Analysis in a Nutshell

- Introduction
- Vector spaces and norms
- Hilbert spaces
- Landau's symbol
- Nonlinear Functions
- Differentiable Functions
- Taylor Expansions

Contents

Objective

- This lecture covers the most important analysis concepts that are needed in the Numerical Analysis course.
- This overview is NOT complete.
- You can use these slides as a check whether you know about all the background stuff.
- If not: start searching online for the keywords in this lecture or re-read your favorite analysis text book.

Objective

- This lecture covers the most important analysis concepts that are needed in the Numerical Analysis course.
- This overview is NOT complete.
- You can use these slides as a check whether you know about all the background stuff.
- If not: start searching online for the keywords in this lecture or re-read your favorite analysis text book.

Objective

- This lecture covers the most important analysis concepts that are needed in the Numerical Analysis course.
- This overview is NOT complete.
- You can use these slides as a check whether you know about all the background stuff.
- If not: start searching online for the keywords in this lecture or re-read your favorite analysis text book.

Objective

- This lecture covers the most important analysis concepts that are needed in the Numerical Analysis course.
- This overview is NOT complete.
- You can use these slides as a check whether you know about all the background stuff.
- If not: start searching online for the keywords in this lecture or re-read your favorite analysis text book.

Contents

Vector space

Let F be a field. The set V together with an addition operation "$+$" and a scalar multiplication "*" is called a vector space, if for all $u, v, w \in V$ and all $a, b \in F$:

1. Associativity: $(u + v) + w = u + (v + w)$
2. Commutativity: $u + v = v + u$
3. There exists $0 \in V$ with $v + 0 = v$
4. There exists $-v \in V$ with $v + (-v) = 0$
5. Compatibility $a * (b * v) = (a * b) * v$
6. There exists $1 \in F$ with $1 * v = v$.
7. Distributivity: $a(u + v) = au + av$ and $(a + b)v = av + bv$

Norms

A norm on a vector space V is a function $\|\cdot\| : V \to \mathbb{R}$ such that for all $a \in F$ and all $u, v \in V$:

1. $\|a * v\| = |a|\|v\|$ (absolute homogeneity),

2. $\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality),

3. $\|v\| = 0$ implies that v is the zero vector.

Examples (finite dimensional)

Vector space $V = \mathbb{R}^n$; examples for norms

1. Euclidean norm: $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$.
2. Maximum norm: $\|x\|_\infty = \max_{i \in \{1,\ldots,n\}} |x_i|$.
3. 1-norm: $\|x\|_1 = \sum_{i=1}^n |x_i|$.

Examples (induced norms)

Vector space $V = \mathbb{R}^{n\times m}$; examples for induced norms

1. Spectral norm:

$$\|A\|_2 = \max_{x\in\mathbb{R}^m} \frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\lambda_{\max}(A^\top A)}$$

2. Matrix ∞-norm:

$$\|A\|_\infty = \max_{x\in\mathbb{R}^m} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1\le i\le n} \sum_{j=1}^{m} |A_{i,j}|$$

3. Matrix 1-norm:

$$\|A\|_1 = \max_{x\in\mathbb{R}^m} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1\le j\le m} \sum_{i=1}^{n} |A_{i,j}|$$

Examples (infinite dimensional)

Vector space $V = L^2[-1,1]$; examples for norms

1. L_2-norm: $\|f\|_{L_2} = \sqrt{\int_{-1}^{1} f(t)^2 \,\mathrm{d}t}$.
2. L_∞-norm: $\|f\|_{L_\infty} = \max_{t\in[-1,1]} |f(t)|$.
3. L_1-norm: $\|f\|_{L_1} = \int_{-1}^{1} |f(t)| \,\mathrm{d}t$.

Equivalence of norms

Let V be a finite dimensional vector space. For any norm $\|\cdot\| : V \to \mathbb{R}$ there exists constants $0 < m < M < \infty$ with

$$\forall x \in V, \qquad m\|x\|_\infty \leq \|x\| \leq M\|x\|_\infty$$

Warning:

- In infinite dimensional spaces norms are not equivalent.
- Example: $V = L^2[0,1]$

$$f(t) = t^n \qquad \Rightarrow \qquad \frac{\|f\|_{L_\infty}}{\|f\|_{L_2}} = \sqrt{2n+1}$$

What happens for $n \to \infty$?

