Analysis 2, Chapter 5

Matthijs Muis

December 2023

1 Differentiability

1.1 Definition of differentiability and the differential

Recall the definition of differentiability in $\mathbb{R} \to \mathbb{R}$:

Definition 1. *Let $\Omega \subset \mathbb{R}$ **open**, then $f : \Omega \to \mathbb{R}$ is called **differentiable at** a if*

$$L = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists in $\mathbb{R}$.

In the scalar case, L is simply a scalar. It is also a **linear map** $L : \mathbb{R} \to \mathbb{R}$ (namely, scalar multiplication with L). This concept is necessary to generalize differentiation to arbitrary **normed vector spaces**. We will discuss it for $\mathbb{R}^N$:

Definition 2. *Let $\Omega \subset \mathbb{R}^M$ **open**, $f : \Omega \to \mathbb{R}^N$ is called **differentiable at** a if there is an $L \in \mathcal{L}(\mathbb{R}^M, \mathbb{R}^N)$ such that:*

$$\lim_{x \to a} \frac{f(x) - (f(a) + L(x-a))}{|x-a|} = 0$$

Notation 1. *In Landau's notation, we say $f = g + o(h)$ for functions $f, g, h : \mathbb{R} \to \mathbb{R}$, if*

$$\lim_{t \to 0} \frac{f(t) - g(t)}{h(t)} = 0$$

*Using this notation, **differentiability writes as***

$$f(x) = f(a) + L(x-a) + o(|x-a|)$$

Notation 2. *We denote the **punctured ball** $B_r^0(x)$ around $x \in X$ with radius $r >$ to be the set*

$$B_r^0(x) = \{y \in X : 0 < d(x,y) < r\}$$

Proposition 1. *If L exists, it is **unique**.*

Proof. Assume for a contradiction that there are two different linear maps, T and L. This implies that there is an $\epsilon > 0$ and $(x_n)_{n\in\mathbb{N}}$ is a sequence such that

$$\forall n \in \mathbb{N} : x_n \in B^0_{\frac{1}{n}}(a) \text{ yet } |L(x_n - a) - T(x_n - a)| \geq \epsilon$$

Clearly, $x_n \to a$ as $n \to \infty$. Moreover, it implies:

$$\forall n \in \mathbb{N} : \frac{|L(x_n - a) - T(x_n - a)|}{|x_n - a|} \geq \epsilon n$$

Now, since

$$\left| \frac{f(x_n) - f(a) - L(x_n - a)}{|x_n - a|} - \frac{f(x_n) - f(a) - T(x_n - a)}{|x_n - a|} \right| = \left| \frac{|L(x_n - a) - T(x_n - a)|}{|x_n - a|} \right|$$

It follows that this sequence is unbounded. On the other hand, by differentiability, both $\frac{f(x_n)-f(a)-L(x_n-a)}{|x_n-a|}$ and $\frac{f(x_n)-f(a)-T(x_n-a)}{|x_n-a|}$ converge to 0, so their sum sequence should converge and thus be bounded. A contradiction. □

Definition 3. *We call the unique linear map the **differential of** f **at** a. Notation: $D_a f = L$*

Proposition 2. *If $f : \Omega \to \mathbb{R}^N$ is differentiable at x, then there is an $M > 0$ and $\delta > 0$ such that*

$$\forall y \in B_\delta(x) : |f(x) - f(y)| \leq M|x - y|$$

*In particular, f is **Lipschitz continuous** ($\implies$ **uniformly continuous**) in an **open neighbourhood** $B_\delta(a)$.*

Proposition 3. *For $\Omega \subset \mathbb{R}^N$ **open** and $a \in \Omega$, $\lambda \in \mathbb{R}$ and $f, g : \Omega \to \mathbb{R}$, **differentiable at** a, we have:*

- *$f + \lambda g$ is **differentiable** at a and $D_a(f + \lambda g) = D_a f + \lambda D_a g$*
- *$f \cdot g$ is **differentiable** at a and $D_a(f \cdot g) = f \cdot D_a g + g \cdot D_a f$*

Note that here, the dimension of the target space is $N = 1$.

Proof.
- These equalities follow because we can interchange limits with linear combinations and products, provided the individual terms form convergent sequences:

$$\lim_{x\to a} \frac{f(x) + \lambda g(x) - f(a) - \lambda g(a) - (D_a f)(x - a) - \lambda(D_a g)(x - a)}{|x - a|} =$$

$$\lim_{x\to a} \frac{f(x) - f(a) - (D_a f)(x - a)}{|x - a|} + \lambda \cdot \lim_{x\to a} \frac{g(x) - g(a) - (D_a g)(x - a) -}{|x - a|} =$$

$$0 + \lambda \cdot 0 = 0$$

- Here, we not only need the lemma that we can take linear combinations of limiting sequences, but also that f and g are continuous (Proposition 2). Also, assume without loss of gne

