger.

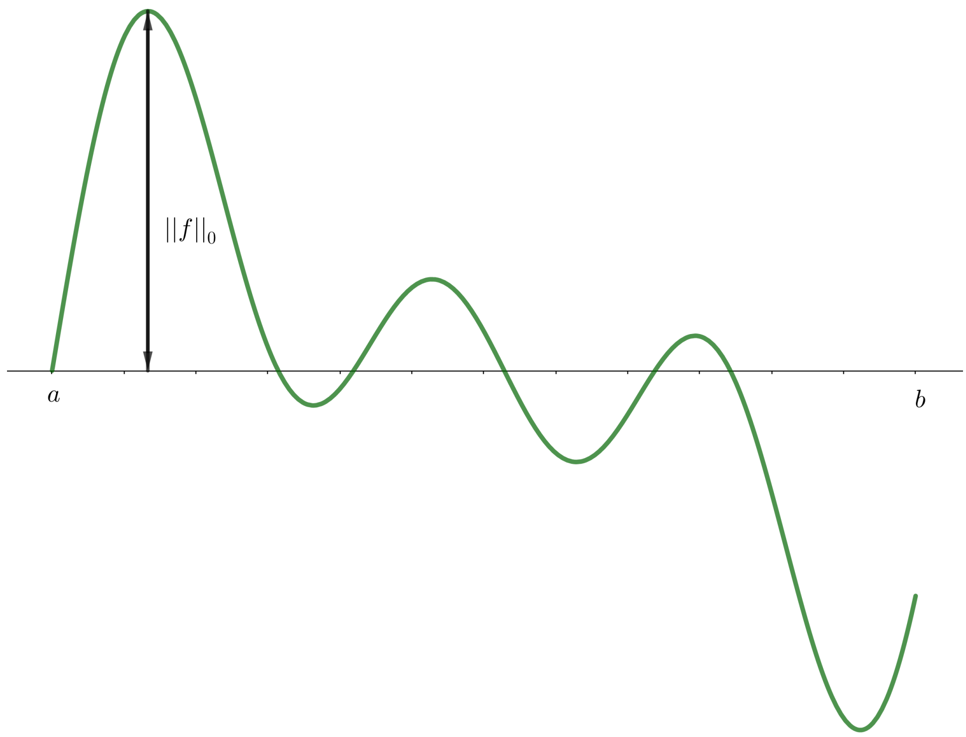
Definition 6.4 *Let $\mathbf{f} : D(\subset \mathbb{R}^n) \to \mathbb{R}^n$ be a function, $\mathbf{a} \in D$ and $||\,.\,||$ a norm on $\mathbb{R}^n$. Then $\mathbf{f}$ is uniformly continuous on D if for every $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that:*