Equivalence of norms

Let V be a finite dimensional vector space. For any norm $\|\cdot\| : V \to \mathbb{R}$ there exists constants $0 < m < M < \infty$ with

$$\forall x \in V, \qquad m\|x\|_\infty \leq \|x\| \leq M\|x\|_\infty$$

Warning:

- In infinite dimensional spaces norms are not equivalent.
- Example: $V = L^2[0,1]$

$$f(t) = t^n \qquad \Rightarrow \qquad \frac{\|f\|_{L_\infty}}{\|f\|_{L_2}} = \sqrt{2n+1}$$

What happens for $n \to \infty$?

Cauchy sequences

Convergent sequence

- A sequence $x_1, x_2, x_3, \ldots$ of real numbers is called convergent to a point $x^* \in \mathbb{R}$ if

$$\lim_{k \to \infty} |x_k - x^*| = 0 \; .$$

Cauchy sequence

- A sequence $x_1, x_2, x_3, \ldots \in \mathbb{R}^n$ is called a Cauchy sequence, if for every $\varepsilon > 0$, there exists $N < \infty$ such that:

$$\forall m, n > N, \qquad \|x_m - x_n\| < \varepsilon \; .$$

Cauchy sequences

Convergent sequence

- A sequence $x_1, x_2, x_3, \ldots$ of real numbers is called convergent to a point $x^* \in \mathbb{R}$ if

$$\lim_{k \to \infty} |x_k - x^*| = 0 \; .$$

Cauchy sequence

- A sequence $x_1, x_2, x_3, \ldots \in \mathbb{R}^n$ is called a Cauchy sequence, if for every $\varepsilon > 0$, there exists $N < \infty$ such that:

$$\forall m, n > N, \qquad \|x_m - x_n\| < \varepsilon \; .$$

Convergence Theorems in $\mathbb{R}^n$

Theorem (Cauchy)

- Every Cauchy sequence in $\mathbb{R}^n$ converges to a $x^* \in \mathbb{R}^n$.

Theorem (Bolzano-Weierstrass)

- Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Convergence Theorems in $\mathbb{R}^n$

Theorem (Cauchy)

- Every Cauchy sequence in $\mathbb{R}^n$ converges to a $x^* \in \mathbb{R}^n$.

Theorem (Bolzano-Weierstrass)

- Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Contents

Hilbert space

The vector space H with inner product $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}$ is called a Hilbert space if for all $x, y \in H$ and all $a, b \in \mathbb{F}$:

1. Symmetry: $\langle y, x \rangle = \langle x, y \rangle$.
2. Linearity: $\langle a x_1 + b x_2, y \rangle = a \langle x_1, y \rangle + b \langle x_2, y \rangle$.
3. Positivity: $\langle x, x \rangle \geq 0$ such that $\|x\|_H = \sqrt{\langle x, x \rangle}$ is a norm.

Cauchy-Schwarz Inequality

In any Hilbert space we have

$$\langle x, y \rangle^2 \;\leq\; \langle x, x \rangle \langle y, y \rangle \;=\; \|x\|_H^2 \|y\|_H^2$$

Proof We may assume $y \neq 0$. Next,

$$\begin{aligned} \|x\|_H^2 &= \left\| \frac{\langle x, y \rangle}{\langle y, y \rangle} y + x - \frac{\langle x, y \rangle}{\langle y, y \rangle} y \right\|_H^2 \\ &= \left| \frac{\langle x, y \rangle}{\langle y, y \rangle} \right|^2 \|y\|_H^2 + \left\| x - \frac{\langle x, y \rangle}{\langle y, y \rangle} y \right\|_H^2 \;\geq\; \frac{\langle x, y \rangle^2}{\langle y, y \rangle} \end{aligned}$$

implies the Cauchy-Schwarz inequality.

Cauchy-Schwarz Inequality

In any Hilbert space we have

$$\langle x, y\rangle^2 \;\leq\; \langle x, x\rangle \langle y, y\rangle \;=\; \|x\|_H^2 \|y\|_H^2$$

Proof We may assume $y \neq 0$. Next,

$$\begin{aligned} \|x\|_H^2 \quad &= \quad \left\| \frac{\langle x, y\rangle}{\langle y, y\rangle} y + x - \frac{\langle x, y\rangle}{\langle y, y\rangle} y \right\|_H^2 \\ &= \quad \left| \frac{\langle x, y\rangle}{\langle y, y\rangle} \right|^2 \|y\|_H^2 + \left\| x - \frac{\langle x, y\rangle}{\langle y, y\rangle} y \right\|_H^2 \;\geq\; \frac{\langle x, y\rangle^2}{\langle y, y\rangle} \end{aligned}$$

implies the Cauchy-Schwarz inequality.