$$\frac{f(x)g(x) - f(a)g(a) - g(a)(D_a f)(x-a) - f(a)(D_a g)(x-a)}{|x-a|} =$$
$$\frac{f(x)g(x) + f(x)g(a) - f(x)g(a) - f(a)g(a) - f(a)(D_a f)(x-a) - g(a)(D_a g)(x-a)}{|x-a|} =$$
$$\frac{f(x)g(x) + f(x)g(a) - f(x)g(a) - f(a)g(a) - f(a)(D_a f)(x-a) - g(a)(D_a g)(x-a)}{|x-a|} =$$
$$\frac{f(x)g(x) - f(x)g(a) - f(x)(D_a f)(x-a)}{|x-a|} - g(a)\frac{f(x) - f(a) - (D_a f)(x-a)}{|x-a|}$$

By differentiability of f,

$$\lim_{x\to a} \frac{f(x) - f(a) - (D_a f)(x-a)}{|x-a|} = 0$$

So it suffices to show, that for

$$\Delta(r) := \sup_{x\in B_r^0(a)} \frac{|f(x)g(x) - f(x)g(a) - f(x)(D_a f)(x-a)|}{|x-a|}, \text{ we have } \lim_{r\to 0} \Delta(r) = 0$$

With Proposition 2, we can already say that for $r < \delta$, it holds

$|f(x) - f(a)| \leq M|x-a|$, therefore

$$\Delta(r) \leq \sup_{x\in B_r^0(a)} \left(\frac{|f(a)g(x) - f(a)g(a) - f(a)(D_a f)(x-a)|}{|x-a|} \right) + \sup_{x\in B_r^0(a)} (M|g(x) - g(a)|)$$

□

1.2 Chain Rule

Proposition 4. *Let $\Omega \subset \mathbb{R}^N$ and $U \subset \mathbb{R}^k$ be **open**, $f : \Omega \to \mathbb{R}^M$ **differentiable at** $a \in \Omega$ and $\varphi : U \to \mathbb{R}^k$ **differentiable at** $f(a) \in U$. Then $\varphi \circ f : \Omega \to \mathbb{R}$ is differentiable at a and*

$$D_a(\varphi \circ f) = D_{f(a)}\varphi \circ D_a f$$

Note that the right-hand-side of the equality is indeed a linear map, as it is the composition of two linear maps.

Remark 1. *This means that $\forall x \in \Omega$,*

$$\varphi \circ f(x) = \varphi \circ f(a) + (D_{f(a)}\varphi \circ D_a f)(x-a) = o(|x-a|)$$

1.3 Gradient, Partial Derivatives

Notation 3. *For the unfamiliar reader, if V is a K-vector space we will make use the* ***dual vector space****, denoted as $V^* = \mathcal{L}(V, K)$.*

Definition 4. *For a linear map $L \in V^*$, there is a* ***unique*** *$w \in V$ such that*

$$L(v) = \langle w, v \rangle$$

If the linear map is $D_x f$, where $f : \Omega \to \mathbb{R}$ is differentiable, we call this w the ***gradient*** *and denote it as $\nabla f(x)$*

Remark 2. *By linearity of $D_x f$ in f, and bilinearity of the inner product, we also have this for $\nabla f(x)$ as a vector:*

$$\nabla(\lambda f + g)(x) = \lambda \nabla f(x) + \nabla g(x)$$
$$\nabla(f \cdot g)(x) = f(x) \cdot \nabla g(x) + g(x) \cdot \nabla f(x)$$

Definition 5. *For $f : \Omega \to \mathbb{R}$, we say that f has directional derivatives at $x \in \Omega$ in the direction v if*

$$\lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}$$

exists and is finite. We denote it $\partial_v f(x)$

Remark 3. *Like with* ***linear continuity****, the existence of directional derivatives is a statement about the behaviour of functions* ***along straight lines.*** *In particular, the existence of partial derivatives of f at x is not at all sufficient for differentiability at x, because f can still behave badly along arbitrary curves. Consider:*

$$f : \mathbb{R}^2 \to \mathbb{R},$$
$$f(x, y) = \begin{cases} \frac{x|y|}{\sqrt{x^2+y^2}} & (x, y) \neq 0 \\ 0 & (x, y) = 0 \end{cases}$$

It has directional derivatives at the origin along every line $\{tv : t \in \mathbb{R}\}$, yet $v \mapsto \partial_v f(0, 0)$ is not linear, while this should be the case according to Proposition 9.