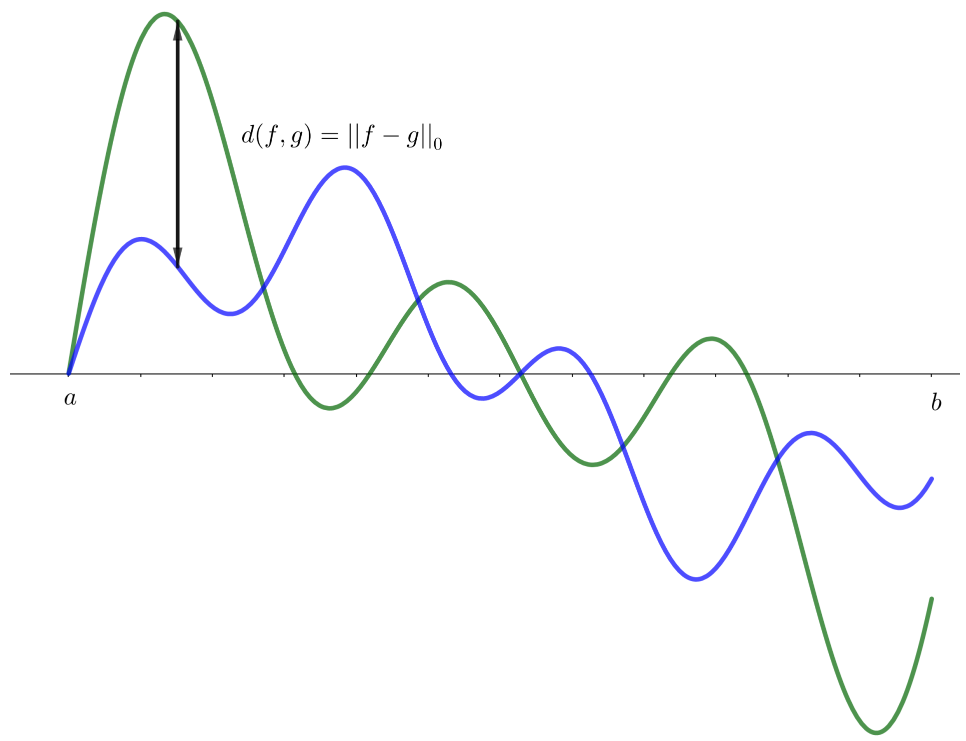
$$\text{For all } \mathbf{x}, \mathbf{y} \in D \text{ with } ||\mathbf{x} - \mathbf{y}|| < \delta \;\Rightarrow\; ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})|| < \epsilon.$$

Example 6.3 *Let $C^0(a,b)$ be the vector space of all continuous functions f on the interval $[a,b]$. Then*

$$||f||_0 := \max_{x\in[a,b]} |f(x)|$$

is a norm on $C^0(a,b)$.





Example 6.4 *Let $C^n(a,b)$ be the vector space of all functions f on the interval $[a,b]$ which are continuous and have continuous derivatives up to order n. Then*

$$||f||_n := \sum_{k=0}^{n} \max_{x\in[a,b]} |f^{(k)}(x)|$$

is a norm on $C^n(a,b)$.

Thus, two functions in $C^1(a,b)$ are regarded as close together if both the functions and their derivatives are close together, since

$$d(f,g) = ||f-g||_1 = \max_{x\in[a,b]} |f(x)-g(x)| + \max_{x\in[a,b]} |f'(x)-g'(x)| \ < \ \epsilon$$

implies that

$$|f(x)-g(x)| \ < \ \epsilon \text{ and } |f'(x)-g'(x)| \ < \ \epsilon \text{ for all } x \in [a,b].$$

6.2 A short look at convergence of functions

A sequence of points converge to a limit if they get physically closer and closer to it. When do functions converge to a limit function? The simplest idea is the following:

Definition 6.5 *A sequence of functions $f_n : [a,b] \to \mathbb{R}$ converges pointwise to a limit function $f : [a,b] \to \mathbb{R}$ if for each $x \in [a,b]$*

$$\lim_{n\to\infty} f_n(x) \quad = \quad f(x).$$

Then f is called the pointwise limit of the sequence (f_n) and we write

$$f_n \;\to\; f \qquad \text{or} \qquad \lim_{n\to\infty} f_n = f.$$

The following requirement of convergence is stronger.