Most important examples

1. Euclidean space $H = \mathbb{R}^n$ with $\langle x, y \rangle = x^\intercal y$.

2. $H = L_2[-1, 1]$: infinite dimensional Hilbert space with

$$\langle f, g \rangle = \int_{-1}^{1} f(t) g(t) \, \mathrm{d}t \, .$$

Most important examples

1. Euclidean space $H = \mathbb{R}^n$ with $\langle x, y \rangle = x^{\mathsf{T}} y$.

2. $H = L_2[-1, 1]$: infinite dimensional Hilbert space with

$$\langle f, g \rangle = \int_{-1}^{1} f(t) g(t) \, \mathrm{d}t \ .$$

Gram-Schmidt Algorithm

Input: k elements $a_1, \ldots, a_k \in H$; $(H, \langle \cdot, \cdot \rangle)$ is a Hilbert space.

Gram-Schmidt Algorithm:

For $i = 1, \ldots, k$:

- Orthogonalization. $\overline{q}_i = a_i - \langle q_1, a_i \rangle q_1 - \ldots - \langle q_{i-1}, a_i \rangle q_{i-1}$.
- Test for dependence. If $\overline{q}_i = 0$, quit.
- Normalization. $q_i = \frac{\overline{q}_i}{\|\overline{q}_i\|_H}$.

If the algorithm does not quit, the vectors a_i are linearly independent.

Gram-Schmidt Algorithm

Input: k elements $a_1, \ldots, a_k \in H$; $(H, \langle \cdot, \cdot \rangle)$ is a Hilbert space.

Gram-Schmidt Algorithm:

For $i = 1, \ldots, k$:

- Orthogonalization. $\overline{q}_i = a_i - \langle q_1, a_i \rangle q_1 - \ldots - \langle q_{i-1}, a_i \rangle q_{i-1}$.
- Test for dependence. If $\overline{q}_i = 0$, quit.
- Normalization. $q_i = \frac{\overline{q}_i}{\|\overline{q}_i\|_H}$.

If the algorithm does not quit, the vectors a_i are linearly independent.

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1,1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1,1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1,1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1, 1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1,1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1,1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Gram-Schmidt Algorithm for Functions

Let $H = L_2[-1, 1]$ be the standard L_2-space.

Example: Start with $a_0(x) = 1$, $a_1(x) = x$, $a_2(x) = x^2$, ..., $a_n = x^n$:

- $q_0(x) = \sqrt{\frac{1}{2}}$.
- $q_1(x) = \sqrt{\frac{3}{2}}x$.
- $q_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$.
- ...
- $q_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\partial^n}{\partial x^n} \left(x^2 - 1\right)^n$. (Exercise)

Contents

Big O-notation

Landau's symbol "O":

- often used in complexity theory, computer science, and mathematics
- describes the asymptotic behavior of functions

Definition Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ be two given functions and $a \in \mathbb{R}$. We write

$$f(x) = \mathbf{O}(g(x)) \qquad \text{for} \quad x \to a$$

if and only if there exist constants $C_1, C_2 \in \mathbb{R}$ such that

$$\|f(x)\| \leq C_1 \|g(x)\| \qquad \text{for all } x \text{ with } \|x - a\| \leq C_2$$

Big O-notation

Remarks:

1. The "for $x \to a$" is skipped if it is clear from the context what a is.
2. In the scalar case, the notation is formally also used for $x \to \infty$:

$$f(x) = \mathbf{O}(g(x)) \qquad \text{for} \quad x \to \infty$$

   if and only if there exist constants $C_1, C_2 \in \mathbb{R}$ such that

$$\|f(x)\| \le C_1 \|g(x)\| \qquad \text{for all } x \text{ with } \; x \ge C_2$$

Big O-notation

Examples:

1. If the run-time of an algorithm for a problem of size n is $T(n) = 5n^2 + 6n + 1$, we write

$$T(n) = \mathbf{O}(n^2)$$

to say that $T(n)$ scales with n^2 for $n \to \infty$.

2. We write

$$e^x = 1 + x + \mathbf{O}(x^2)$$

assuming that it is clear from the context that we mean $x \to 0$.

Contents

Continuous functions

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at a point $a \in \mathbb{R}^n$ if for any convergent sequence $x_1, x_2, \ldots \in \mathbb{R}$ we have

$$\lim_{k\to\infty} x_k = a \qquad \text{implies} \qquad \lim_{k\to\infty} f(x_k) = f(a)\ .$$

If f is called continuous if it is continuous at all points $a \in \mathbb{R}^n$.