Proposition 5. ***Lagrange's Mean Value Theorem***

Let $f : \Omega \to \mathbb{R}$ be a function and $x, y \in \Omega$. Let $v = y - x$ and assume $S = \{x + tv : t \in [0, 1]\} \subset \Omega$, and that f ***has directional derivatives*** *(does* ***not*** *need to be* ***differentiable****) at all the points of S. Then:*

$$f(y) - f(x) = \partial_v f(z)$$

or also

$$f(y) - f(x) = \langle \nabla f(z), y - x \rangle$$

for some $z \in S$

A more general form of the theorem requires us to define a $C^1(\Omega)$ curve:

Definition 6. *A function $f : \Omega \to \mathbb{R}^M$ is called* ***of class*** $C^r(\Omega)$ *if it has partial derivatives up to and including order r in directions of all basis vectors $e_1, .., e_N$ at all $x \in \Omega$ and these partial derivatives $x \mapsto \partial^\alpha f(x)$ are continuous with respect to x.*

Remark 4. *We then often denote $\partial_{e_i} f$ simply as ∂_i. (Like in quantum mechanics, where one often just denotes the n-th eigenstate ψ_n as $|n\rangle$.*

Proposition 6. *$f \in C^1(\Omega)$ and $\gamma : [0,1] \to \Omega$ s.t. $\gamma_i \in C^1([0,1])$. Then*

$$f(y) - f(x) = \int_0^1 (D_{\gamma(t)} f)(\gamma'(t)) dt$$

As a corollary we have the estimate for **straight line segments** $S = x + [0,1]v$:

Proposition 7.

$$|f(y) - f(x)| \leq |x - y| \sup_{z \in S} |\nabla f(z)|$$

Proposition 8. *Let $\Omega \subset \mathbb{R}^M$ be open and $f : \Omega \to \mathbb{R}^M$, where we write $f = (f_1, .., f_M)$.*

Then f is ***differentiable at*** *$x \in \Omega$ if and only if each of its components $f_i : \Omega \to \mathbb{R}$ is differentiable in $x \in \Omega$.*

Now is the time to investigate the relation between differentiability and the **existence** and **regularity** of partial derivatives. First, we discuss what differentiability implies for the partial derivatives, and then we discuss that continuous partial derivatives imply differentiability.

Proposition 9. *Let $f : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^N$ is* ***open****. If f is* ***differentiable*** *at $x \in \Omega$, then*

(i) f ***has directional derivatives*** *at x in every direction $v \in \mathbb{R}^N \backslash \{0\}$*

(ii) $\partial_v f(x) = (D_x f)(v)$, or equivalently (by definition of $\nabla f(x)$ through identification by duality)

(iii) $\partial_v f(x) = \langle \nabla f(x), v \rangle$

(iv) $v \mapsto \partial_v f(x)$ is ***linear****.*

In particular, we can conclude that if $v \mapsto \partial_v f(x)$ is **not linear** at some $x \in \Omega$, then f is **not differentiabl**e at x.

Remark 5. *However, existence of partial derivatives in all directions v and their linearity in v is not sufficient either: consider the function f defined as:*

$$f(x,y) = \begin{cases} 1 & \textit{if } y = x^2 \neq 0 \\ 0 & \textit{else} \end{cases}$$

Then f has directional derivatives in all directions at the origin, since

$$\frac{f(tv_1, tv_2) - f(0,0)}{t} = \frac{0-0}{t} = 0, \quad \forall t \neq 0$$

And in particular, the directional derivatives are all 0, hence linear in v. Yet f is not even continuous at the origin!

In general, we have the strict inclusion

partial derivatives $\supset$ linear partial derivatives $\supset$ differentiable $\supset$ of class C^1

The final $\supset$ is called the **Total Differential theorem**

Theorem 1. ***Total differential theorem***
*Let $f : \Omega \to \mathbb{R}$ be **open** and let $f \in C^1(\Omega)$. Then, f is **differentiable** at each point of Ω and moreover, we have:*

$$\forall z \in \Omega : \lim_{x \neq y \to z} \frac{f(y) - f(x) - D_z f(y-x)}{|x-y|} = 0$$

*Note that we could also have used the **gradient** notation, which is equivalent by definition:*

$$\forall z \in \Omega : \lim_{x \neq y \to z} \frac{f(y) - f(x) - \langle \nabla f(z), (y-x) \rangle}{|x-y|} = 0$$

*Moreover, the gradient **can be directly computed** as*

$$\nabla f(x) = (\partial_1 f(x), ..., \partial_N f(x))$$

So this theorem is really about showing the following:

$$\forall z \in \Omega : \lim_{x \neq y \to z} \frac{f(y) - f(x) - \sum_{i=1}^N \partial_i f(z) \cdot (y_i - x_i)}{|x-y|} = 0$$