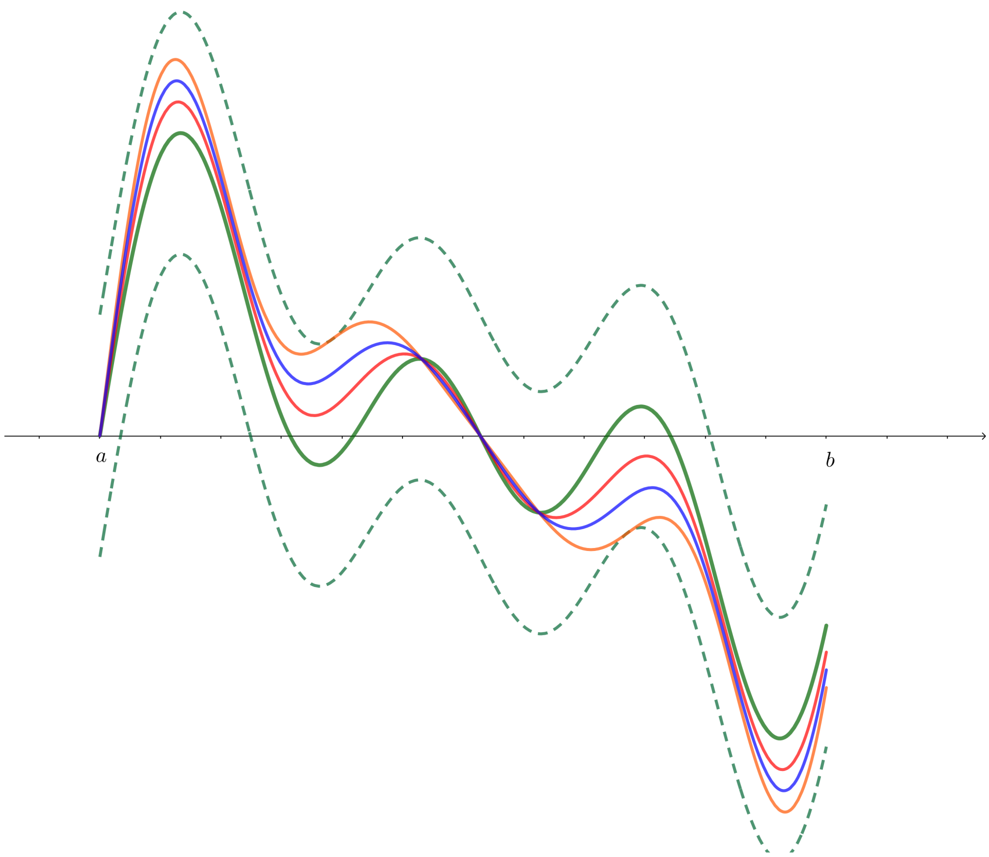
Definition 6.6 *A sequence of functions $f_n : [a,b] \to \mathbb{R}$ converges uniformly to a limit function $f : [a,b] \to \mathbb{R}$ if for each $\epsilon > 0$ there is an $N = N(\epsilon)$ such that for all $n \geq N$ and* **all** *$x \in [a,b]$*

$$|f_n(x) - f(x)| \quad < \quad \epsilon.$$

Then f is called the uniform limit of the sequence (f_n) and we write

$$f_n \;\rightrightarrows\; f \qquad \text{or} \qquad \text{unif}\lim_{n\to\infty} f_n = f.$$

The intuition about uniform convergence is crucial. Draw a tube T of vertical distance ϵ around the graph of f. For n large enough, the graph of f_n should lie completely in T.

