Analysis II – Lecture Notes

Colin Rothgang

April 26, 2017

Contents

Chapter 1

Differentiation in several variables

Lecture 1

Motivation

It's important to study the way functions in several variables change (leading to the notions of continuity and differentiability).

Example 1.1.

- *Newton's gravitational law* $F = G\frac{Mm}{r^2}$
- *Lagrange energy field* $L(q_1, q_2, \ldots, q_n, \dot{g_1}, \dot{g_2}, \ldots, \dot{g_n})$*, where* $G_i = g_i(t), \forall i \in \{1, \ldots, n\}$ *are positive variables and* $\dot{q_1}, \dot{q_2}, \ldots, \dot{g_n}$ *are the temporal derivatives of the position. Now we can write equations like* $L = K - V$*, where* K *denotes the kinetic energy and* V *the potential energy.*

Base idea Understand differentiation as linearisation. In one dimension this corresponds to finding the tangent line to the graph of a function. In two dimensions we can look either for tangent lines (partial differentiability) or a tangent plane (total differentiability). We will see that total differentiability $\implies$ partial differentiability.

Norms, Metrics and Banach and Hilbert spaces

Definition 1.2. *Let* $\mathcal{X}$ *be a set. Now* $\circ : \mathbb{R}^n \to \mathbb{R}^{\geq 0}$ *is called a norm iff:*

1. $||x|| = 0 \Leftrightarrow x = 0.$
2. $||\lambda x|| = |\lambda| \, ||x|| \, , \forall \lambda \in \mathbb{R}$
3. $||x + y|| \leq ||x|| + ||y|| \, , \forall x, y \in \mathcal{X}$

Definition 1.3. *On $\mathbb{R}^n$ we define the p-norm $||\cdot||_p : \mathbb{R}^n \to \mathbb{R}{\geq 0}$ by:*

$$\forall p \in \mathbb{N},\, \forall x \in \mathbb{R}^n.\;\; ||x||_p := ||(x_1, x_2 \ldots, x_n)||_p := \left(\left|\sum_{i=1}^{n} |x_i|^p\right|\right)^{\frac{1}{p}}.$$

For $p = 1$ we call the norm the Manhatten norm, for $p = 2$ we call the norm the Euclidean norm, which is defined by the inner product $\langle\cdot|\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \langle x|y\rangle := \sum_i x_i y_i$ in the sense that

$$||x||_2 := (|\langle x|x\rangle|)^{\frac{1}{2}}$$

Exercise 1. *The p-norm is a norm on $\mathbb{R}^n$.*

Definition 1.4. *A vector space equipped with a norm is called normed vector space.*

Definition 1.5. *A complete normed vector space is called Banach space.*

Definition 1.6. *A Banach space with an inner product $\langle x|y\rangle$ is called a Hilbert space iff its norm is defined by $||x|| := \sqrt{\langle x|x\rangle}$.*

Definition 1.7. *Let $\mathcal{X}$ be a set, $\mathrm{dist} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}{\geq 0}$ with*

1. $\mathrm{dist}(x, y) \geq 0$ *;non-negativity*
2. $\mathrm{dist}(x, y) = 0 \Leftrightarrow x = y$ *;identity of indiscernibles*
3. $\mathrm{dist}(x, y) = \mathrm{dist}(y, x)$ *;symmetry*
4. $\mathrm{dist}(x, z) \leq \mathrm{dist}(x, y) + \mathrm{dist}(y, z)$ *;triangle inequality*

, then dist *is called a metric on $\mathcal{X}$ and $(\mathcal{X}, \mathrm{dist})$ is called a metric space.*

Exercise 2. *If $\mathcal{X}$ is a normed vector space, then $\mathrm{dist}(x, y) := ||x - y||$ is a metric. In particular all p-norms on $\mathbb{R}^2$ are equivalent i.e.*

$$\exists C \in \mathbb{R},\, C = C(x, p, q) \geq 1 :\; \frac{1}{C} \leq \frac{||x||_p}{||x||_q} \leq C, \qquad \forall x \in \mathbb{R}^n.$$

This implies that all these norms define the same topology on $\mathbb{R}^n$.

Remark. *Given a metric space $(\mathcal{X}, \mathrm{dist})$, for any $R \in \mathbb{R} \geq 0, x \in \mathcal{X}$ sets of the form*

$$\mathrm{B}(x, R) := \{y \in \mathcal{X} :\, \mathrm{dist}(x, y) < R\},$$

are called "open balls in $\mathcal{X}$" (centered at x with radius R). For example, we define a topology in $\mathbb{R}^n$ by saying that a set $\mathcal{O} \subset \mathbb{R}^n$ is open iff $\forall x \in \mathcal{O}, \exists R = R(x) > 0.\, \mathrm{B}(x, R) \subset \mathcal{O}$.

Definition 1.8. *A sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ converges to $\mathcal{X} \in \mathbb{R}^n$ iff $||x_k - x|| \overset{k\to\infty}{\longrightarrow} 0$, or equivalently if $\forall \epsilon > 0.\, \exists N \in \mathbb{N}, \forall n \geq N \in \mathbb{N}.\; ||x_n - x|| < \epsilon$ or equivalently $\forall x \in \mathbb{R}^n, \forall \mathcal{O}$ open, $x \in \mathcal{O}.\, \exists N \in \mathbb{N}. \forall n \geq N \in \mathbb{N}.\; x_n \in \mathcal{O}$. If the limit exists we write $\lim_{k\to\infty} x_k$.*

Lemma 1.9. *Let* $x =: (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n, (x_k)_{k\in\mathbb{N}} = (x_{k_1}, x_{k_2}, x_{k_n}), k \in \mathbb{N}$. *Now we have*

$$\lim_{n\to\infty} x_k = x \Leftrightarrow \forall j \in \{1, 2, \ldots, n\}.\ \lim_{n\to\infty} x_{n_j} = x_j.$$

Proof. "$\Rightarrow$": We easily observe

$$\forall x \in \mathbb{R}^n, \forall i \in \{1, 2, \ldots, n\}.\, ||x|| = \sqrt{x_1^2 + \ldots + x_n^2} \geq \sqrt{x_i^2} = |x_i|$$

Fix $\epsilon > 0$, for any $i \in \{1, 2, \ldots, n\}$ we have $\epsilon_i' := \frac{\epsilon}{\sqrt{n}}$, $\exists\, |x_{k_i} - x_i| < \epsilon' = \frac{\epsilon}{\sqrt{n}}$. $\quad \forall k \geq N_\epsilon^i$. Now take $N_\epsilon := \max\{N_\epsilon^1, N_\epsilon^2, \ldots, N_\epsilon^n\}$, then

$$||x_k - x|| = \left(\sum_i (x_{k_i - x_i})^2\right)^{\frac{1}{2}} < \sqrt{\sum_i \epsilon'^2} = \epsilon \qquad \forall k \geq N_\epsilon.$$

□

Example 1.10. $x_k := (\frac{1}{k}, \frac{1}{k})_{k\in\mathbb{N}} \Rightarrow \lim_{k\to\infty} x_k = (0,0) = 0$ *We can use this sequence, by applying sandwich theorem, on the components sequences of a sequence of vectors.*

Definition 1.11. *Let* $\mathcal{X}$ *be any set. Now a point* $a \in \mathcal{X}$ *is called an accumulation point of* $\mathcal{X}$ *iff* $\forall \epsilon' > 0 : (\mathrm{B}(a, \epsilon') \setminus \{a\}) \cap \mathcal{U} \neq \emptyset$.

Definition 1.12. *Let* $f : \mathcal{U} \to \mathbb{R}^n, \mathcal{U} \subset \mathbb{R}^n$, *let* a *be an accumulation point for* $\mathcal{U}$. *If* $\exists L \in \mathbb{R}, \forall \epsilon > 0. \exists \delta > 0.\ ||f(x) - L|| < \epsilon \qquad \forall x \in \mathcal{U} \cap (\mathrm{B}(a, \delta) \setminus \{a\})$, *then* L *is the limit of* f *at* a, *denoted by* $L = \lim_{x\to a} f(x)$.

Remark. *One could also write (equivalent definition):* $\forall \epsilon > 0, \exists \delta > 0.\ x \neq a \Rightarrow ||x - a|| < \delta \ \Rightarrow ||f(x) - L|| < \epsilon$.

Lemma 1.13 (Anouther equivalent definition)**.** *Given a function* $f : \mathcal{U} \to \mathbb{R}^n$ *with* $\mathcal{U} \subset \mathbb{R}^n$ *and given an accumulation point* a *for* $\mathcal{U}$, *we have*

$$\exists \lim_{x\to a} f(x) = L \in \mathbb{R}^n \Leftrightarrow \forall (x_k) \in \mathcal{U} \| \lim_{k\to\infty} x_k = a \Rightarrow \lim_{k\to\infty} f(x_k) = L.$$

Example 1.14.

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

  – *Approach* $(0, 0)$ *along the x-axis:* $\longrightarrow f(x_k, y_k) = f(\frac{1}{k}, 0) = \frac{\frac{1}{k}\cdot 0}{\frac{1}{k^2}+0^2} = 0 \overset{k\to\infty}{\longrightarrow} 0$.

  – *Approach* $(0, 0)$ *along the y-axis:* $\longrightarrow f(x_k, y_k) = f(0, \frac{1}{k}) = \frac{0\cdot\frac{1}{k}}{0^2+\frac{1}{k^2}} = 0 \overset{k\to\infty}{\longrightarrow} 0$.

  – *Approach* $(0, 0)$ *along the main diagonal:* $\longrightarrow f(x_k, y_k) = f(\frac{1}{k}, \frac{1}{k}) = \frac{\frac{1}{k^2}}{\frac{1}{k^2}+\frac{1}{k^2}} = \frac{1}{2} \overset{k\to\infty}{\longrightarrow} \frac{1}{2}$.

  Hence, f *is not continuous at* 0.

- $g : \mathbb{R}^n \to \mathbb{R}$, $g(x,y) := \begin{cases} \frac{xy^3}{x^4+y^4} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$
  - *Take $y_k := cx_k$ and yield.*

$$g(x_k, y_k) = \frac{x_k c x_k^3}{x_k^4 + c^4 x_k^4} = \frac{c^3}{1+c^4} \Rightarrow \nexists \mathit{limit}.$$

- $h : \mathbb{R}^n \to \mathbb{R}$, $h(x,y) := \begin{cases} \frac{x^2}{x+y} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$.

  As it turns out this does not converge either.

Lecture 2

Definition 2.15. *Let $f : \mathcal{U} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ and assume let $a \in \mathbb{R}^n$ be an accumulation point. Now if $a \in \mathcal{U}$ and if $\exists \lim_{n\to\infty} f(x) = f(a)$, then we say that f is continuous at a. If all points in $\mathcal{U}$ are accumulation points for $\mathcal{U}$ and f is continuous in every point of $\mathcal{U}$, then we call f continuous on $\mathcal{U}$.*

Example 2.16. *1.* $f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) := \begin{cases} \frac{xy}{x^2+y^2} & (x,y) \neq 0 \\ 0 & (x,y) = 0 \end{cases}$. *We have seen that $\nexists \lim_{(x,y)\to(0,0)} f(x,y)$, hence f is not continuous in 0. How about the other points in the domain?*

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) \Leftrightarrow \forall y_n \overset{x_n \to x_0}{\to} y_0 . \exists \lim_{n\to\infty} f(x_n, y_n)$$

By the usual properties of sequences of reals and their limits, we obtain:

$$\begin{matrix} x_n \overset{n\to\infty}{\longrightarrow} x_0 \neq 0 \\ y_n \overset{n\to\infty}{\longrightarrow} y_0 \neq 0 \end{matrix} \Rightarrow \frac{x_n y_n}{x_n^2 + y_n^2} \overset{n\to\infty}{\longrightarrow} \frac{x_0 y_0}{x_0^2 + y_0^2},$$

hence f is continuous on $\mathbb{R}^2 \setminus \{(0,0)\}$, since $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \lim_{(x,y)\to(x_0,y_0)} \frac{xy}{x^2+y^2} = \frac{x_0 y_0}{x_0^2+y_0^2} = f(x_0,y_0)$

2. $f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) := \begin{cases} \frac{x^4+2x^2y^2}{x^2+y^2} & (x,y) \neq 0 \\ 0 & (x,y) = 0 \end{cases}$. *As before, we can conclude that g is continuous on $\mathbb{R}^2 \setminus \{0\}$. Furthermore, we have*

$$g(x,y) = \frac{x^2+2y^2}{x^2+y^2} x^2 \leq \frac{x^4+2x^2y^2+x^2}{x^2+x^2} = x^2+y^2 = ||(x,y)||^2 \overset{(x,y)\to(x_0,y_0)}{\longrightarrow} 0.$$

Thus g is continuous on its domain.