Definition 7. *Ifr $f : \Omega \to \mathbb{R}^M$ is differentiable at $x \in \Omega$, where $\Omega \subset \mathbb{R}^M$ is open, and we denote $e_1, ..., e_N$ the standard basis vectors of $\mathbb{R}^N$, we define the matrix isomorphism*

$$\mathsf{mat} : L(\mathbb{R}^N, \mathbb{R}^M) \to \mathbb{R}^{M \times N}$$

*Which identifies a **linear map** $L : \mathbb{R}^N \to \mathbb{R}^M$ with a **matrix** $M \in \mathbb{R}^{M \times N}$ where*

$$M_{ij} = \langle e_i, L e_j \rangle$$

*With this, we can define the **Jacobi matrix** via mat as follows:*

$$J_x f = \mathsf{mat}(D_x f)$$

Proposition 10.

$$J_x f = \begin{pmatrix} \nabla f_1(x)^T \\ \nabla f_2(x)^T \\ ... \\ \nabla f_N(x)^T \end{pmatrix}$$

Definition 8. *for $a \in \mathbb{R}^N$ and $b \in \mathbb{R}^N$, define the **tensor product** (here: simply matrix outer product) $a \otimes b$ as the $M \times N$ matrix:*

$$(a \otimes b)_{ij} = a_i b_j$$

Remark 6. *In general, a **rank** (r, s) **tensor** is a **multilinear form** $T : (V^*)^r \times V^s \to K$ where V is an K-vector space and V^* is its dual vector space. for S another tensor of rank (t, u), we define its **tensor product** $T \otimes S$ as the **unique rank** $r + t, s + u)$ **tensor** such that:*

$$(T \otimes S)(v^1..v^r, w^1..w^t, x_1..x_s, y_1..y_u) = T(v^1..v^r, x_1..x_s)S(w^1..w^t, y_1..y_u)$$

Proposition 11. ***Product rule for scalar multiplication***
For $\varphi : \mathbb{R}^N \to \mathbb{R}$, $f : \mathbb{R}^N \to R^M$ differentiable functions at a point $x \in R^N$, we have that

$$\varphi \cdot f : \mathbb{R}^N \to R^M, \text{ defined as } x \mapsto \varphi(x) \cdot f(x)$$

is differentiable at x and:

$$J_x(\varphi \cdot f) = \nabla\varphi(x) \otimes f + \varphi \cdot (J_x f)$$

1.4 Leibniz' integration theorem

Proposition 12. ***Leibniz' Integral Rule I***
*For $\psi : [a, b] \times \Omega \to \mathbb{R}$ a **continuous** function where $\Omega \subset \mathbb{R}^N$ is **open**:*

(i) The function $\phi : \Omega \to \mathbb{R}$ defined as

$$\psi(x) = \int_a^b \psi(t, x)dt$$

*is **continuous**.*

*(ii) If $v \in \mathbb{R}^N \backslash \{0\}$ and $\partial_v\psi$ is continuous on $[a, b] \times \Omega$, then the **Leibniz' formula** holds:*

$$\partial_v\phi(x) = \int_a^b \partial_v\psi(t, x)dt$$

Proposition 13. ***Leibniz' Integral Rule II***
*If $a, b \in C^0(\Omega; [c, d])$ and $\psi : [c, d] \times \Omega \to \mathbb{R}$ is a **continuous** function, where $\Omega \subset \mathbb{R}^N$ is **open**.*

(iii) Then $\phi : \Omega \to \mathbb{R}$, defined through:

$$\phi(x) = \int_{a(x)}^{b(x)} \psi(t, x)dt$$

*is **continuous**.*

(iv) *Moreover, if $a, b \in C^1(\Omega; [c,d])$ and for some $v \in \mathbb{R}^N \backslash \{0\}$, we have that $\partial_v \phi$ is* ***continuous*** *on $[c,d] \times \Omega$, then*

$$\begin{aligned}\partial_v \phi(x) = &\int_{a(x)}^{b(x)} \partial_v \psi(t,x) dt \\ &+ \psi(b(x), x)\partial_v b(x) \\ &- \psi(a(x), x)\partial_v a(x)\end{aligned}$$

Proof. The proof is for the **Riemann integral**. There is a similar theorem for the Lebesgue integral, but it requires a definition limiting theorems for the Lebesgue integral, which will follow in Chapter 11, 12 & 13.

(i) Let $x_n \to x$, for $(x_n)_{n\in\mathbb{N}}$ in Ω and fix $\epsilon > 0$. We need to prove there is a $N \in \mathbb{N}$ with, up to a constant factor,

$$\forall n \geq N : |\phi(x_n) - \phi(x)| < \epsilon$$

Since Ω is open, find a $R > 0$ with $\overline{B_R(x)} \subset \Omega$ and conclude that ψ is uniformly continuous on $[a,b] \times \overline{B_R(x)}$, by compactness of this set. This means that there exists a $\delta > 0$ with:

$$\forall u, v \in [a,b],\ \forall y, z \in \overline{B_R(x)} : |\psi(u,y) - \psi(v,z)| < \epsilon$$