Theorem A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at a if and only if there exists for every $\epsilon > 0$ a $\delta > 0$ such that

$$\|x - a\| \leq \delta \qquad \text{implies} \qquad \|f(x) - f(a)\| < \epsilon\ .$$

Continuous functions

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at a point $a \in \mathbb{R}^n$ if for any convergent sequence $x_1, x_2, \ldots \in \mathbb{R}$ we have

$$\lim_{k\to\infty} x_k = a \qquad \text{implies} \qquad \lim_{k\to\infty} f(x_k) = f(a) \; .$$

If f is called continuous if it is continuous at all points $a \in \mathbb{R}^n$.

Theorem A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at a if and only if there exists for every $\epsilon > 0$ a $\delta > 0$ such that

$$\|x - a\| \leq \delta \qquad \text{implies} \qquad \|f(x) - f(a)\| < \epsilon \; .$$

Examples

1. Any norm, $f(x) = \|x\|$, is a continuous function.
2. For two continuous functions f, g with compatible dimensions their sum $f + g$ and product $f * g$ is continuous.
3. Polynomials are examples for continuous functions.

Exercise: write down formal proofs!

Examples

1. Any norm, $f(x) = \|x\|$, is a continuous function.
2. For two continuous functions f, g with compatible dimensions their sum $f + g$ and product $f * g$ is continuous.
3. Polynomials are examples for continuous functions.

Exercise: write down formal proofs!

Examples

1. Any norm, $f(x) = \|x\|$, is a continuous function.
2. For two continuous functions f, g with compatible dimensions their sum $f + g$ and product $f * g$ is continuous.
3. Polynomials are examples for continuous functions.

Exercise: write down formal proofs!

Extrema

Theorem (Weierstrass)

- If $D \subseteq \mathbb{R}^n$ is compact and $f : D \to \mathbb{R}$ a continuous function, then there exist points $x_{\text{min}} \in D$ and $x_{\text{max}} \in D$ with

$$f(x_{\text{max}}) = \sup_{x \in D} f(x) \qquad \text{and} \qquad f(x_{\text{min}}) = \inf_{x \in D} f(x) \; .$$

Uniformly continuous functions

Every continuous function is uniformly continuous on a compact domain; that is, there exists for every $\varepsilon > 0$ a $\delta > 0$ such that

$$\forall x, y \in D \text{ with } \|x - y\| < \delta \quad \text{we have} \quad \|f(x) - f(y)\| \le \varepsilon \ .$$

Uniform convergence

Definition

- A sequence of continuous functions $f_1, f_2, \ldots$ is said to converge uniformly on D if

$$\sup_{x \in D} \|f_k(x) - f(x)\| \to 0 \qquad (k \to \infty) .$$

Theorem

- If the continuous function sequence $f_1, f_2, \ldots$ converges uniformly on D, then the limit function f is continuous on D, too.

Uniform convergence

Definition

- A sequence of continuous functions $f_1, f_2, \ldots$ is said to converge uniformly on D if

$$\sup_{x \in D} \|f_k(x) - f(x)\| \to 0 \qquad (k \to \infty) .$$

Theorem

- If the continuous function sequence $f_1, f_2, \ldots$ converges uniformly on D, then the limit function f is continuous on D, too.

Lipschitz continuous functions

Definition

- A function $f : D \to \mathbb{R}^m$ is Lipschitz continuous on D, if there exists a $L < \infty$ with

$$\forall x, y \in D, \qquad \|f(x) - f(y)\| \leq L\|x - y\| \ .$$

Application:

- The contraction of so-called fixed point iterations, given by

$$x_{k+1} = g(x_k) \ ,$$

if often analyzed for Lipschitz continuous functions $g : \mathbb{R}^n \to \mathbb{R}^n$.

Lipschitz continuous functions

Definition

- A function $f : D \to \mathbb{R}^m$ is Lipschitz continuous on D, if there exists a $L < \infty$ with

$$\forall x, y \in D, \qquad \|f(x) - f(y)\| \leq L\|x - y\| \ .$$

Application:

- The contraction of so-called fixed point iterations, given by

$$x_{k+1} = g(x_k) \ ,$$

  if often analyzed for Lipschitz continuous functions $g : \mathbb{R}^n \to \mathbb{R}^n$.