Theorem 2.17. *Let $f, g : \mathcal{U} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and a be an accumulation point for $\mathcal{U}$. Then assuming that the limits on the right-hand-side exist, we obtain the following identities:*

(i)

$$\lim_{x\to a} f(x) + g(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

(ii)

$$lim_{x \to a} \langle f(x)|g(x) \rangle = \left\langle \lim_{x \to a} f(x) \middle| \lim_{x \to a} g(x) \right\rangle$$

(iii)

$$\lim_{x \to a} f(x)g(x) = \left(\lim_{x \to a} f(x) \right) \left(\lim_{x \to a} g(x) \right)$$

If $g(x) \neq 0$, $\forall x \in \mathcal{U}$ and if $\exists \lim_{x \to a} g(x)$, then we further have:

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

1. *If we write $f : \mathcal{U} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with $f = (f_1, f_2, f_3 \ldots, f_n)$ with $f_j : \mathcal{U} \to \mathbb{R}$ $\forall j \in \{1, 2, \ldots, n\}$ then*

   $\exists \lim_{x \to a} f(x) = L \in \mathbb{R}^m = (L_1, L_2, \ldots, L_n) \Leftrightarrow \forall j \in \{1, 2, \ldots, m\}. \exists \lim_{x \to a} f_j(x) = L_j$

Proof. (i), (iii) follow from the definitions, from lemma 1.9 and the additivity of limits on $\mathbb{R}$. (iv) follows from:

$$||f(x) - L|| \leq |f_j(x) - L_j| \qquad \forall j \in \{1, 2, \ldots, n\}.$$

(iii) follows from (i), (iii) and (iv). □

Remark. *The analogous rules also hold for continuity i.e. sums, products, inner product, components of continuous functions are also continuous. Furthermore, sum, inner product, product with scalar are continuous themselves.*

Definition 2.18 (Cauchy Sequence)**.** *Let $(\mathcal{X}, \text{dist})$ be a metric space and let a_n in $\mathcal{X}$.Now a_n is called Cauchy (or a Cauchy sequence) iff $\forall \epsilon > 0.\ \exists N \in \mathbb{N}.\forall n, m \geq N.\, \text{dist}(a_n, a_m)) < \epsilon$.*

Definition 2.19 (A complete metric space)**.** *A metric space $(\mathcal{X}, \text{dist})$ is called complete or Banach iff every Cauchy sequence in the space is convergent.*

Theorem 2.20. $(\mathbb{R}^n, ||\cdot||)$ *is Banach.*

Proof. Let (x_k) be a Cauchy sequence. Then $\exists N_\epsilon \in \mathbb{N}.\, ||x_k - x_l|| < \epsilon$, $\forall k, l \geq N_\epsilon \Rightarrow (x_{k_i})$ Cauchy $\forall i \in \{1, 2, \ldots, n\}$. Then x_{k_i} converges to some $x_i \in \mathbb{R}\, \forall i \in \{1, 2, \ldots, n\}$. Now put $x := (x_1, x_2, \ldots, x_n)$ and apply lemma 1.9. It follows $x_k \to x$. Therefore, $\mathbb{R}^n, ||\cdot||)$ is Banach. □

Lecture 3

Remark. *An inner product, allows us to "measure" angles between vectors.*

Differentiability; linear maps, matrices and the operator norm

Definition 3.21. *Let $\mathcal{X}$, $\mathcal{Y}$ be two vector spaces over $\mathbb{R}$. A map $A : \mathcal{X} \to \mathcal{Y}$ is called linear iff $A(\alpha x + \beta y) = \alpha A(x) + \beta A(x) \qquad \forall \alpha, \beta \in \mathbb{R}, \forall x, y \in \mathcal{X}$. The set of linear maps $\mathcal{X} \to \mathcal{Y}$ is denoted by $L(\mathcal{X}, \mathcal{Y})$. $L(\mathcal{X}, \mathcal{X})$ is denoted $L(\mathcal{X})$. Often we write $A\ x$ instead of $A(x)$.*

Facts:

1. $A\,0 = 0 \qquad \forall A \in L(\mathcal{X}, \mathcal{Y})$.
2. If $\mathcal{X}$ is finite dimensional (e.g. $\mathbb{R}^n$), then $A \in L(\mathcal{X})$ bijective $\Leftrightarrow$ A injective $\Leftrightarrow$ A surjective.

   Proof Sketch. A surjective $\Leftrightarrow \langle A\mathcal{X}\rangle = \langle \mathcal{X}\rangle \Leftrightarrow \langle A^{-1}\mathcal{X}\rangle = \langle \mathcal{X}\rangle \Leftrightarrow A$ injective (by linearity). □

3. $\forall A_1, A_2 \in L(\mathcal{X}, \mathcal{Y})$, $\forall \alpha_1, \alpha_2 \in \mathbb{R}$:
$$(\alpha_1 A_1 + \alpha_2 A_2) \in L(\mathcal{X}, \mathcal{Y}).$$
   Thus $L(\mathcal{X}, \mathcal{Y})$ is a vector space.

Definition 3.22. *The set $\mathcal{M}^{m\times n}(\mathbb{R})$ of $(m \times n)$-matrices $\forall m, n \in \mathbb{N}$ fixed consists of all vectors in $\mathbb{R}m \cdot n$ written in a "separable form" i.e.*

$$A := (a_{i,j})_{\substack{1 \le i \le n \\ 1 \le j \le m}} =: \begin{pmatrix} a_{1,1}, a_{1,2}, \dots, a_{1,n} \\ a_{2,1}, a_{2,2}, \dots, a_{2,n} \\ \vdots \\ a_{m,1}, a_{m,2}, \dots, a_{m,n} \end{pmatrix}.$$

Definition 3.23. *Matrix multiplication is defined by: Let $A \in \mathcal{M}^{n\times m}(\mathbb{R}), B \in \mathcal{M}^{m\times p}(\mathbb{R})$. Then $C := A \cdot B := (c_{k,j})_{\substack{1 \le j \le n \\ 1 \le k \le p}}$, where $c_{k,j} = \sum_{i=1}^m a_{i,j} b_{k,i} \in \mathcal{M}^{n\times p}(\mathbb{R})$. Furthermore, $A \in \mathcal{M}^{m\times n}(\mathbb{R})$ is called invertible iff $\exists A^{-1} \in \mathcal{M}^{n\times m}(\mathbb{R})$. $AA^{-1} = A^{-1}A = I_n := \begin{pmatrix} 1\ 0\ 0\ \dots 0 \\ 0\ 1\ 0\ 0 \dots \\ \vdots \\ 0\ 0\ \dots 0\ 1 \end{pmatrix}$ ($\Rightarrow m = n$), where I_n is the identity matrix.*

Remark. *Matrix multiplication is not commutative (only associative, as is function composition in general).*

Remark. *$\mathcal{M}^{m\times n}$ is a vector space over $\mathbb{R}$ for all $m, n \in \mathbb{N}$.*

4. Let $\mathcal{X}, \mathcal{Y}$ be finite dimensional vector spaces with basis $\{x_1, x_2, \dots, x_n\}$, $\{y_1, y_2, \dots, y_m\}$. We can now use these basis to define the natural isomorphism between $\mathcal{M}^{m\times n}$ and
$$L(\mathcal{X}, \mathcal{Y}) : A_{i,j} \to \left((x_k)_{k\in\{1,2,\dots,n\}} \to \left(\sum_{k=1}^m a_{i,k} \right)_{i\in\{1,2,\dots,n\}} \right).$$
   By linearity and item (2) we see that this mapping is a linear bijection, thus it is an isomorphism. Sometimes we will denote the matrix associated to an linear map (given a fixed basis) A by $[A]$. Usually we will not distinguish between A and $[A]$.

5. Let $A \in L(\mathcal{X}, \mathcal{Y}), B \in L(\mathcal{Y}, \mathcal{Z})$ be linear maps, then $B \circ A \in L(\mathcal{X}, \mathcal{Z})$ and $[B \circ A] = [B] \cdot [A]$.
6. $A \in L(\mathcal{X})$ inveritble $\Leftrightarrow$ $[A]$ invertible (under multiplication) $\Leftrightarrow \det(A) \neq 0$.

Remark. *$\mathcal{M}^{m \times n}$ is topologically just $\mathbb{R}m \cdot n$.*

Definition 3.24 (operator norm)**.** *For $A \in L(\mathcal{X}, \mathcal{Y})$, define the operator norm by $||A|| := \sup_{||x|| \leq 1} ||Ax||$.*

Remark. *operator norm not isometric to $||\cdot||_2$.*

Lemma 3.25. *If $\alpha \in \mathbb{R}$ has the property that $||Ax|| \leq ||x_i||$ $\forall x \in \mathbb{R}^n$, then $||A|| \leq \alpha$.*

Proof. $\forall x \in \mathbb{R}^n$, $||u|| \leq 1$, we have that $||Au|| \leq \alpha ||u|| \leq \alpha$. Thus, by definition of the operator norm $||A|| = \sup_{||u|| \leq 1} ||Au|| \leq \alpha$. □

Theorem 3.26 (Properties of the operator norm)**.**

1. $||Ax|| \leq ||A|| \, ||x|| \,, \forall \in L(\mathbb{R}, \mathbb{R}^n)$ $\forall x \in \mathbb{R}^n$.

2. $||A|| < \infty, \qquad \forall A \in L(\mathbb{R}^n, \mathbb{R}^m)$.

3. $\forall A \in L(\mathbb{R}^n, \mathbb{R}^m)$. A Lipschitz i.e. $\exists$ a real constant (here the operator norm) s.t. $||Ax - Ay|| \leq ||A|| \, ||x - y|| \,, \forall x, y \in \mathbb{R}^n$.

   Exercise 3. *This implies uniform continuity.*

4. The operator norm is a norm i.e. $\forall A, B \in L(\mathbb{R}^n, \mathbb{R}^m), \forall \alpha \in \mathbb{R}$, we have:

   (a) $||A + B|| \leq ||A|| + ||B||$, the triangular inequality

   (b) $||\alpha A|| = |\alpha| \, ||A||$

5. $\forall A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$. $||BA|| \leq ||B|| \, ||A||$.

6. With $I(\mathbb{R}^n) \subset L(\mathbb{R}^n) := \{A \in L(\mathbb{R}^n) \text{ invertible}\}$ we have that for $A \in I(\mathbb{R}^n)$, $B \in L(\mathbb{R}^n)$ and $||B - A|| \left|\left|A^{-1}\right|\right| < 1$ it follows $B \in I(\mathbb{R}^n)$. In other words, $\forall A \in I(\mathbb{R})$, we have that $\mathrm{B}(A, \frac{1}{||A^{-1}||}) \subset I(\mathbb{R}^n)$. Thus, $I(\mathbb{R}^n)$ is open in $\mathbb{R}^n$.