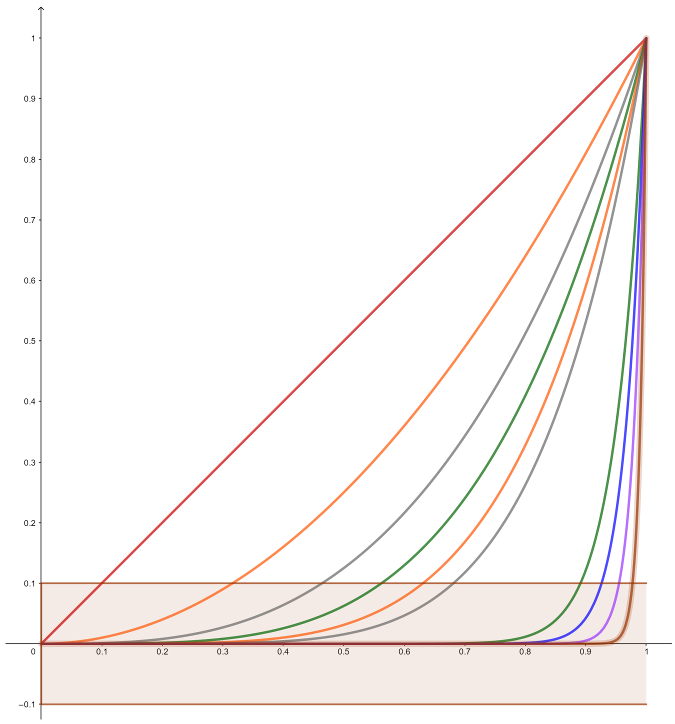


Clearly: If $f_n \rightrightarrows f$ then $f_n \to f$.

The following standard example explains the difference of the two definitions.

Example 6.5 *Define $f_n : (0,1) \to \mathbb{R}$ by $f_n(x) = x^n$. For each $x \in (0,1)$ it is clear, that $f_n(x) = x^n \to 0$. The sequence of functions converges pointwise to the zero function $f(x) = 0$.*

But it does not converge uniformly! Take $\epsilon = 0.1$. The point $x_n = \sqrt[n]{0.5} \in (0,1)$ is mapped to 0.5 by f_n.



We also see: The sequence of continuous functions $f_n(x) = x^n$ on $[0,1]$ converges pointwise (but not uniformly) to the noncontinuous function

$$f(x) = \begin{cases} 0 & \text{if} \quad 0 \leq x < 1 \\ 1 & \text{if} \quad x = 1 \end{cases}$$

The pointwise limit of a sequence of continuous functions has not to be continuous.

We may ask the natural questions: **Which properties of functions are preserved under uniform convergence?** The answers are found in

Reference 1: *Charles Chapman Pugh, Real mathematical analysis, Spinger, Chapter 4*

We only remark the following:

Theorem 6.1 *The uniform limit of continuous functions is continuous.*

What is the connection between uniform convergence and the maximum norm on the normed space $C^0(a,b)$? Let $f_n, f : [a,b] \to \mathbb{R}$ be continuous functions on $[a,b]$. If

$$d(f_n, f) \;=\; ||f_n - f||_0 \;=\; \max_{x\in[a,b]} \,|f_n(x) - f(x)| \;\to\; 0 \text{ then } f_n \Rightarrow f$$

and conversely.

Theorem 6.2 *Convergence with respect to the norm* $||\;.\;||_0$ *is equivalent to uniform convergence.*

Theorem 6.3 *The normed space* $(C^0(a,b), ||\;.\;||_0)$ *is complete.*

7 Inner product spaces

Definition 7.1 *Given a real vector space V. An inner product on V is a map*

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

such that for all $u, v, w \in V$ and all $\lambda, \mu \in \mathbb{R}$:

1. $\langle u, v \rangle = \langle v, u \rangle$
2. $\langle \lambda \cdot u + \mu \cdot v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle$
3. $\langle v, v \rangle > 0$ *for all* $v \neq 0$

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

An inner product $\langle \cdot, \cdot \rangle$ on V induces a norm on V by $||v|| := \sqrt{\langle v, v \rangle}$. Inner product spaces are normed spaces (and hence metric spaces).

Definition 7.2 *A Hilbert space is a complete inner product space.*

Example 7.1 $V = \mathbb{R}^n$ *and* $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \bullet \mathbf{v} = u_1 v_1 + \cdots + u_n v_n$

Example 7.2 *Let $C^0(a, b)$ be the vector space of all continuous functions f on the interval $[a, b]$. Then*

$$\langle f, g \rangle = \int_a^b f(t) g(t) dt$$

is an inner product on $C^0(a, b)$.

Example 7.3 *Let V be the real vector space of all random variables $X : \Omega \rightarrow \mathbb{R}$. Then $\langle X, Y \rangle = E(X \cdot Y)$ is an inner product.*