Banach's fixed point theorem

Theorem If $g : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with Lipschitz constant $L < 1$, then the equation

$$g(x^*) = x^*$$

has a solution $x^* \in \mathbb{R}^n$ and the fixed point iteration

$$\forall k \in \mathbb{N}, \qquad x_{k+1} = g(x_k) \ ,$$

satisfies

$$\forall k \in \mathbb{N}, \qquad \|x_k - x^*\| \leq \frac{L^k}{1-L} \|x_1 - x_0\| \ .$$

Contents

Partial Derivatives

Definitions:

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is called partially differentiable at point $x \in \mathbb{R}^n$ in the i-th coordinate direction e_i if the limit

$$\frac{\partial f}{\partial x_i}(x) \;=\; \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}$$

  exists. If this limit exists for all x and all directions i, f is called partially differentiable.

- If the functions $\frac{\partial f}{\partial x_i}(x)$ are all continuous, f is called continuously differentiable.

Mixed second order derivatives

Theorem:

- If a function $f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable, then we have
$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(x) = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(x)$$
for all $x \in \mathbb{R}^n$ and all $i, j \in \{1, \ldots, n\}$.

Gradient Vector and Hessian Matrix

Gradient:

- The gradient of a scalar function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by

$$\nabla_x f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix} .$$

Hessian:

- The Hessian of a scalar function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by

$$\nabla_x^2 f(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \dots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \dots & \frac{\partial}{\partial x_n \partial x_n} f(x) \end{pmatrix} .$$

If f is twice continuously differentiable, $\nabla_x^2 f(x)$ is symmetric.

Jacobian Matrix

Jacobian:

- For a vector valued function $f : \mathbb{R}^n \to \mathbb{R}^m$ the Jacobian matrix is denoted by

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x) & \dots & \frac{\partial}{\partial x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x) & \dots & \frac{\partial}{\partial x_n} f_m(x) \end{pmatrix} .$$

  For scalar functions: $\nabla_x f(x) = \left(\frac{\mathrm{d}}{\mathrm{d}x} f(x) \right)^\mathsf{T}$.

  (don't forget the "transpose" !!!).

Directional derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable vector valued function.

Forward Differentiation

- For a given direction $\lambda \in \mathbb{R}^n$ the limit

$$\lim_{h \to 0} \frac{f(x + h\lambda)}{h} = \left(\frac{\mathrm{d}}{\mathrm{d}x} f(x) \right) * \lambda$$

  is called the (forward) directional derivative.

Backward Differentiation

- For a given direction $\mu \in \mathbb{R}^m$ the term

$$\mu^{\mathsf{T}} * \left(\frac{\mathrm{d}}{\mathrm{d}x} f(x) \right)$$

  is called the backward (or adjoint) directional derivative.

Directional derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable vector valued function.

Forward Differentiation

- For a given direction $\lambda \in \mathbb{R}^n$ the limit

$$\lim_{h \to 0} \frac{f(x + h\lambda)}{h} = \left(\frac{\mathrm{d}}{\mathrm{d}x} f(x) \right) * \lambda$$

is called the (forward) directional derivative.

Backward Differentiation

- For a given direction $\mu \in \mathbb{R}^m$ the term

$$\mu^{\mathsf{T}} * \left(\frac{\mathrm{d}}{\mathrm{d}x} f(x) \right)$$

is called the backward (or adjoint) directional derivative.

Mean Value Theorem

Scalar functions:

- If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then

$$\forall x, y \in \mathbb{R}^n, \quad f(y) - f(x) = \left\langle \left(\int_0^1 \nabla_x f(x + s(y - x)) \mathrm{d}s \right), y - x \right\rangle .$$

Vector-valued functions:

- If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, $J(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x)$, then

$$\forall x, y \in \mathbb{R}^n, \quad f(y) - f(x) = \left(\int_0^1 J(x + s(y - x)) \mathrm{d}s \right) (y - x) .$$

Mean Value Theorem

Scalar functions:

- If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then

$$\forall x, y \in \mathbb{R}^n, \quad f(y) - f(x) = \left\langle \left(\int_0^1 \nabla_x f(x + s(y - x)) \mathrm{d}s \right), y - x \right\rangle \, .$$

Vector-valued functions:

- If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, $J(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x)$, then

$$\forall x, y \in \mathbb{R}^n, \quad f(y) - f(x) = \left(\int_0^1 J(x + s(y - x)) \mathrm{d}s \right) (y - x) \, .$$

Implicit Function Theorem

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be a continuously differentiable function with $f(x^*, y^*) = 0$ and let the Jacobian

$$\frac{\partial f(x^*, y^*)}{\partial x}$$

be invertible. Then there exists an $\varepsilon > 0$ and a continuously differentiable function $g : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$f(g(y), y) = 0$$

for all $y \in \mathbb{R}^n$ with $\|y - y^*\| \leq \varepsilon$.