7. $\forall A \in I(\mathbb{R}^n)$. A^{-1} is continuous.

Proof.

1. Fix $x \in \mathbb{R}, x \neq 0$ then for $\frac{x}{||x||}$ we obtain $\left|\left|\frac{x}{||x||}\right|\right| = \frac{1}{||x||} ||x|| = 1$. By definition $||A|| \geq \left|\left|A\left(\frac{x}{||x||}\right)\right|\right| = \frac{1}{||x||} ||A(x)|| \Rightarrow ||A|| \, ||x|| \geq ||A||$. The case $x = 0$ is trivial.
2. $\forall A \in L(\mathbb{R}^m, \mathbb{R}^n)$. $||A|| \leq m \cdot n \cdot \max(A) < \infty$.
3. Follows from linearity and item 2.
4. (a) Obvious.
   (b) By linearity (preserved by sup)

5. By item 1 and Lemma 3.25:

$$||BAx|| \leq ||BA|| \, ||x|| \leq ||B|| \, ||A|| \, ||x|| \tag{1.1}$$

6. Using item 1 and linearity:

$$\frac{||x||}{||A^{-1}||} = \frac{||A^{-1}Ax||}{||A^{-1}||} \overset{\text{item 1}}{\leq} \frac{||A^{-1}||}{||A^{-1}||} \, ||Ax|| = ||Ax||$$
$$= ||(A-B)x + Bx|| \overset{\text{item 4}}{\leq} ||A-B|| \, ||x|| + ||Bx||$$

Therefore by assumption $0 < \frac{1}{||A^{-1}||} \, ||x|| \leq ||Bx||$. Thus, $x \neq 0 \Rightarrow 0 < ||Bx|| \Rightarrow B$ invertible.

7. By item 5 and item 6: $||BA|| \leq ||B|| \, ||A||$.

$\square$

Lemma 3.27. $||A|| = 0 \Rightarrow A$ *not invertible.*

Proof. Assume $||A|| = 0 \Rightarrow \forall x \in \mathbb{R}^n.\ 0 = ||A|| \, ||x|| \geq ||Ax|| \Rightarrow ||Ax|| = 0$. Thus A is not invertible at all. $\square$

Lecture 4

Definition 4.28. *Let $f : \mathcal{U} \to \mathbb{R}$ be some function and x_0 an accumulation point of $\mathcal{U}$. We say that f is differentiable in x_0 iff $\exists \lim_{x \to x_0} \frac{f(x)-f(x_0)}{x-x_0} =: f'(x)$. Another way to write this is $\exists \lim_{h \to 0} \frac{f(x_0+h)-f(x_0)}{h} =: f'(x)$.*

Remark. *Equivalently, f is differentiable in $x_0 \Leftrightarrow \exists$ linear map $A \in L(\mathbb{R})$ s.t. $\lim_{x \to x_0} \frac{f(x)-f(x_0)-A(x-x_0)}{|x-x_0|} = 0$.*

So how can we generalize this to several dimensions?

Remark. *From now on we will mostly consider functions on ope sets, thus all points in the domain will be open.*

Definition 4.29. *Let $f : \mathcal{U} \to \mathbb{R}^n, \mathcal{U} \subset \mathbb{R}^m$ open, be a map and let $x_0 \in u$. The map f is called (totally) differentiable in x_0 iff $\exists A \in L(\mathbb{R}^n, \mathbb{R}^m)$, so that*

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - A(x - x_0)}{||x - x_0||} = 0,$$

or

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - A(x_0)h}{h} = 0.$$

The linear approximation A of f is called the total differential of f in x_0 and is denoted by $D_x f$ or $Df(x_0)$ or $\mathrm{d}f(x_0)$ of $f'(x_0)$.

Lemma 4.30 (Uniqueness)**.** *If $A_1, A_2 \in L(\mathbb{R}^n, \mathbb{R}^m)$ both sotisfy the above condition, then $A_1 = A_2$.*

Proof. Before applying the definition of differentiability twice, we observe:

$$\begin{aligned}&||(A_1 - A_2)h||\\ =&||A_1 h - A_2 h||\\ =&||f(x_0+h) - f(x_0) - A_2 h - (f(x_0+h) - f(x_0) - A_1 h)||\\ \leq&||f(x_0+h) - f(x_0) - A_2 h|| + ||f(x_0+h) - f(x_0) - A_1 h||\end{aligned}$$

Now divide by $||h||$ (assuming $h \in \mathbb{R}^n \setminus \{0\}$):

$$0 \leq \frac{||(A_1 - A_2)h||}{||h||} \leq \frac{||f(x_0+h) - f(x_0) - A_1 h||}{h} + \frac{||f(x_0+h) - f(x_0) - \delta h||}{h}.$$

Thus, $\lim_{h\to 0} \frac{||(A_1 - A_2)h||}{||h||} = 0$.

Consider some (fixed but arbitrary) $h \in \mathbb{R}^n$ and look the $th, \forall t \in \mathbb{R}$. The limit being 0 means in particular that $\lim_{t\to 0} \frac{||A_1 - A_2||(th)}{th} = 0$. By linearity of A_1, A_2: $\frac{||t(A_1-A_2)h||}{th} = \frac{||(A_1-A_2)h||}{||h||}$. But doesn't depend on t anymore. $\forall h \in \mathbb{R}^n \setminus \{0\} \Rightarrow \frac{||(A_1-A_2)h||}{||h||} = 0$ and $||(A_1 - A_2)h|| = 0$. It follows that $A_1 = A_2$. □

Remark. *In the definition of differentiability one can also write* $\lim_{x\to x_0} \frac{||f(x)-f(x_0)-A(x-x_0)||}{||x-x_0||} = 0$.

Remark. $\lim_{x\to x_0} ||g(x)|| = c \neq 0 \not\Rightarrow \exists \lim_{x\to x_0} g(x)$.

Counter-Example 4.31. $h(x) := \begin{cases} 1, & x \in (0, \frac{1}{2}) \\ 0, & x \in [\frac{1}{2}, 1] \end{cases}$. *Define* $g : [0,1] \to \mathbb{R}^2$ *by* $g(x) := \begin{pmatrix} c\cos(x) \\ c\sin(x) \end{pmatrix}$, *and* $||g||$ *at* $x = \frac{1}{2}$: $\lim_{x\to\frac{1}{2}} ||g(x)|| = 0$, *but* $\exists \lim_{x\to\frac{1}{2}} g(x)$.

Example 4.32.

- *The differential of* $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ *in any point is* A *itself.*
- *Affine maps: Consider any affine map* $\mathbb{R}^n \to \mathbb{R}^m, x \to Ax + b, \forall b \in \mathbb{R}^m, A \in L(\mathbb{R}^n, \mathbb{R}^m)$. *Here the differential is* A *again.*

Proposition 4.33. $f : \mathcal{U} \to \mathbb{R}^m$ *is differentiable in* $x_0 \in \mathcal{U} \Leftrightarrow \exists \varphi : \mathcal{U} \to \mathbb{R}^m$. $\exists A \in L(\mathbb{R}^n, \mathbb{R}^m)$ *s.t.* $f(x) = f(x_0) + A(x - x_0) + \varphi(x - x_0)$, *where* $\lim_{x\to x_0} \frac{\varphi(x-x_0}{||x-x_0||} = 0$.

Lecture 5

Proposition 5.34. *If any function* $f : \mathcal{U} \subset \mathbb{R}^n \to \mathbb{R}^m$, *for* $\mathcal{U}$ *open is differentiable in* $x_0 \in \mathcal{U}$, *it is also continuous in* x_0.

Proof. By differentiability in x_0, $\exists D$ s.t.

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \varphi(x - x_0)$$

with $lim_{x\to x_0} \frac{\varphi(x-x_0)}{||x-x_0||}$. Thus $\lim_{x\to x_0} \varphi(x - x_0) = 0$, so $\lim_{x\to x_0} f(x) = f(x_0) + Df(x_0)(x - x_0) = f(x_0)$. □

Proposition 5.35 (linearity)**.** *Let $f, g : \mathcal{U} \to \mathbb{R}^m$ differentiable in $x_0 \in \mathcal{U}, \alpha, \beta \in \mathbb{R}$, $\mathcal{U} \subset \mathbb{R}^n$ open, then $\alpha f + \beta g : \mathcal{U} \to \mathbb{R}^m$ is differentiable in x_0 and $D(\alpha f(x_0) + \beta g(x_0) = \alpha Df(x_0) + \beta Dg(x_0)$.*

Proof. Plugging in the definition of differentiability, we yield:

$$\begin{aligned}
&\lim_{x \to x_0} f(\alpha x + \beta x) \\
=&\lim_{x \to x_0} f(\alpha x_0 + \beta x_0) + Df(\alpha x + \beta x - \alpha x_0 - \beta x_0) + \varphi(\alpha x + \beta x - \alpha x_0 - \beta x_0) \\
\overset{\varphi \to 0}{=}&\lim_{x \to x_0} f(\alpha x_0 + \beta x_0) + Df(\alpha x + \beta x - \alpha x_0 - \beta x_0)
\end{aligned}$$

By definition of differentiability Df is linear:

$$\Rightarrow Df(\alpha x + \beta x - \alpha x_0 - \beta x_0) = \alpha Df(x_0) + \beta Dg(x_0)$$

□

Theorem 5.36 (chain rule)**.** *Let $f : \mathcal{U} \to \mathbb{R}^m, \mathcal{U} \in \mathbb{R}^n$ open be differentiable in $x_0 \in \mathcal{U}$ and $g : \mathcal{V} \to RR^p$, for $\mathcal{V} \in \mathbb{R}^n$ open with $y_0 := f(x_0) \in \mathcal{V}$. Then $g \circ f : \mathcal{U} \to \mathbb{R}^n$ is differentiable in x_0 and $D(g \circ f)(x_0) = Dg(f(x_0)) \circ Df(x_0)$.*

Proof. From differentiable of f in x_0 and g in y_0, we obtain:

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \varphi(x - x_0), \lim_{x \to x_0} \frac{\varphi(x - x_0)}{||x - x_0||} \tag{1.2}$$

$$f(y) = f(y_0) + Df(y_0)(y - y_0) + \varphi(y - y_0), \lim_{y \to y_0} \frac{\varphi(y - y_0)}{||y - y_0||} \tag{1.3}$$

Putting $y := f(x), x \in \mathcal{U}$, we get (allowed since f continuous):

$$y - y_0 = f(x) - f(x_0) = Df(x_0)(x - x_0) + \varphi(x - x_0) \tag{1.4}$$

Using this equations, we wan to give a linear approximation of $g \circ f$ in $g(f(x_0)) = g(y_0)$:

$$\begin{aligned}
&g(f(x)) \\
\overset{(1.3)}{=}&g(f(x_0)) + (Dg(f(x_0)(x - x_0) + \varphi(x - x_0)) + \varphi(f(x) - f(x_0)) \\
\overset{(1.2)}{=}&g(f(x_0)) + Dg(f(x_0))\,(Df(x_0)(x - x_0) + \varphi(x - x_0)) + \varphi(f(x) - f(x_0)) \\
\overset{\text{linearity}}{=}&g(f(x_0)) + Dg(f(x_0))\,(Df(x_0)(x - x_0)) + Dgf(x_0)\varphi(x - x_0) + \psi(f(x_0) - f(x_0)).
\end{aligned}$$

Defining $\psi(\theta(fx - x_0) := \psi(f(x) - f(x_0)) + Dg(f(x_0))\varphi(x - x_0)$:

$$=(g \circ f)(x_0) + Df(x_0) \circ D(g \circ f)(x_0)(x - x_0) + \theta(x - x_0).$$