Therefore, let $N \in \mathbb{N}$ be such that $\forall n \geq N : |x_n - x| < \delta$. Then, for all $n \geq N$:

$$\begin{aligned}|\phi(x_n) - \phi(x)| &= \left| \int_a^b [\psi(t,x_n) - \psi(t,x)] dt \right| \leq \int_a^b |\psi(t,x_n) - \psi(t,x)| dt \\ &\leq |a-b| \cdot \sup_{t\in[a,b] y \in B_\delta(x)} |\psi(t,y) - \psi(t,x)| \\ &< \delta |a-b|\end{aligned}$$

(ii) We need to show

$$\lim_{t\to 0} \frac{\phi(x) - \phi(x+hv)}{h} = \int_a^b \partial_v \psi(t,x) dt$$

By openness of Ω, for $|h|$ sufficiently small, we have $x + hv \in \Omega$, so we reason with that. On the compact neighbourhood $[a,b] \times \overline{B_r(x)}$, we have that $\partial_v \phi(t,v)$ is **uniformly continuous**, which will be useful: for all $\epsilon > 0$ there is a $\delta > 0$ with

$$\begin{gathered}\forall s, t \in [a,b], \forall h, l \text{ such that } x + hv,\ x + lv \in \Omega : \\ |s - t| < \delta,\ |h - l| < \frac{\delta}{|v|} \implies |\phi(t, x+hv) - \phi(s, x+lv)| < \epsilon\end{gathered}$$

$$\frac{\phi(x+hv)-\phi(x)}{h} = \frac{1}{h}\int_a^b [\psi(t,x+hv)-\psi(t,x)]dt$$

By Lagrange's Mean Value Theorem, applied to $l \mapsto \phi(t, x+lhv)$, $[0,1] \to \mathbb{R}$, there exists, for each $t \in [a,b]$, h as above, a $\lambda_t \in [0,1]$ such that

$$\psi(t,x+hv)-\psi(t,x) = h\cdot(\partial_v\psi)(t,x+\lambda_t hv)$$

Therefore, we rewrite:

$$\begin{aligned}|\psi(t,x+hv)-\psi(t,x)-h\partial_v\psi(t,x)| &= |h\cdot(\partial_v\psi)(t,x+\lambda_t hv)-h\cdot\partial_v\psi(t,x)|\\ &= |h||\partial_v\psi(t,x+\lambda_t hv)-\partial_v\psi(t,x)| \leq \epsilon|h|\end{aligned}$$

Namely, the latter equality holds for sufficiently small $|h| < \delta$. This is sufficient to derive the required approximation: for $|h| < \delta$, we have:

$$\begin{aligned}\left|\frac{\phi(x+hv)-\phi(x)}{h} - \int_a^b \partial_v\phi(t,x)dt\right| &= \left|\frac{1}{h}\int_a^b [\psi(t,x+hv)-\psi(t,x)-h\partial_v\phi(t,x)]dt\right|\\ &\leq \frac{1}{|h|}|a-b| \sup_{t\in[a,b],|h|<\delta} [\psi(t,x+hv)-\psi(t,x)-h\partial_v\phi(t,x)]\\ &\leq |a-b|\frac{|h|}{|h|}\epsilon = \epsilon|a-b|\end{aligned}$$

This was to be shown.

□

1.5 Higher order differentiability and Taylor expansion

Definition 9. *$N \in \mathbb{N}$. An N-**multi-index** α is an element of $\mathbb{N}^N$. For this element, we define:*

$$|\alpha| := \sum_{i=1}^N \alpha_i \qquad\qquad \alpha! = \prod_{i=1}^k \alpha_i!$$

Moreover, for $x \in \mathbb{R}^N$, we define

$$x^\alpha := \prod_{i=1}^N x_i^{\alpha_i}$$

Definition 10. *For $k \in \mathbb{N}$, $\Omega \subset \mathbb{R}^N$ open. $f \in C^k(\Omega, \mathbb{R}^N)$. $\alpha \in \mathbb{N}^N$ with $|\alpha| \leq k$. Then we define*

$$\partial^\alpha f(x) := \partial_{x_i}^{\alpha_1}..\partial_{x_N}^{\alpha_N} f(x)$$

where we define $\partial_{x_i}^{\alpha_i} g(x) := \partial_{x_i}..\partial_{x_i} h(x)$, i.e. α_i times.

Theorem 2. ***Taylor Series with Peano Remainder***

*Let $k \in \mathbb{N}$, and $\Omega \subset \mathbb{R}^N$ **open**, and fix $x \in \Omega$. Assume that $\partial^\alpha f(x)$ exists for all $\alpha \in \mathbb{N}^N$ with $|\alpha| \leq k$.*

*Then fix $x, y \in \Omega$ and assume that the **straight line segment** $S = x + [0,1](y-x)$ joining them is contained in Ω. Then, the following approximation holds:*

$$f(y) = \sum_{\alpha \in \mathbb{N}^N : |\alpha| \leq k-1} \frac{1}{\alpha!} \partial^\alpha f(x)(y-x)^\alpha + o(|y-x|^k)$$

Theorem 3. ***Taylor Series with Integral Remainder***

$k \in \mathbb{N}$, $\Omega \subset \mathbb{R}^N$ open, $x \in \Omega$ fixed. Assume that $\partial^\alpha f(x)$ exists for all $\alpha \in \mathbb{N}^N$ with $|\alpha| \leq k+1$.