Implicit Function Theorem (continued)

The derivative of the continuously differentiable function $g : \mathbb{R}^m \to \mathbb{R}^n$ at $y = y^*$ can be worked out explicitly by using the equation

$$0 = \frac{\mathrm{d}}{\mathrm{d}y} f(g(y), y) = \frac{\partial f(x^*, y^*)}{\partial x} * \frac{\partial g(y^*)}{\partial y} + \frac{\partial f(x^*, y^*)}{\partial y} \ ,$$

which implies

$$\frac{\partial g(y^*)}{\partial y} = - \left[\frac{\partial f(x^*, y^*)}{\partial x} \right]^{-1} * \frac{\partial f(x^*, y^*)}{\partial y} \ .$$

Contents

Scalar Taylor Expansions

Let $f : \mathbb{R} \to \mathbb{R}$ be a $(r+1)$-times continuously differentiable function.

- Taylor series

$$f(x+h) = \sum_{k=0}^{r} \frac{f^{(k)}(x)}{k!} h^k + R(x,h)$$

- Remainder term in differential form:

$$R(x,h) = \frac{f^{(r+1)}(x+\theta h)}{(r+1)!} h^{r+1} \quad \text{for a} \quad \theta \in [0,1]$$

- Remainder term in integral form:

$$R(x,h) = \frac{h^{r+1}}{r!} \int_0^1 f^{(r+1)}(x+sh)(1-s)^r \, \mathrm{d}s \; .$$

Multi-Index Notation

Definition

- A tuple $(\alpha_1, \ldots, \alpha_n)$ with $\alpha_i \in \mathbb{N}$ is called a multi-index.
- The order/factorial of a multi-index are denoted by

$$|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n \quad \text{and} \quad \alpha! = \alpha_1! * \ldots * \alpha_n! \, .$$

Example: for the case $n = 2$:

$$\sum_{|\alpha|=2} \frac{\partial^{|\alpha|}}{\partial x^\alpha} f(x) = \frac{\partial^2}{\partial x_1^2} f(x) + \frac{\partial^2}{\partial x_2^2} f(x) + \frac{\partial^2}{\partial x_1 \partial x_2} f(x)$$

Multi-Index Notation

Definition

- A tuple $(\alpha_1, \ldots, \alpha_n)$ with $\alpha_i \in \mathbb{N}$ is called a multi-index.
- The order/factorial of a multi-index are denoted by

$$|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n \quad \text{and} \quad \alpha! = \alpha_1! * \ldots * \alpha_n! \,.$$

Example: for the case $n = 2$:

$$\sum_{|\alpha|=2} \frac{\partial^{|\alpha|}}{\partial x^\alpha} f(x) = \frac{\partial^2}{\partial x_1^2} f(x) + \frac{\partial^2}{\partial x_2^2} f(x) + \frac{\partial^2}{\partial x_1 \partial x_2} f(x)$$

General Taylor Expansions

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a $(r+1)$-times continuously differentiable function.

- General Taylor series with $x, h \in \mathbb{R}^n$, $h^\alpha = h_1^{\alpha_1} * \ldots * h_n^{\alpha_n}$,

$$f(x+h) = \sum_{k=0}^{r} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f(x)}{\partial x^\alpha} h^\alpha + \mathbf{O}(\|h\|^{r+1})$$

- Important example for $m = 1$ and $r = 2$:

$$f(x+h) = f(x) + \nabla_x f(x)^\intercal h + \frac{1}{2} h^\intercal \nabla_x^2 f(x) h + \mathbf{O}(\|h\|^3) \,.$$

General Taylor Expansions

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a $(r+1)$-times continuously differentiable function.

- General Taylor series with $x, h \in \mathbb{R}^n$, $h^\alpha = h_1^{\alpha_1} * \ldots * h_n^{\alpha_n}$,

$$f(x+h) = \sum_{k=0}^{r} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f(x)}{\partial x^\alpha} h^\alpha + \mathbf{O}(\|h\|^{r+1})$$

- Important example for $m = 1$ and $r = 2$:

$$f(x+h) = f(x) + \nabla_x f(x)^\intercal h + \frac{1}{2} h^\intercal \nabla_x^2 f(x) h + \mathbf{O}(\|h\|^3)\ .$$