Firstly, since $Dg(f(x_0)) \in L(\mathbb{R}^m, \mathbb{R}^n)$, it follows from

$$\lim_{x \to x_0} \frac{\varphi(x - x_0)}{||x - x_0||} = 0, \text{ that}$$

$$\lim_{x \to x_0} \frac{Dg(f(x_0))\varphi(x - x_0)}{||x - x_0||} = 0$$

Secondly, define φ_0 by $\varphi(y-y_0) =: ||y-y_0||\,\Psi_0(y-y_0)$, so $\lim_{x-x_0}\Psi_0(y-y_0) = 0$ and use (1.2), the triangular inequality and the operator norm to estimate:

$$\begin{aligned}
&||\Psi(f(x)-f(x_0)|| = ||f(x)-f(x_0)||\,||\Psi_0(f(x)-f(x_0)|| = ||df(x_0)(x-x_0)+\varphi(x-x_0)|| \\
&= ||\Psi_0(f(x))-f(x_0)|| \leq (||Df(x_0)(x-x_0)||+||\varphi(x-x_0)||)\,||\Psi_0(f(x)-f(x_0)|| \\
&= ||Df(x_0)||\,||x-x_0||
\end{aligned}$$

Division by $||x-x_0||$ yields:

$$\begin{aligned}
&0 \\
&\leq \frac{\Psi(f(x)-f(x_0)}{||x-x_0||} \\
&\leq \left(||Df(x_0)||+\frac{||\varphi(x-x_0)||}{x-x_0}\right)||\Psi(f(x)-f(x_0)||
\end{aligned}$$

Since $\lim_{z\to 0}\Psi_0(z) = 0$ and f is continuous in x_0, so $\lim_{x\to x_0} f(x) = f(x_0)$, we obtain $\lim_{x\to x_0}\Psi_0(f(x)-f(x_0)) = 0$. We can conclude

$$\lim_{x\to x_0}\text{RHS} = 0 \Rightarrow \lim_{x\to x_0}\frac{\Psi(f(x)-f(x_0)}{||x-x_0||} = 0.$$

□

Proposition 5.37 (Lipschitz continuity and $Df = 0 \Rightarrow f$ constant)**.** *Let $f: \mathcal{U} \to \mathbb{R}^m$, $\mathcal{U} \subset \mathbb{R}^n$ open and convex, i.e. $\forall u,v \in \mathcal{U}, \forall t \in [0,1] : (1-t)u+tv \in \mathcal{U}$, be differentiable in $\mathcal{U}$ s.t. $\exists M \geq 0$ with $||Df(x)|| \leq M, \forall x \in \mathcal{U}$. Then, f is M-Lipschitz continuous on $\mathcal{U}$, i.e. $||f(x)-f(y)|| \leq M\,||x-y||\,, \forall x,y \in \mathcal{U}$. In particular, if $M = 0$ meaning that Df is constant on $\mathcal{U}$ then f is a constant function.*

Proof. To be continued... □

Remark. *The special case is a generalisation of hte statement that if $f : I \to \mathbb{R}$ is differentiable with $f \cong 0$, then f must be constant.*

Definition 5.38. *Let $f : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open and consider $x_0 \in \mathcal{U}$ and $v \in \mathbb{R}^n$ so that $||v|| = 1$. If the limit exists $D_v f(x_0) := \lim_{t\to 0}\frac{f(x_0+tv)-f(x_0)}{t}$, it is called the directional derivative of f in x_0 in direction of v.*

Example 5.39. $f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) := \begin{cases}\frac{xy}{x^2+y^2} & (x,y) \neq 0\\ 0 & \textit{otherwise}\end{cases}$ *. We can check that the partial and even the total differential exist in all $(x,y) \neq 0$, but f is not continuous in 0. It is, however, more interesting to look for f in 0. So we will later come back to this example.*

Definition 5.40. *If $v = e_j \quad \forall j \in \{1,2,\dots,n\}$ in previous definition, then the corresponding directional derivatives the j-th partial derivatives of f in x_0 denoted by $f_{x_j}(x_0), D_j(f(x_0)), \frac{\partial f}{\partial x_j}(x_0)$.*

Example 5.41.

1. $f : \mathbb{R} \to \mathbb{R}, f(x,y) := x^2+y^2$, $\frac{\partial f}{\partial x}((x,y)) = 2x, \frac{\partial f}{\partial y}((x,y)) = 2y$

2. $g : \mathbb{R}^2 \to \mathbb{R}, g(x, y) := \begin{cases} \frac{xy}{x^2+y^2} & (x, y) \neq 0 \\ 0 & otherwise \end{cases}$. $\frac{\partial g}{\partial x}((0, 0)) = \lim_{t\to 0} \frac{g((t,0))}{t} = \lim_{t\to 0} \frac{0}{t} = 0$, $\frac{\partial g}{\partial y}((0, 0)) = \lim_{t\to 0} \frac{g((0,t))}{t} = \lim_{t\to 0} \frac{0}{t} = 0$. *However, g is not continuous in 0 let alone differentiable:* $y := \alpha x_i, \lim_{x\to 0} \frac{x\alpha x}{x^2+\alpha^2 x^2} = \frac{\alpha}{1+\alpha^2}$

3. *Show that f has partial derivatives everywhere apart from the origin.*

Definition 5.42. *$f : \mathcal{U} \to \mathbb{R}$, $\mathcal{U} \subset \mathbb{R}$ open is called partially differentiable iff partial derivatives exist and are continuous on $\mathcal{U}$.*

Example 5.43.
1. *$f : \mathbb{R}^2 \to \mathbb{R}, f((x, y)) := x^2+y^2$, is continuously partially differentiable on $\mathbb{R}^2$, since both $\mathbb{R}^2 \to \mathbb{R}, (x, y) \to 2x, (x, y) \to 2y$ are continuous on $\mathbb{R}^2$.*

2. *Norms on $\mathbb{R}^k$ are continuously partially differentiable on $\mathbb{R}^k \setminus \{0\}$.*

Definition 5.44 (Jacobean)**.** *$f : \mathcal{U} \to \mathbb{R}^m$, $\mathcal{U} \subset \mathbb{R}^n$ open, so that all partial derivatives $\frac{\partial f_i}{\partial x_j}(x_0), \forall i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, n\}$ in $\times_o$ exist. Then the matrix $J_f(x_0) := \left(\frac{\partial f_i}{\partial x_j}\right)^{1\leq m}_{1\leq n} \in \mathcal{M}^{m\times n}(\mathbb{R})$ is called the Jacobean matrix of f in x_0.*

Remark. *We will see that in the special case $m = 1$, the Jacobean is also called gradient of f in x_0 and has a special geometric meaning.*

Example 5.45. *$f : \mathbb{R}^3 \to \mathbb{R}^2, f((x, y)) := \begin{pmatrix} x^2 + y^2 + z^2 \\ xyz \end{pmatrix}$. f is continuously partially differentiable everywhere and the Jacobean is given by $J_f((x, y, z)) = \begin{pmatrix} 2x & 2y \\ 2z & yz \\ xz & xy \end{pmatrix} \in \mathcal{M}^{2\times 3}(\mathbb{R})$.*

Lecture 6

Theorem 6.46 (total differentiability implies partial differentiability)**.** *Let $f : \mathcal{U} \to \mathbb{R}^m, \mathcal{U}$ open be totally differentiable in $x_0 \in \mathcal{U}$. Then f is partially in x_0 and $[df(x_0)] = J_f(x_0)$.*

Proof. f is totally differentiable in $x_0 \Rightarrow \forall h \in \mathbb{R}^n. f(x_0+h) = f(x_0)+Ah+\varphi(h)$, where $\lim_{h\to 0} \frac{\varphi(h)}{h} = 0$. Define $[Df(x_0)] := A(x_0) = (a_{i,j})^{1\leq i\leq m}_{1\leq i\leq n}$. Note we can write $Df(x_0)(h) = Ah = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1,j}h_j \\ \vdots \\ \sum_{j=1}^n a_{m,j}h_j \end{pmatrix}$. Written componentwise, this means for $i \in \{1, 2, \dots, m\} : f_i(x_0 + h) = f_i(x_0) + \sum_{j=1}^n a_{i,j}h_j + \varphi(h)$, where $\varphi\left(\begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi \end{pmatrix}\right)$ and $\lim_{h\to 0} \frac{\varphi_i(h)}{||h||} = 0$. Now putting $h = te_j, t \in \mathbb{R} \Rightarrow x_0 + te_j \in \mathcal{U}, \forall j \in \{1, 2, \dots, n\}$. Thus, $f_j(x_0 + te_j) = f_i(x_0)+\sum_{l=1}^n a_{il}th_i+\varphi_i(te_j) = f_j(x_0)+ta_{ij}+\varphi(te_j)$. Finally, let us look at the

j-th partial derivatives $\frac{\partial f_i}{\partial x_j}(x_0)$ of the i-th component f_i of f in x_0. By definition $\frac{\partial f_i}{\partial x_j}(x_0) = \lim_{t\to 0} \frac{f_j(x_0+te_j)-f(x_0)}{t} = \lim_{t\to 0} \frac{ta_{ij}+\varphi_i(te_j)}{t} = a_{ij}t\lim_{t\to 0}\frac{\varphi_i(te_j)}{||te_j||} = 0$, since $\lim_{h\to 0}\frac{\varphi_i(h)}{||h||} = 0$.

Thus, the matrix A giving the differential is exactly the Jacobean in x_0. $\square$

Theorem 6.47 ($\exists$ and continuity of partial derivatives $\Rightarrow$ total differentiability)**.**

Remark. *Now given a partial differentiable function we can first determine the Jacobean, check its partial derivatives for continuity and if they are construct the total differential using the Jacobean.*

Example 6.48. $f : \mathbb{R}^3 \to \mathbb{R}^2, f((x,y)) = \begin{pmatrix} x^2+y^2+z^2 \\ xyz \end{pmatrix}, J_f((x_0,y_0,z_0)) = \begin{pmatrix} 2x_0 & 2y_0 & 2z_0 \\ y_0z_0 & x_0z_0 & x_0y_0 \end{pmatrix} \in \mathcal{M}^{3\times 2}(\mathbb{R})$. *Therefore,* $Df((x_0,y_0,z_0)) \in \mathrm{L}(\mathbb{R}^n,\mathbb{R}^m)$ *is defined as* $Df((x_0,y_0,z_0))(x,y,z) = \begin{pmatrix} 2(x_0x+y_0y+z_0z) \\ xy_0z_0+x_0yz_0+x_0y_0z \end{pmatrix}$.

Remark. *The converse is not true,* f *totally differentiable* $\not\Rightarrow$ *continuous partially differentiable. For instance in case of*

$$f:\mathbb{R}^2\to\mathbb{R}, f((x,y)) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & (x,y)=0 \\ 0 & \textit{otherwise} \end{cases}.$$

Remark. *One can show that continuous partial differentiability is equivalent to continuous total differentiability in the sense that* $x \to Df(x)$ *is continuous.*

Lecture 7

Theorem 7.49. *Let* $f : \mathcal{U} \to \mathbb{R}^m, \mathcal{U} \subset \mathbb{R}^m$ *open be any function and let its partial derivatives* $\frac{\partial f}{\partial x_i}\mathcal{U} \to \mathbb{R}$ *iff* $\imath \in \{1,2,\ldots,m\}, j \in \{1,2,\ldots,1,n\}$ *exist and are continuous (f is continuously differentiable). Then f is totally differentaible and the total differential is given as* $[Df(x_0)] = J_f(x_0) \quad \forall x_0 \in \mathcal{U}$.