Then fix $x, y \in \Omega$ and assume that the line segment $S = x + [0,1](y-x)$ joining them is contained in Ω. The following approximation holds:

$$f(y) = \sum_{\alpha \in \mathbb{N}^N : |\alpha| \leq k} \frac{1}{\alpha!} \partial^\alpha f(x)(y-x)^\alpha + R_k(x,y)$$

*Where the **remainder term** $R_k(x,y)$ satisfies:*

$$R_k(x,y) = \sum_{\alpha \in \mathbb{N}^N : |\alpha| = k+1} \frac{(y-x)^\alpha}{\alpha!} \int_0^1 (k+1)(1-t)^k \partial^\alpha f(x + t(y-x))dt$$

Remark 7. *For $N = 1$, this gives the familiar **Lagrange theorem** with integral remainder:*

$$f(y) = \sum_{i=0}^{k} \frac{1}{i!} f^{(i)}(x)(y-x)^i + R_k(x,y)$$

where

$$\begin{aligned} R_k(x,y) &= \frac{(y-x)^{k+1}}{(k+1)!} \int_0^1 (k+1)(1-t)^k f^{(k+1)}(x + t(y-x))dt \\ &= \frac{(y-x)^{k+1}}{(k+1)!} \int_x^y (k+1) \left(1 - \frac{u-x}{y-x}\right)^k f^{(k+1)}(u) \left(\frac{u-x}{y-x}\right) du \\ &= \frac{(y-x)^{k+1}}{(k+1)!} \int_x^y (k+1) \left(\frac{y-u}{y-x}\right)^k \left(\frac{u-x}{y-x}\right) f^{(k+1)}(u)du \\ &= \frac{1}{k!} \int_x^y (y-u)^k \left(\frac{u-x}{y-x}\right) f^{(k+1)}(u)du \end{aligned}$$

Definition 11. *If a function $f : \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^N$ is of $C^\infty(\Omega)$, then its **Taylor series** $T(f)(a)$ at $a \in \Omega$ is defined as:*

$$T(f)(a)(x) := \sum_{k=0}^{\infty} \sum_{\alpha \in \mathbb{N}^N : |\alpha| = k} \frac{1}{\alpha!} \partial^\alpha f(a)(x-a)^\alpha$$

Note*: there is no natural order in which to enumerate the terms $(x-a)^\alpha$, and therefore, we adopt no ordering convention, This means that this series is only defined well if it converges* ***absolutely*** *(in that case, permutations of the terms do not change the convergence properties); moreover, not every $C^\infty(\Omega)$-function equals its Taylor series. We call function $f \in C^\infty(\Omega)$* ***analytic*** *at $a \in \Omega$ if that is the case:*

$$T(f)(a) = f$$

Example 1. *Consider $f : \mathbb{R} \to \mathbb{R}$, defined through*

$$f(x) := \begin{cases} e^{-\frac{1}{x^2}} & x > 0 \\ 0 & x \leq 0 \end{cases} \implies f^{(n)}(x) = \begin{cases} 2 \cdot \left(\sum_{k=3}^{n+2} \frac{(-1)^k}{x^k} \right) e^{-\frac{1}{x^2}} & x > 0 \\ 0 & x < 0 \\ 0 & x = 0 \end{cases}$$

This writes as:

$$f^{(n)}(x) = \begin{cases} 2 \cdot \left(\sum_{k=3}^{n+2} \frac{(-1)^k}{x^k} \right) f(x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Which is clearly continuous, also at the origin as $f \to 0$ faster than any polynomial. However, at $a = 0$, the Taylor-series is 0, which is not equal to f in any open neighbourhood of 0, let alone at $\mathbb{R}$.

2 Homework

Exercise 1 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$f(x, y) := \begin{cases} \frac{x|y|}{\sqrt{x^2+y^2}} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

Prove that:

(i) f has directional derivative in all directions at the origin;

(ii) The map

$$v \mapsto \partial_v f(0, 0)$$

is not linear. In particular, prove that f is not differentiable at the origin.