Proof. Wlog. let $m=1$ (component-wise total differentiability is equivalent to total differentiability as already shown). WE need to show (given fixed x_0) $f(x_0+h) = f(x_0)+Ah+\varphi(h)$ with $\lim_{h\to 0}\frac{\varphi(h)}{||h||} = 0$, $\forall h \in \mathbb{R}^n$, put $h^0 := 0 \in \mathbb{R}^n, \ldots h^j := \sum_{l=1}^{j} h_le_l = (h_1e_1,\ldots,h_je-j,0,\ldots,0),\ldots,h^n = h$, where $h = \sum_{l=1}^{n} h_1e_1$ with $h_j \in \mathbb{R}$ and $\{e_1,e_2,\ldots,e_n\}$ the standard basis in $\mathbb{R}^n$. We can write

$$f(x_0+h)-f(x_0) = f(x_0+h)-f(x_0+h^{n-1})+f(x_o+h^{n-1})-\ldots+f(x_0+h^1)-f(x_0)$$

$$= \sum_{j=1}^{n} f(x_0+h^j)-f(x_0+h^{j-1}) \tag{1.5}$$

Applying the mean value theorem to f restricted to the segments $[x_0+h^j, x_0+h^{j+1}]$ leads to $\exists \xi_j \in (x_0+h^{j-1}, x_0+h^j)$, so that $\frac{\partial f}{\partial x_j}(x_0+h^j+\xi_j) = \frac{f(x_0+h^j)-f(x_0+h^{j-1})}{h_j}$.

Local definition 7.49.1. *Here a segment in* $\mathbb{R}^n$ *is defined as* $[x,y] := \{(1-t)x+ty : t \in [0,1]\} \quad \forall x,y \in \mathbb{R}^n$.

Substituting these equalities into equation 1.5 leads to:

$$f(x_0+h)-f(x-0) = \sum_{j=1}^{n} \left[f(x_0+h^j) - f(x_0+h^{j-1})\right] = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\left(x_0+h^{j-1}+\xi^j h_j e_j\right) h_j.$$

Coming back to the definition we want to show that $\lim_{h\to 0}\frac{\varphi(h)}{||h||} = 0$, where

$$S\varphi(h) = f(x_0)-f(x_0+h)-Ah = \sum_{j=1}^{n}\left(\frac{\partial f}{\partial x_j}\left(x_0+h^{j-1}+\xi^j h_j e_j\right) - \frac{\partial f}{\partial x_j}(x_0)\right) h_j.$$

By construction, $(x_0+h^{j-1}+\xi^j h_j e_j) \xrightarrow{h\to 0} x_0$. Therefore, by continuity of $\frac{\partial f}{\partial x_j} \Rightarrow (x_0+h^{j-1}+\xi^j h_j e_j) \to \frac{\partial f}{\partial x_j}(x_0)$ and hence $\lim_{h\to 0}\varphi(h) = 0$. Now putting $\varphi_j(h) := \frac{\partial f}{\partial x_j}(x_0+h^{j-1}+\xi^j h_j e_j) - \frac{\partial f}{\partial x_j}(x_0)$ we obtain a function $\Psi : h \to \sum_{j=1}^{n}\Psi_j(h)e_j$, with $\varphi(h) = \langle\Psi(h)|h\rangle$. Then, we can estimate $|\varphi(h)| = \langle\Psi(h)|h\rangle \leq ||\Psi(h)|| \cdot ||h||$, where the last estimation is by Cauchy-Schwartz-Inequality. Hence,

$$0 < \frac{|\varphi(h)|}{||h||} \leq \frac{||\varphi(h)||\,||h||}{||h||} = ||\Psi(h)|| \xrightarrow{h\to 0} 0.$$

□

Corollary 7.50. *Recalling the chain rule for the total differential:* $\mathbb{R}^m \supset \mathcal{U} \xrightarrow{f} \mathcal{V} \subset \mathbb{R}^n \xrightarrow{g} \mathbb{R}^p$, *$f$ differentiable in x_0, g differentiable in $y_0 := f(x_0) \Rightarrow g \circ f$ differentiable in x_0 and $Dg \circ f(x_0) = Dg(x_0) \circ Df(x_0)$. Knowing $[Df(x_0)] = J_f(x_0), [Dg(x_0)] = J_g(y_0)$ and $[Dg \circ f(x_0)] = J_{g\circ f}(x_0)$ we can conclude that $J_{g\circ f}(x_0) = J_g(f(x_0)) \cdot J_f(x_0)$.*

Example 7.51.

1. *$f : \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} x^2+y^2 \\ 2xy \end{pmatrix}, g : \begin{pmatrix} u \\ v \end{pmatrix} \to u+v$, then $g \circ f : x^2+2xy+y^2 = (x+y)^2$ and $J_{g\circ f}(x,y) = \big(2(x+y), 2(x+y)\big) = 2\left(x+y, x+y\right)$. On the other hand: $J_g(u,v) = (1,1)$ and $J_f(x,y) = \begin{pmatrix} 2x & 2y \\ 2y & 2x \end{pmatrix} = 2\begin{pmatrix} x & y \\ y & x \end{pmatrix} \Rightarrow J_g(f(x,y)) \cdot J_f(x,y) = 2\left(1,1\right) \cdot \begin{pmatrix} x & y \\ y & x \end{pmatrix} = 2\left(x+y, x+y\right)$.*

2. $$f : (x,y)\frac{1}{xy}, g : t \to \begin{pmatrix} t \\ t^2 \end{pmatrix}$$

   . $J_f\big((x,y)\big) = \left(\frac{1}{x^2y}, \frac{1}{xy^2}\right), J_g\big((x,y)\big) = \begin{pmatrix} 1 \\ 2t \end{pmatrix} \overset{chain\ rule}{\Longrightarrow} J_{g\circ f}\big((x,y)\big) = J_g(f\big((x,y)\big)) \cdot J_f\big((x,y)\big) = \begin{pmatrix} 1 \\ \frac{2}{xy} \end{pmatrix}\left(-\frac{1}{x^2y}, -\frac{1}{xy^2}\right) = \begin{pmatrix} -\frac{1}{x^2y} & -\frac{1}{xy} \\ -\frac{2}{x^3y^2} & -\frac{2}{x^2y^3} \end{pmatrix}$. On the other hand: $g \circ f\big((x,y)\big) = g(f\big((x,y)\big)) = g(\frac{1}{xy}) = \begin{pmatrix} \frac{1}{xy} \\ \frac{1}{x^2y^2} \end{pmatrix} \Rightarrow J_{g\circ f}\big((x,y)\big) = \begin{pmatrix} -\frac{1}{x^2y} & -\frac{1}{xy} \\ -\frac{2}{x^3y^2} & -\frac{2}{x^2y^3} \end{pmatrix}$.

3. $f : t \to \begin{pmatrix} f \\ \frac{1}{t} \end{pmatrix}, g : (u,v) \to uv \Rightarrow J_f(\begin{pmatrix} 1 \\ -\frac{1}{t^2} \end{pmatrix}), J_g((u,v)) = (v,u) \overset{chain\ rule}{\Longrightarrow}$
   $J_{g\circ f}(t) = J_g \cdot J_f = (\frac{1}{t}, t)\begin{pmatrix} 1 \\ \frac{1}{t^2} \end{pmatrix} = \frac{1}{t} - \frac{1}{t} = 0.$

Definition 7.52. *Let $f : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open, partially differentiable in $x_0 \in \mathcal{U}$. Then, the Jacobean of f in x_0 is also called the gradient of f in x_0 an denoted by $\nabla f(x_0)$. Thus, we have $\nabla f(x_0) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(x_0)e_j = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)$.*

Remark. *We will see that the gradient can be used to compare directional derivatives and we can interpret it geometrically as the direction of maximal increase of f in x_0.*

Proposition 7.53. *Let $f : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open, be continuously differentiable on $\mathcal{U}$. Then, $\forall x \in \mathcal{U}, \forall v \in \mathbb{R}^n$ with $||v|| = 1$ we have $D_v f(x) = \langle \nabla f(x) | v \rangle$.*

Proof. Consider the line $\{x + tv : t \in \mathbb{R}\}$ in $\mathbb{R}^n$, which is parallel to the unit vector v. Since $\mathcal{U}$ is open, $\exists \epsilon > 0$ so that $x + tv \in \mathcal{U}, \forall t \in (-\epsilon, \epsilon)$. Define $\varphi : (-\epsilon, \epsilon) \to \mathcal{U}; \varphi(t) := x + tv$ and consider $F := f \circ \varphi : (-\epsilon, \epsilon) \to \mathbb{R}$. Applying the chain rule in this situation yields $J_f(t) = J_F(\varphi(t)) \cdot J_g(t) = \nabla f(\varphi(t)) \cdot \begin{pmatrix} \varphi_1'(t) \\ \vdots \\ \varphi_n'(t) \end{pmatrix} \overset{\text{definition}}{=} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \langle \nabla f((\varphi(t))) | v \rangle$. In particular for $t = 0$, $J_F(0) = \langle \nabla f(x) | v \rangle$. By definition of hte directional derivative, we have that $D_v f(x) = \lim_{t\to 0} \frac{f(x+tv)-f(x)}{t} = \lim_{t\to 0} \frac{F(t)-f(0)}{t} = F'(0)$. $\square$

Remark. *The assumption of continuity of partial derivatives is not necessary.*

Corollary 7.54. *In the situation of the last proposition the gradient gives the direction of maximal increase in each point in $x \in \mathcal{U}$. This means that $D_v f(x)$ is maximal for $v = \frac{\nabla f(x)}{||\nabla f(x)||}$. Furthermore, if $\nabla f(x)$ is not zero $D_v f(x) = 0$ iff $v \perp \nabla f(x)$, and always if $\nabla f(x) = 0$.*

Proof. Assume $\nabla f(x) \neq 0$ (otherwise the claim is obvious). Thus $\theta := \angle(\nabla f(x), v)$ is well defined and we obtain $\cos(\theta) = \frac{\langle \nabla f(x) | v \rangle}{||\nabla f(x)|| ||v||}$ and thus $D_v f(x) = \langle \nabla f(x) | v \rangle = ||\nabla f(x)|| \cos(\theta)$, which is maximal iff $\cos(\theta) = 1 \Leftrightarrow \theta = \frac{\pi}{2}$. Also $v \perp f(x) \Leftrightarrow \theta = \frac{\pi}{2} \Leftrightarrow \cos(\theta) = 0 = \langle \nabla f(x) | v \rangle$. $\square$

Remark. *We can visualize this statement by the "Hill Billy" example; here the gradient corresponds to the direction of steepest ascend $\perp$ constant height lines. Also a river bed $g(t)$ satisfies $\varphi'(t) = -\nabla f(\varphi(t))$.*

Example 7.55. 1. *the paraboloid: $f : \mathbb{R}^2 \to \mathbb{R}, f((x,y)) = x^2 + y^2$. What is the direction $\frac{v}{||v||}$ and magnitude $||v||$ of maximal increase v of f in $(1,1)$?*

$$\nabla f((1,1)) = (2,2) \Rightarrow \frac{v}{||v||} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \wedge ||v|| = 2\sqrt{2}.$$

2. *$f : \mathbb{R}^2 \to \mathbb{R}, f((x,y)) := x^2 + e^{xy}\sin(y)$. Some question; Find the direction $\frac{v}{||v||}$ and magnitude $||v||$ of maximal increase v of f in $(0,1)$.*

$$\nabla f = \left(2xy + ye^{xy}\sin(y), x^2 + xe^{xy}\sin(y) + e^{xy}\cos(y)\right)$$
$$\Rightarrow \nabla f((1,0)) = (0,2) \Rightarrow \frac{v}{||v||} = (0,1) \wedge ||v|| = 2.$$

Lecture 8

Example 8.56. *1. $f : \mathbb{R}^3 \to \mathbb{R}, f(x, y, z) := x^4 - 2xy + z^3$ find directional derivative of f at $1, 0, 1)$ in direction $(-3, 6, -2)$.*

2. What is the direction of maximal increase of f in $(1, 0, 1)$ and what is the rate of change in that direction.

The proof is left as an exercise to the reader.

Definition 8.57. *Let $\mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open be some function and consider $l \in \mathbb{R}$. The level set (also called contour-line, ...) of f at level l is defined as $\mathcal{N}_f(l) := \{x \in \mathcal{U} : f(x) = l\}$.*

Proposition 8.58. *Let $f : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open be differentiable on w. Then, $\nabla f(x)$ with is orthogonal to $\mathcal{N}_f(f(x_0)$ i.e. for any continuous differentiable mapping $\varphi : (-\epsilon, \epsilon) \to \mathcal{N}_f(f(x_0)$, with $\varphi(0) = x_0$ we have $\langle \nabla f(x_0) | \varphi'(0) \rangle$.*

Proof. Analogously to the proof of $D_v(x_0) = \langle \nabla f(x_0) | v \rangle$ we apply the chain rule suitable. Define $F : (-\epsilon, \epsilon) \to \mathbb{R}, F(t) := f(\varphi(t))$. Then, since φ only has values in $\mathcal{N}_f(x_0)$ it follows that f is hte constant function $f(x_0)$ thus $F'(t) = 0 \quad \forall t \in (-\epsilon, \epsilon)$. But the chain rule for $F = f \circ \varphi$ gives $0 = F'(t) = \left(\frac{\partial f}{\partial x_1}(\varphi(t)), \ldots, \frac{\partial f}{\partial x_n}(\varphi(t))\right) \begin{pmatrix} \varphi_1'(t) \\ \vdots \\ \varphi_n'(t) \end{pmatrix} = \langle \nabla f(\varphi(t)) | \varphi'(t) \rangle$. In particular, for $t = 0$, we obtain $F'(0) = 0$. □

Remark. *We have seen:*

- $\nabla f(x_0) = J_f(x_0)$ *for* $w = 1$.
- $D_v f(x_0) = \langle \nabla f(x_0) | v \rangle$.

In general, for $f : \mathcal{U} \to \mathbb{R}^m$, we have $J_f(x_0) \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \ldots & \frac{\partial f_1}{\partial x_n} \\ \ldots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \ldots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$ *for $v \in \mathbb{R}^n$, $||v|| = 1$.*

Definition 8.59 (higher order derivative)**.** *$f : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open. Now iff f is partially differentiable, we say f is 1-times partially differentiable. If f is k-times differentiable for $k \in \mathbb{N}$, we call f $k+1$-times partially differentiable iff the k-th partial derivatives are all partially differentiable.*

Notation

$\forall$finite sequences$i_1, i_2, \ldots, i_k$ we write $\frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_n}} := \frac{\partial}{\partial x_{i_1}}\left(\frac{\partial}{\partial x_{i_2}}\left(\ldots\left(\frac{\partial f}{\partial x_{i_n}}\right)\right)\right)$.
In particular for $k = 2$, we have $\frac{\partial f}{\partial x_i \partial x_j} := \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) \qquad \forall i, j \in \{1, 2, \ldots, n\}$.

Proposition 8.60 (Schwartz theorem)**.** *If $f : \mathcal{U} \to \mathbb{R}, \mathcal{U} \in \mathbb{R}^n$ open, is k-times partially continuously differentiable (of class $C^k, k \in \mathbb{N}$), then the order of taking partial derivatives in $\frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_n}}$ is irrelevant. I particular for $k = 2$, we have $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2}{\partial x_j \partial x_i}, \forall i, j \in \{1, 2, \ldots, n\}$.*

Proof sketch. Apply mean value theorem twice after reducing to the case $n = 2, k = 2$. □

Example 8.61.

1. *assignment 2*
2. *"Counterexample partially differentiable, but not continuous" ($k = n = 2$): Consider $f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) := \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & (x, y) \neq 0 \\ 0 & otehrwise \end{cases}$ This is twice partially differentiable but not of class C^2, since the second order partial derivative is not continuous in $(0, 0)$. One can check that $\frac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \frac{\partial^2 f}{\partial y \partial x}$.*

Lecture 9

The inverse function theorem

Motivation In analysis of functions in one variable we had that

Theorem 9.62. *Given*

1. *$f : I \to \mathbb{R}$. $I \subset \mathbb{R}$ (I being an open interval)*
2. *f is continuously differentiable on its domain I*
3. *$f'(x) \neq 0$ $\forall x \in I$ (i.e. f is strictly increasing or decreasing)*

Then f is invertible (injective), its inverse f^{-1} is also continuously differentiable and

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)}$$

Remark. *Can be derived as*

$$f^{-1}(f(x)) = x \xrightarrow{differentiate} (f^{-1})'(f(x)).f'(x) = 1 \Leftrightarrow (f^{-1})'(f(x)) = \frac{1}{f'(x)}$$

We wish to generalize this statement to functions of multiple variables.

Local definition 9.62.1. *We call $f|_u$ a restriction of f at u; we only consider f on the subset u of the domain and its image from u. More precisely, using the definition of functions as certain subsets of cartesian products, we define $f|_u := \{(x, y) \in f \text{ s.t. } x \in u\}$.*

Theorem 9.63 (Inverse function theorem for functions of several variables)**.** *Given*

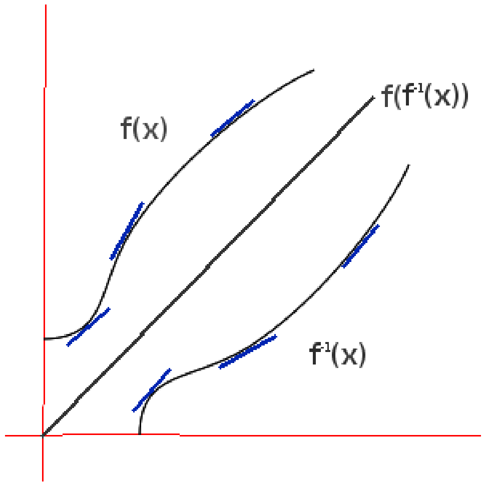

Figure 1.1: Geometric idea of inverse function theorem

1. *$f : u \to \mathbb{R}^n . u \subset \mathbb{R}^n$ (u is an open n-ball)*
2. *f is differentiable on u, or, equivalently, $f|_u$ is differentiable*
3. *$Df : u \to L(\mathbb{R}^n)$ is continuous.*
4. *$Df(x_0)$ is invertible $x_0 \in u$, or, equivalently, the Jacobean $J_f(x_0)$ is invertible in $M^{n\times n}$*

Then $\exists$ neighbourhood $u_0 \subset u$ of x_0 so that $f|_{u_0}$ is bijective onto its image $v_0 = f(u_0)$ (invertible), the inverse $g = (f|_{u_0})^{-1}$ is also differentiable on v_0 and $\forall y = f(x) \in v_0$

$$Dg(y) = (Df(x))^{-1}$$

or, equivalently

$$J_g(y) = (J_f(x))^{-1}$$

Proof sketch:

We will show this in three major steps:

I $\exists u_0$ neighbourhood of x_0 such that $f|_{u_0}$ is bijective with inverse $g : v_0 \to u_0$, where $v_0 = f(u_0)$ (using *Banach fixed point theorem*)

II v_0 is open

III g is differentiable on v_0 and $D_g(y) = (Df(g(y)))^{-1}, \forall y \in v_0$

Firstly, we need to define some preliminaries and prove the Banach fixed point theorem.

Definition 9.64. *A function defined from metric space $(\mathcal{X}, d)$ to itself $\kappa : \mathcal{X} \to \mathcal{X}$ is called a contraction if*

$$d(\kappa(x), \kappa(y)) < cd(x, y).\ \forall x, y \in \mathcal{X}.\ c \in [0, 1)$$

Remark. *Intuitively, a contraction κ is a function, whose the preimage always changes faster than the image; its graph would be 'flatter' the smaller c is.*

Theorem 9.65 (Banach fixed point theorem)**.** *Let $(\mathcal{X}, d)$ be a complete metric space. If $\kappa : \mathcal{X} \to \mathcal{X}$ is a contraction, then κ has a unique fixed point $x \in \mathcal{X}$, i.e. a point $x \in \mathcal{X}$ where $\kappa(x) = x$.*

Remark. *The intuitive picture of Banach fixed point theorem: Because κ is a contraction (the preimage x is always changing faster than the image $\kappa(x)$), we know that moving towards either its maximal point (e.g: $+\infty$) or minimal point (e.g: $-\infty$) the value of x will eventually catch up to $\kappa(x)$*

Proof.

1. Take some $x_1 \in \mathcal{X}$
2. Define sequence (x_p) inductively such that $x_{p+1} = \kappa^p(x_1)$. $p \in \mathbb{N}$
3. Define inequality
$$d(x_{p+1}, x_p) \leq c^{p-1} d(x_2, x_1).\ p \in \mathbb{N}.\ c \in [0, 1) \tag{1.6}$$
where c is the contraction constant for κ.
4. Proving above inequality by induction:

Base case $(p = 1)$:
$$d(x_{1+1}, x_1) \leq c^{1-1} d(x_2, x_1)$$
$$d(x_2, x_1) \leq (c^0) d(x_2, x_1) \Rightarrow d(x_2, x_1) = d(x_2, x_1)$$

Inductive step $(p \to p + 1)$
$$d(x_{p+2}, x_{p+1}) = d(\kappa(x_{p+1}), \kappa(x_p))$$
Since κ is a contraction
$$= d(\kappa(x_{p+1}), \kappa(x_p)) \leq c d(x_{p+1}, x_p)$$
Activate induction
$$\leq c c^{p-1} d(x_2, x_1) = c^p d(x_2, x_1)$$

5. We will show the sequence (x_p) is cauchy. Let $m, n \in \mathbb{N}$. $m > n$
6. By triangle inequality
$$d(x_m, x_n) \leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n)$$
By inequality (1.6)
$$\leq c^{m-2} d(x_2, x_1) + c^{m-3} d(x_2, x_1) + \cdots + c^{n-1} d(x_2, x_1)$$

$$= c^{n-1} d(x_2, x_1) \sum_{k=0}^{m-n-2} c^k$$

$$\leq c^{n-1} d(x_2, x_1) \sum_{k=0}^{\infty} c^k = c^{n-1} d(x_2, x_1) \frac{1}{1-c}$$

So finally

$$d(x_m, x_n) = c^{n-1} \frac{d(x_2, x_1)}{1-c}$$

7. Let $\epsilon = c^{N-1} \frac{d(x_2,x_1)}{1-c} > 0 \Leftrightarrow N = \lceil \log_c(\frac{\epsilon(1-c)}{d(x_2,x_1)}) \rceil + 1$

8. ϵ is arbitrary, so the sequence (x_p) is cauchy. And since the metric space it is defined on, $(\mathcal{X}, d)$, is complete, it converges to some $x^* \in \mathcal{X}$.

9. κ is a contraction $\Rightarrow$ Lipschitz (with constant < 1) $\Leftrightarrow$ Lipschitz continuous. Thus we can have

$$\kappa(\lim_{p\to\infty} x_p) = \lim_{p\to\infty} x_{p+1}$$

$$\kappa(x^*) = x^*$$

So x^* is a fixed point of κ

10. Suppose there exists another $y \in \mathcal{X}$ such that $\kappa(y) = y$, then

$$0 \leq d(x^*, y) = d(\kappa(x^*), \kappa(y)) \leq c d(x^*, y)$$

Since $|c| < 1$, $d(x^*, y) = 0$

$$0 \leq d(\kappa(x^*), \kappa(y)) \leq 0$$

$$x^* = \kappa(x^*) = \kappa(y) = y$$

So x^* is the unique fixed point.