Proof

(i) let $v = (v_x, v_y) \in \mathbb{R}^2 \backslash \{0\}$. Then

$$\frac{f((0,0) + t(v_x, v_y)) - f(0,0)}{t} = \frac{t|t||v_x||v_y|}{t|t|\sqrt{v_x^2 + v_y^2}}$$

And this is a constant for any t, hence goes to $\frac{v_x|v_y|}{\sqrt{v_x^2+v_y^2}}$ as $t \to 0$. So partial derivatives exist in all directions and are calculated to be:

$$\partial_v f(0,0) = \frac{v_x|v_y|}{\sqrt{v_x^2 + v_y^2}}$$

(ii) The map

$$v \mapsto \partial_v f(0,0) = \frac{v_x|v_y|}{\sqrt{v_x^2 + v_y^2}}$$

is clearly not linear: $\partial_{(0,1)} f(0,0) = 0$ $\partial_{(1,0)} f(0,0) = 0$ but $(1,1) = (0,1), +(1,0)$, yet $\partial_{(1,1)} f(0,0) 1/\sqrt{2} \neq 0+0 = \partial_{(0,1)} f(0,0) + \partial_{(1,0)} f(0,0)$ In particular, we see f is not differentiable at the origin, since then it follows $\partial_v f(0,0) = \langle \nabla f(0,0), v \rangle$, which would be linear in v.

Exercise 2 In this exercise we want to compute the derivative of the determinant. We will do that by seeing the determinant as a **multilinear map of the columns**. Since, recall, the determinant is the unique mapping $L : \mathbb{R}^{N\times N} \to \mathbb{R}$ that is **alternating**, **N-linear on the columns (rows)**, and has $L\text{Id} = 1$. We will only make use of the second property, and maybe the third.

(i) Consider an N-linear map $L : (\mathbb{R}^N)^N \to \mathbb{R}$. Namely, a map such that for all $i \in \{1, ..., N\}$, all $V = (v1, ..., vn) \in (R^N)^N$, all $w \in \mathbb{R}^N$, and all $\lambda \in \mathbb{R}$ it holds

$$L(v_1, ..., v_i + \lambda w, ..., v_N) = L(v_1, ..., v_i, ..., v_N) + \lambda L(v_1, ..., w, ..., v_N)$$

. Fix $V = (v_1, ..., v_N)$ and $W = (w_1, ..., w_N)$, where $v_i, w_i \in \mathbb{R}^N$ for each $i = 1, ..., N$. Compute the **directional derivative of L at V** in the **direction W** . Namely, compute

$$\lim_{t\to 0} \frac{L(V + tW) - L(V)}{t} = \lim_{t\to 0} \frac{L(v_1 + tw_1, ..., v_N + tw_N) - L(v_1, ..., v_N)}{t}$$

Solution

(i) We will denote the matrix arising when the i-th column of V is replaced by the i-th column of W, i.e. $(v_1, ...v_{i-1}, w_i, v_{i+1}, ..., v_N)$, with V^iW. With this, we prove by induction that

$$\partial_W L(V) = \sum_{i=1}^{N} L(V^iW)$$

Induction Basis: for $N = 1$, L is just a linear map. L not only has partial derivatives, but even a total derivative, as can be seen from the fact that

$$\lim_{x\to v} \frac{L(x) - L(v) - T(x - v)}{|x - v|} = 0 \text{ trivially holds for } T = L$$

From this, we can conclude $D_v L = L$, and therefore $\partial_w L(v) = Lw = L(V^1 W)$ for $V = (v)$, $W = (w)$.

Induction Hypothesis: suppose that for $n-1$-linear maps $L : (\mathbb{R}^N)^{n-1} \to \mathbb{R}$ we have the result that for any two sequences of column vectors $V = (v_1, ..., v_{n-1})$, $W = (w_1, ..., w_n)$, the partial derivative $\partial_W L(V)$ exists and equals $\sum_{i=1}^n L(V^i W)$.

Induction Step Then, let $L : (\mathbb{R}^N)^n \to \mathbb{R}$. Consider, for $t \neq 0$,

$$\frac{L(v_1 + tw_1, ..., v_n + tw_n) - L(v_1, ..., v_n)}{t}$$

We use linearity in the N-th component to rewrite this as

$$\frac{L(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, v_n) + tL(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, w_n) - L(v_1, ..., v_n)}{t}$$

$$= \frac{L(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, v_n) - L(v_1, ..., v_n)}{t} + L(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, w_n)$$

Next, define $\tilde{L} : (\mathbb{R}^N)^{n-1}$ as $\tilde{L}(x_1, ..., x_{n-1}) := L(x_1, ..., x_{n-1}, v_n)$. This is a $n-1$-linear map, so the induction hypothesis holds for this map; what is more, we can substitute it everywhere where L is given v_n as its n-th argument:

$$\frac{L(v_1 + tw_1, ..., v_n + tw_n) - L(v_1, ..., v_n)}{t}$$

$$= \frac{\tilde{L}(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}) - \tilde{L}(v_1, ..., v_{n-1})}{t} + L(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, w_n)$$

Now note that in the lower expression, we can take two limits separately: by the induction hypothesis,

$$\lim_{t \to 0} \frac{\tilde{L}(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}) - \tilde{L}(v_1, ..., v_{n-1})}{t} = \sum_{i=1}^{n-1} L(V^i N)$$