□

Proof of the inverse function theorem. :

I '$\exists u_0$ neighbourhood of x_0 such that $f|_{u_0}$ is bijective with inverse $g : v_0 \to u_0$, where $v_0 = f(u_0)$'

(a) $A = Df(x_0) \in L(\mathbb{R}^n)$ is invertible (by assumption), $\Rightarrow ||A|| \neq 0 \Leftrightarrow ||A^{-1}|| \neq 0$. Define

$$\lambda = \frac{1}{2||A^{-1}||} > 0$$

(b) The differential map $Df : u \to L(\mathbb{R}^n)$ is continuous in x_0 (by assumption); there exists (for the previously defined $\lambda > 0$) an open neighbourhood $u_0 \subset u$ of x_0 so that

$$||Df(x) - Df(x_0)|| < \lambda \, \forall x \in u_0 \tag{1.7}$$

(c) $\forall y \in \mathbb{R}^n$ define a map $\kappa = \kappa_y : u \to \mathbb{R}^n$ by

$$\kappa(x) = x + A^{-1}(y - f(x)) \tag{1.8}$$

Remark. *Contrary to what the notation suggests, we do not have $f(x) = y$ always; this is the case exactly when x is a fixed point of κ. Note also here that κ is not a contraction yet either, but the plan to show that κ is a contraction.*

(d) Finding the derivative of κ by the chain rule

$$D\kappa(x) = I + A^{-1} \circ (-Df(x)) = I - A^{-1} \circ Df(x) = A^{-1} \circ (A - Df(x))$$

(e) By properties of the operator norm we get $\forall x \in u_0$

$$||D\kappa(x)|| = ||A^{-1} \circ (A - Df(x))|| \leq ||A^{-1}|| \, ||A - Df(x)||$$

recall that $A = Df(x_0)$, and by inequality (1.7) is

$$< ||A^{-1}||\lambda = ||A^{-1}|| \frac{1}{2||A^{-1}||} = \frac{1}{2} < 1$$

(f) By proposition, since the differential is bounded $\Rightarrow$ map is Lipschitz. We have that $||\kappa(u) - \kappa(v)|| \leq \frac{1}{2}||u - v||$, $\forall u, v \in u_0 \Rightarrow \kappa$ is a contraction.

(g) By Banach fixed point theorem, $\forall y \in \mathbb{R}^n$, there is at most one fixed point $x \in u_0$ of $\kappa = \kappa_y \Rightarrow f(x) = y$. In particular, $\forall y \in v_0 = f(u_0)$ there exists exactly one [1] $x \in u_0$ with $f(x) = y \Rightarrow f|_{u_0}$ is injective $\Rightarrow f|_{u_0}$ is bijective.

II 'v_0 is open'

(a) Let $y_1 \in v_0$ and let $x_1 = f^{-1}(y_1)$.

(b) Since u_0 is open, $\exists \rho > 0$ so that closed ball $\overline{B(x_1, \rho)} = \{u \in \mathbb{R}^n : ||x_1 - u|| \leq \rho\} \subset u_0$

(c) We will show that $B(y_1, \lambda\rho) \subset v_0$ which would imply that v_0 is open; we need to prove that

$$||y - y_1|| < \lambda\rho = \frac{\rho}{2||A^{-1}||} \Rightarrow y \in v_0 = f(u_0)$$

(d) Let $y \in B(y_1, \lambda\rho)$. For $\kappa = \kappa_y$, we can have

$$||\kappa(x_1) - x_1|| = ||A^{-1}(y - f(x_1))||$$

by property of the operator norm is

$$\leq ||A^{-1}|| \, ||y - f(x_1)|| < ||A^{-1}||\lambda\rho = ||A^{-1}|| \frac{\rho}{2||A^{-1}||} = \rho/2$$

thus, $\forall x \in \overline{B(x_1, \rho)} := \overline{B}$,

$$||\kappa(x) - x_1|| = ||\kappa(x) + (-\kappa(x_1) + \kappa(x_1)) - x_1||$$

by triangle inequality

$$\leq ||\kappa(x) - \kappa(x_1)|| + ||\kappa(x_1) - x_1|| \leq \frac{\rho}{2} + \frac{\rho}{2} = \rho$$

[1] Banach fixed point theorem says that this implies that there exists *at most one* x, but I think it's exactly one since it works for all y.

(e) But this means that $\kappa(\overline{B}) \subset \overline{B} \Rightarrow \kappa$ is a contraction of a complete metric space $\overline{B}$ with the euclidean metric.

(f) By Banach fixed point theorem, $\exists! x \in \overline{B} : \kappa(x) = x \Leftrightarrow f(x) = y \Leftrightarrow y \in v_0 \Rightarrow v_0$ is open.

III 'g is differentiable on v_0 and $D_g(y) = (Df(g(y)))^{-1}, \forall y \in v_0$'

(a) Let $g = (f|_{u_0})^{-1}$

(b) Since $f|_{u_0}$ is bijective, we have a one to one correspondence

$$v_0 \to \begin{cases} y & \leftrightarrow \quad x = g(y) \in u_0 \\ y + k & \leftrightarrow \quad x + h = g(y + k) \in u_0 \end{cases}$$

(c) With $\kappa = \kappa_y$ defined in (1.8), and knowing $f(x) = y, f(x+h) = y+k$, we now have

$$\begin{aligned} \kappa(x + h) - \kappa(x) &= x + h + A^{-1}(y - f(x + h)) - x - A^{-1}(y - f(x)) \\ &= h + A^{-1}(f(x) - f(x + h)) = h - A^{-1}k \end{aligned}$$

(d) We can have

$$||h|| = ||x + h - x||$$

since κ is a contraction with constant $\frac{1}{2}$, it is

$$\geq 2||\kappa(x + h) - \kappa(x)|| = 2||h - A^{-1}k||$$

(e) On the other hand

$$||h|| = ||h - A^{-1}k + A^{-1}k||$$

by triangle inequality is

$$\leq ||h - A^{-1}k|| + ||A^{-1}k||$$

by the previous inequality is

$$\leq \frac{||h||}{2} + ||A^{-1}k||$$

(f) Thus, $||A^{-1}k|| \geq \frac{||h||}{2}$ and finally

$$||h|| \leq 2||A^{-1}k|| \leq 2||A^{-1}|| \; ||k|| = \frac{2||k||}{2\lambda}$$

$$\Rightarrow ||h|| \leq \frac{||k||}{\lambda} \tag{1.9}$$

(g) We need to show that $Df(x)$ is invertible. Recall from property 6 of the operator norm that if $||B - A|| < \frac{1}{||A^{-1}||} \Rightarrow B$ is invertible. We have by inequality (1.7) (recall $A = Df(x_0)$)

$$||Df(x) - A|| < \lambda = \frac{1}{2||A^{-1}||} < \frac{1}{||A^{-1}||}$$

$\Rightarrow Df(x)$ is invertible.

(h) Now to identify the differential of g in $y = f(x)$, let $B = (Df(x))^{-1}$ (note $h = Ih = (BDf(x))h$)

$$\begin{aligned} g(y+k) - g(y) - Bk =& x + h - x - Bk = h - B(y + k - y) \\ =& BDf(x)h - B(f(x+h) - f(x)) \\ =& -B(f(x+h) - f(x) - Df(x)h) \end{aligned}$$

(i) Then g is differentiable if following limit equals 0 as $k \to 0$

$$0 \leq \frac{||g(y+k) - g(y) - Bk||}{|k|} = \frac{1}{||k||}||B(f(x+h) - f(x) - Df(x)h)||$$

$$\leq \frac{||B||}{||k||}||f(x+h) - f(x) - Df(x)h||$$

by inequality (1.9) is

$$\leq \frac{||B||}{\lambda}\frac{f(x+h) - f(x) - Df(x)h}{||h||}$$

(j) As $k \to 0, h \to 0$ by inequality (1.9), and as f is differentiable $\frac{||B||}{||k||}||f(x+h)-f(x)-Df(x)h|| \xrightarrow{h\to 0} 0$. $\frac{||B||}{\lambda}$ is a constant (by definition, $\lambda > 0$). So, finally,

$$\lim_{k\to 0}\frac{||g(y+k) - g(y) - Bk||}{|k|} = 0$$

thus g is differentiable on v_0, and the differential is $B = (Df(x))^{-1}$.

□

Example 9.66. *of a map that is locally invertible in every point but not globally injective is a map of polar coordinates in the plane* $f : (0,\infty) \times \mathbb{R} \to \mathbb{R}^2$ *defined by* $f(r,\theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$ *Take* $r > 0, \theta \in \mathbb{R}$*, then* $J_f(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$. $\det J_f = r\cos^2\theta + r\sin^2\theta \Rightarrow f$ *is locally invertible in* (r,θ) *But* $f(r,\theta) = f(r,\theta + 2\pi)$*, so* f *is not injective.*

Remark. *One can strengthen the statement of the theorem by assuming f is continuously differentiable on all of u, then one can show (using property 7 of operator norm) that g is also continuously differentiable.* [2]

Lecture 10

Implicit function theorem

Motivation

Let F be a differentiable function in $x, y \in \mathbb{R}$. Given $F(x,y) = 0$, on can ask, when there is a differentiable (locally defined) function $g(x) = y$, s.t. $F(x, g(x)) = 0$. $F(x,y) = 0$ could for instance describe a level set.

[2]Refer to Rudin

We need that $\frac{\partial F}{\partial y}(x,y) \neq 0$, then (in $\mathbb{R}$)

$$g'(x_0) = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}.$$

Example 10.67. *$F(x,y) = 0$, for $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) = x^2 + y^2 - r^2 \Rightarrow \frac{\partial F}{\partial x} = 2x$, $\frac{\partial F}{\partial y} = 2y$, $\frac{\partial F}{\partial y} = 0 \Leftrightarrow x = \pm r$, thus for $x \neq \pm r$, we obtain $g'(x) = -\frac{x}{y}$.*

Before stating the general implicit function theorem, let us frst consider a linearized version.

Notation:

For $\mathbb{R}^m \ni x = (x_1, x_2, \ldots, x_m), \mathbb{R}^n \ni x = (y_1, y_2, \ldots, y_n)$, we write $(x,y) := (x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n) \in \mathbb{R}m + n$. For $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^m)$, we define $A_x \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $Ay \in L(\mathbb{R}^m)$, by $\begin{cases} A_x(u) := A(u,0) & \forall u \in \mathbb{R}^n \\ A_y(v) := A(0,v) & \forall v \in \mathbb{R}^m \end{cases}$. Thus, $A(u,v) = A_x(u) + A_y(v) \qquad \forall (u,v) \in \mathbb{R}n + m.$

Theorem 10.68. *Let $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ be a linear map s.t. $A_y \in \mathcal{I}(\mathbb{R}^m)$, i.e. is invertible. Then $\forall u \in \mathbb{R}^n . \exists^1 v \in \mathbb{R}^m$ so that $A(u,v) = 0$, namely $v := -(A_y^) - 1 \circ A_x(u)$.*

Proof.

$$\begin{aligned} & A(u,v) = 0 \\ \Leftrightarrow & A_x u + A_y v = 0 \\ \Leftrightarrow & A_y^{-1} A_x u + A_y^{-1} A_y v = 0 \\ \overset{A_y \in \mathcal{I}(\mathbb{R}^m)}{\Longleftrightarrow} & A_y^{-1} A_x u + v = 0 \\ \Leftrightarrow & v = -A_y^{-1} A_x u. \end{aligned}$$