While, by continuity of linear maps,

$$\lim_{t \to 0} L(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, w_n) = L(v_1, ..., v_{n-1}, w_n) = L(V^n W)$$

Since sums of limits equal limits of sums, we discover that

$$\lim_{t \to 0} \frac{L(v_1 + tw_1, ..., v_N + tw_N) - L(v_1, ..., v_N)}{t} =$$

indeed exists, and equals

$$\lim_{t \to 0} \frac{\tilde{L}(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}) - \tilde{L}(v_1, ..., v_{n-1})}{t}$$
$$+ \lim_{t \to 0} L(v_1 + tw_1, ..., v_{n-1} + tw_{n-1}, w_n) =$$
$$\sum_{i=1}^{n-1} L(V^i W) + L(V^n W) = \sum_{i=1}^{n} L(V^i W)$$

This completes the induction step and thereby the proof. To repeat, the partial derivative $\partial_W L(V)$ exists and equals $\sum_{i=1}^{n} L(V^i W)$, for L an N-linear map where $N \geq 1$.

(ii) We now use that the determinant on $N \times N$-matices is N-linear on the columns. We use this to compute $\partial_B \det(A) = \lim_{t\to 0} \frac{\det(A+tB)-\det(A)}{t}$. This equals, by (i), $\sum_{i=1}^{N} \det(A^i B)$. This is the first part of the proof.

Next, taking $A = I$, we can write:

$$\lim_{t\to 0} \frac{\det(I+tB)-\det(I)}{t} = \sum_{i=1}^{N} \det(I^i B) = \lim_{t\to 0} \frac{t\sum_{i=1}^{N} \det(I^i B)}{t},$$

therefore, subtracting the right limit from the left (using additivity of limits) and using $\det(I) = 1$:

$$\lim_{t\to 0} \frac{\det(I+tB) - 1 - t\sum_{i=1}^{N} \det(I^i B)}{t} = 0, \text{ therefore, by definition of } o:$$

$$\det(I+tB) = 1 + t\sum_{i=1}^{N} \det(I^i B) + o(t)$$

So if we can show $\sum_{i=1}^{N} \det(I^i B) = \operatorname{trace}(B)$, we are done.

Let's calculate $\det(I^i B)$ by using the **Laplace expansion** to the i-th **row**: $\det(M) = \sum_{j=1}^{N} (-1)^{i+j} M_{ij} \det(\tilde{M}_{ij})$, where $\tilde{M}_{ij}$ is the $N-1 \times N-1$-matrix arising from M by deletion of the i-th row and j-th column. If we apply this to row i of $I^i B$, we notice that all elements $(I^i B)_{ij} \neq 0$ if and only if $i = j$, in which case the element is $(I^i B)_{ii} = B_{ii}$, and deletion of column i and row i will leave the $N-1 \times N-1$ identity matrix, which has determinant 1.

Hence,

$$\det(I^i B) = \det(M) = \sum_{j=1}^{N} (-1)^{i+j} (I^i B)_{ij} \det(\tilde{(I^i B)}_{ij})$$
$$= (-1)^{i+i} B_{ii} \det(I_{N-1}) = B_{ii}$$

. Therefore,

$$\sum_{i=1}^{N} \det(I^i B) = \sum_{i=1}^{N} B_{ii} = \operatorname{trace}(B)$$

, and we are done.

Exercise 3 Let $f : \Omega \to \mathbb{R}$ be a function that is differentiable at all $x \in \Omega$. Assume that Ω is an open connected set. Namely, for each $x, y \in \Omega$ it is possible to find a curve

$$\gamma : [0,1] \to \Omega$$

with $\gamma \in C^1([0,1]; \mathbb{R}^N)$, such that $\gamma(0) = x$, and $\gamma(1) = y$. Assume that $df(x) = 0$ for all $x \in \Omega$ Prove that f is constant

Proof Fix arbitrary $x, y \in \Omega$. It suffices to show that $f(x) = f(y)$. To this end, consider a $\gamma \in C^1([0,1];\mathbb{R}^N)$ with $\gamma(0) = x$, $\gamma(1) = y$. By Lagrange's Mean Value Theorem, we have:

$$f(y) - f(x) = \int_0^1 \langle \nabla f(\gamma(t), \gamma'(t) \rangle dt$$

Since $\gamma(t) \in \Omega$ for all $\in [0,1]$, and $df(x) = 0$ for all $x \in \Omega$, it follows $\nabla f(\gamma(t)) = 0$ for all $t \in [0,1]$. Therefore:

$$f(y) - f(x) = \int_0^1 \langle 0, \gamma'(t) \rangle dt = \int_0^1 0 dt = 0$$

Hence $f(y) = f(x)$, as was to be shown.