□

Theorem 10.69 (implicit function theorem)**.** *Let $F : \mathcal{U} \to \mathbb{R}^m, \mathcal{U} \subset \mathbb{R}n + m$ open be continuously differentiable s.t. $F(x_0, y_0) = 0$, for some $x_0, y_0 \in \mathcal{U}$, $x_0 \in \mathbb{R}^n, y_0 \in \mathbb{R}^m$, where $A := Df(x_0, y_0)$ and $A_y \in \mathcal{I}(\mathbb{R}^m)$. Then $\exists x \in \mathcal{V}_0$ neighborhood of x_0, $\exists \mathcal{U}_0 \subset \mathcal{U}$ neighborhood of (x_0, y_0), so that $\forall x \mathcal{V}_0 \exists^1 y \in \mathbb{R}^m :$ $(x,y) \in \mathcal{U}_0 \wedge F(x,y) = 0$. If $g(x)$ denotes this y by, $g : \mathcal{V}_0 \to \mathbb{R}^m$ is continuously differentiable and satisfies $F(x, g(x)) = 0$ and $Dg(x_0) = -A_y^{-1} \circ A_x$.*

Lecture 11

Proof. Define $f : \mathcal{U} \to \mathbb{R}^{n+m}, f(x,y) := (x, f(x,y))$, with $\mathcal{U} \subset \mathbb{R}^n$ open. f is continuously differentiable by definition. Showing $Df(x_0, y_0) \in \mathcal{I}(\mathbb{R}n + m)$, we obtain using f differentiable in (x_0, y_0) that:

$$F(x_0 + h, y_0 + k) = F(x_0, y_0) + A(n,k) + \varphi(n,k).$$

Looking at f in x_0, y_0):

$$\begin{aligned}
&f(x_0+h, y_0+k) - f(x_0, y_0)\\
\overset{\text{def. of } f}{=}&(x_0+h-x_0, y_0+k)\\
\overset{\text{def. of } F}{=}&(h, A(n,k)+\varphi(h,k))\\
=&(h, A(n,k)) + (0, \varphi(n,k)).
\end{aligned}$$

So $Df(x_0, y_0) = (h, A(h,k))$. Now we need to show that $Df(x_0, y_0)(n,k) = (h, A(n,k))$ is invertible. We will show this, by proving $Df(x_0, y_0)(h,k) = 0 \Rightarrow (h,k) = (0,0)$, then for $h = 0 \overset{\text{A inv.}}{\Longrightarrow} A(h,k) = 0 \Rightarrow k = 0$. Then the invertibility of A follows (by linear algebra). Now applying the inverse function theorem 9.63 to f in (x_0, y_0): $\exists \mathcal{U}_0 \subset \mathcal{U}$open neighbourhood of (x_0, y_0) so that $f(\mathcal{U}) \subset \mathcal{I}(\mathbb{R}n+m), (f|\mathcal{U}_0)^{-1}$ is continuous. differentiable on $f(\mathcal{U}_0)$ and $D(f|\mathcal{U}_0)^{-1}(f(x,y)) = (Df(x,y)) \qquad \forall (x,y) \in \mathcal{U}_0$.

Define $\mathcal{V}_0 := \{x \in \mathbb{R}^n : (x,0) \in f(\mathcal{U}_0)\}$, $\mathcal{V}_0$ open since $f(\mathcal{U}_0)$ is (since f is open and $\mathcal{U}_0$ is open).

What does it mean that $f|\mathcal{U}_0 : \mathcal{U}_0 \to f(\mathcal{U}_0)$ is invertible? It means:

$$\exists^1 (x,y) \in \mathcal{U}_0 \text{ s.t. } f(x,y) = (x,z)$$

or equivalently (definition of $\mathcal{V}_0$ and looking at $z = 0$):

$$\Leftrightarrow \forall x \in \mathcal{V}_0 \, \exists^1 y \in \mathbb{R}m \text{ s.t. } F(x,y) = 0.$$

We will call this unique y just $g(x)$ and obtain a function $g : \mathcal{V}_0 \to \mathbb{R}^m$ so that $F(x, g(x)) = 0, \qquad \forall x \in \mathcal{V}_0$.

Why is g continuously differentiable? By definition of f, we have $f(x, g(x)) = (x, F(x, g(x))) = (x, 0) \qquad \forall x \in \mathcal{V}_0$. Since $(f|\mathcal{U}_0)^{-1}$ continuously differentiable and $(x, g(x)) = (f|\mathcal{U}_0)^{-1}(x, 0) \qquad \forall x \in \mathcal{V}_0$. We can compute the differential of g using chain rule: Defining $\varphi : \mathcal{V}_0 \to \mathbb{R}n+m, \varphi(x) := (x, g(x))$ we get on the one hand

$$D\varphi(x)(u) = (u, Dg(x))(u).$$

and on the other hand:

$$\forall x \in \mathcal{V}_0.\ 0 = F(x, g(x)) = F \circ \varphi(x),$$

so by chain rule:

$$0 = DF(x, g(x)) \circ D\varphi(x).$$

Finally, applying both equalities, we obtain: $0 = Df(x_0, y_0) \circ D\varphi(x_0)u = A \circ (D\varphi(x_0)u) = A(u, Dg(x_0)(u)) = A(u, 0) + A(0, Dg(x_0)(u)) \overset{\text{def. of } A_x, A_y}{=} A_x(u) + A_y(Dg(x_0)(u))$.

Since $A_y \in \mathcal{I}(\mathbb{R}^m)$ we get $Dg(x_0) = -A_y^{-1} \circ A_x$. □

Lecture 12

Lecture 13

local extrema with constraints

Lagrange multipliers: $\nabla f = \lambda \nabla g$

Idea:

Imagine we are running up a mountain;
Q: What happens in the point where the trajectory has a local maximum?
A: The trajectory is parallel to the level lines in that local maximum

Example 13.70. *Look at the "mathematical mountain", say the graph of the function* $f : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) := 16 - x^2 - 4y^2$. *Now consider the "trajectory"* $r : (0,\infty) \to \mathbb{R}^2, r(t) := \begin{pmatrix} t \\ \frac{1}{2} \end{pmatrix}$. *We will look for a* $t_0 > 0$ *s.t.* $f(r(t_0))$ *has a local maximum and verify this statement, in the following steps:*

(i) Find all $t_0\colon h := f \circ r : (0,\infty) \to \mathbb{R}$ *has a local extremum (we will actually only check for a critical point) at* t_0.

(ii) Check that $\nabla f(r(t_0)) \perp r'(t_0)$

(ii') We already know that ∇f *is orthogonal to the level set of* f *i.e.* $\nabla f(r(t_0)) \perp N_f(f(r(t_0)))$, *but we want to check it anyway*

(iii) Write path r *"implicitly" as level set of* $g : (0,\infty)^2 \to \mathbb{R}$ *and check* $\nabla f(r(t_0)) \perp N_g(r(t_0)) = N_g(0)$

(iv) Find $\lambda \in \mathbb{R} \setminus \{0\}\colon \nabla f(r(t_0)) = \lambda \nabla g(r(t_0))$.

(I) $h : (0,\infty) \to \mathbb{R},\ h(f) = f(r(t)) = f((t, \frac{1}{t})) = 16 - t^2 - \frac{4}{t^2}$. *Therefore,* $h'(t) = -2t + \frac{8}{t^3} \Rightarrow (f \text{ critical}) \Rightarrow -2t = \frac{8}{t^3} \Rightarrow t = \pm\sqrt{2} \overset{t \in (0,\infty)}{\Longrightarrow} t = \sqrt{2}$. *It can easily be verified, that this critical point corresponds to a local maximum.*

(II) $\nabla f(x,y) = \begin{pmatrix} -2x \\ -8y \end{pmatrix}, \nabla f(r(t_0)) = f\left(\begin{pmatrix} \sqrt{2} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\right) = \begin{pmatrix} -2\sqrt{2} \\ -4\sqrt{2} \end{pmatrix}$. *The velocity vector of* r *in* t_0 *is:* $r'(t) = \begin{pmatrix} 1 \\ -\frac{1}{t^2} \end{pmatrix} \Rightarrow r'(t_0) = r'(\sqrt{2}) = \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}$. *Is* $\begin{pmatrix} -2\sqrt{2} \\ -4\sqrt{2} \end{pmatrix} \perp \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}$? *Yes, since* $\left\langle \begin{pmatrix} -2\sqrt{2} \\ -4\sqrt{2} \end{pmatrix} \middle| \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix} \right\rangle = -2\sqrt{2} + 2\sqrt{2} = 0$.

(II') $f(r(t_0)) = h(t_0) = h(\sqrt{2}) = 12$. *We yield:* $N_f(12) = \{(x,y) \in \mathbb{R}^2 \colon f(x,y) = 12\}$, *the equation of an ellipse* $16 - x^2 - 4y^2 = 12 \Leftrightarrow x^2 + 4y^2 = 4 \Rightarrow r(\sqrt{2}) = r(t_0)$.

Now we only need to check: $\nabla f(r(t_0)) \perp$ *to the velocity vector of* f *on the ellipse* $x^2 + 4y^2 = 4$ *at some point. We can calculate this velocity vector*

γ' and find the desired point:

$$y^2 = \frac{4x^2}{4} \Rightarrow y = \sqrt{1 - \frac{x2}{4}},$$

$$\Rightarrow \gamma(t) := \begin{pmatrix} 1 \\ \sqrt{1 - \frac{x^2}{4}} \end{pmatrix},$$

$$\Rightarrow \gamma'(t) = \begin{pmatrix} 1 \\ \frac{1}{2\sqrt{1-\frac{t^2}{4}}(-\frac{t}{2})} \end{pmatrix} = \begin{pmatrix} 1 \\ -\frac{t}{4\sqrt{1-\frac{t^2}{4}}} \end{pmatrix}$$

$$\Rightarrow \gamma'(\frac{1}{2}) = \begin{pmatrix} 1 \\ -\frac{\sqrt{2}}{4\sqrt{1-\frac{1}{2}}} \end{pmatrix} = \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}.$$

We can now check:

$$\left\langle -2\sqrt{2} \middle| -4\sqrt{2} \right\rangle.$$

(III) *Rewriting r we obtain the equation of a hyperbola $xy = 1$, so we choose $g : (0, \infty)^2 \to \mathbb{R}, g(x, y) = xy - 1$. Then, $N_g(0)$ is simply $r((0, \infty))$. Thus, $\nabla g(x, y) = \begin{pmatrix} y \\ x \end{pmatrix}, \nabla g(r(t_0)) = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \sqrt{2} \end{pmatrix}$ and we only need to check that $\nabla g(r(t_0)) \perp r'(t_0)$:*

$$\left\langle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \sqrt{2} \end{pmatrix} \middle| \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix} \right\rangle = \frac{1}{2} - \frac{2}{2} = 0.$$

(IV) *$\nabla f(r(t_0)) = \begin{pmatrix} -2\sqrt{2} \\ -4\sqrt{2} \end{pmatrix}, \nabla g(r(t_0)) = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \sqrt{2} \end{pmatrix} \Rightarrow \lambda = -4$ and $\nabla f = \lambda \nabla g$ as desired.*

Now let us make this idea a rigorous statement.

Definition 13.71. *For $f, g : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open, we say that f has a local maximum (minimum) under the constraint $g(x) = 0$ at t_0, iff*

$$\begin{cases} x_0 \in N_g(0) := \{x \in \mathcal{U} \colon g(x) = 0\} \\ \exists \text{neighbourhood } \mathcal{U}_0 \subset \mathcal{U} \text{ of } x_0 \text{ such that } f(x) \le f(x_0) (f(x) \ge f(x_0), \text{ resp.}) & \forall x \in \mathcal{U} \cap N_g(0) \end{cases}.$$

Remark. *We will look for such local extrema under constraints for the values of functions in such points.*

Theorem 13.72 (method of Lagrange multipliers)**.** *Let $f, g : \mathcal{U} \to \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open be continuously differentiable functions. Assume f has a local extremum at $x_0 \in \mathcal{U}$ under constraint $g(x) = 0$. Then, assuming that $\nabla f(x_0), \nabla g(x_0) \neq 0$, if follows that $\exists \lambda \in \mathbb{R} \setminus \{0\}$ with $\nabla f(x_0) = \lambda \nabla g(x_0)$.*