

Analysis II – Lecture Notes

Colin Rothgang

April 16, 2017

Contents

1	Differentiation	2
	Lecture 1	2
	Motivation	2
	Norms, Metrics and Banach and Hilbert spaces	2
	Lecture 2	5
	Lecture 3	6
	Differentiability	6
	Lecture 4	9
	Lecture 5	10
	Lecture 6	13
	Lecture 7	14
	Lecture 8	17
	The inverse function theorem	17

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 $L(q_1, q_2, \dots, q_n, \dot{q}_1, \dot{q}_2, \dots, \dot{q}_n)$, where $G_i = g_i(t), \forall i \in \{1, \dots, n\}$ are positive variables and $\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n$ 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 $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^{\geq 0}$ 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 $\|\cdot\|_p : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ by:

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

For $p = 1$ we call the norm the Manhattan norm, for $p = 2$ we call the norm the Euclidean norm, which is defined by the inner product $\langle \cdot | \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $\langle x | y \rangle := \sum_i x_i y_i$ 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, $\text{dist} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ with

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

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

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

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

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

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

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

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 $\forall x \in \mathcal{O}, \exists R = R(x) > 0. B(x, R) \subset \mathcal{O}$.

Definition 1.8. A sequence $(x_k)_{k \in \mathbb{N}}$ in \mathbb{R}^n converges to $x \in \mathbb{R}^n$ iff $\|x_k - x\| \xrightarrow{k \rightarrow \infty} 0$, 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 $\lim_{k \rightarrow \infty} x_k$.

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

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

Proof. “ \Rightarrow ”: We easily observe

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

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

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

□

Example 1.10. $x_k := (\frac{1}{k}, \frac{1}{k})_{k \in \mathbb{N}} \Rightarrow \lim_{k \rightarrow \infty} x_k = (0, 0) = 0$ 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 $\forall \epsilon' > 0 : (B(a, \epsilon') \setminus \{a\}) \cap \mathcal{U} \neq \emptyset$.

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

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

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

Example 1.14.

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

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

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

$$\text{-- Approach } (0, 0) \text{ along the main diagonal: } \longrightarrow f(x_k, y_k) = f(\frac{1}{k}, \frac{1}{k}) = \frac{\frac{1}{k^2}}{\frac{1}{k^2} + \frac{1}{k^2}} = \frac{1}{2} \xrightarrow{k \rightarrow \infty} \frac{1}{2}.$$

Hence, f is not continuous at 0.

$$\bullet \ g : \mathbb{R}^n \rightarrow \mathbb{R}, g(x, y) := \begin{cases} \frac{xy^3}{x^4+y^4} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

– Take $y_k := cx_k$ and yield.

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

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

As it turns out this does not converge either.

Lecture 2

Definition 2.15. Let $f : \mathcal{U} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ and assume let $a \in \mathbb{R}^n$ be an accumulation point. Now if $a \in \mathcal{U}$ and if $\exists \lim_{n \rightarrow \infty} f(x) = f(a)$, 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. $f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(x, y) := \begin{cases} \frac{xy}{x^2+y^2} & (x, y) \neq 0 \\ 0 & (x, y) = 0 \end{cases}$. We

have seen that $\nexists \lim_{(x,y) \rightarrow (0,0)} f(x, y)$, hence f is not continuous in 0. How about the other points in the domain?

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

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

$$\begin{matrix} x_n & \xrightarrow{n \rightarrow \infty} & x_0 \neq 0 \\ y_n & \xrightarrow{n \rightarrow \infty} & y_0 \neq 0 \end{matrix} \Rightarrow \frac{x_n y_n}{x_n^2 + y_n^2} \xrightarrow{n \rightarrow \infty} \frac{x_0 y_0}{x_0^2 + y_0^2},$$

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

$$2. \ f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(x, y) := \begin{cases} \frac{x^4 + 2x^2 y^2}{x^2 + y^2} & (x, y) \neq 0 \\ 0 & (x, y) = 0 \end{cases}. \text{ As before, we can}$$

conclude that g is continuous on $\mathbb{R}^2 \setminus \{0\}$. Furthermore, we have

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

Thus g is continuous on its domain.

Theorem 2.17. Let $f, g : \mathcal{U} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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)

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

(ii)

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

(iii)

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

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

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

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

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

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

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

(iii) follows from (i), (iii) and (iv). \square

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 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 $\forall i \in \{1, 2, \dots, n\}$. Then x_{k_i} converges to some $x_i \in \mathbb{R} \forall i \in \{1, 2, \dots, n\}$. Now put $x := (x_1, x_2, \dots, x_n)$ and apply lemma 1.9. It follows $x_k \rightarrow x$. Therefore, $(\mathbb{R}^n, \|\cdot\|)$ is Banach. \square

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 $A : \mathcal{X} \rightarrow \mathcal{Y}$ is called linear iff $A(\alpha x + \beta y) = \alpha A(x) + \beta A(y) \quad \forall \alpha, \beta \in \mathbb{R}, \forall x, y \in \mathcal{X}$. The set of linear maps $\mathcal{X} \rightarrow \mathcal{Y}$ is denoted by $L(\mathcal{X}, \mathcal{Y})$. $L(\mathcal{X}, \mathcal{X})$ is denoted $L(\mathcal{X})$. Often we write Ax instead of $A(x)$.

Facts:

1. $A0 = 0 \quad \forall A \in L(\mathcal{X}, \mathcal{Y})$.
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). \square

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.

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

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 $C := A \cdot B := (c_{k,j})_{\substack{1 \leq j \leq n \\ 1 \leq k \leq p}}$, 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 $\exists A^{-1} \in \mathcal{M}^{n \times m}(\mathbb{R})$.

$$AA^{-1} = A^{-1}A = I_n := \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots \\ & & \vdots & & \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \quad (\Rightarrow m = n), \text{ where } I_n \text{ is the identity}$$

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

$$L(\mathcal{X}, \mathcal{Y}) : A_{i,j} \rightarrow \left((x_k)_{k \in \{1,2,\dots,n\}} \rightarrow \left(\sum_{k=1}^m a_{i,k} \right)_{i \in \{1,2,\dots,n\}} \right).$$

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. $A \in L(\mathcal{X})$ invertible $\Leftrightarrow [A]$ 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 $\|Ax\| \leq \alpha \|x\| \forall x \in \mathbb{R}^n$, then $\|A\| \leq \alpha$.

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

Theorem 3.26 (Properties of the operator norm). 1. $\|Ax\| \leq \|A\| \|x\|, \forall x \in \mathbb{R}^n$.

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

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 $\|B - A\| \|A^{-1}\| < 1$ it follows $B \in I(\mathbb{R}^n)$. In other words, $\forall A \in I(\mathbb{R})$, we have that $B(A, \frac{1}{\|A^{-1}\|}) \subset I(\mathbb{R}^n)$. 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 $\left\| \frac{x}{\|x\|} \right\| = \frac{1}{\|x\|} \|x\| = 1$. By definition $\|A\| \geq \left\| A \left(\frac{x}{\|x\|} \right) \right\| = \frac{1}{\|x\|} \|A(x)\| \Rightarrow \|A\| \|x\| \geq \|A(x)\|$. 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:

$$\|BAx\| \leq \|BA\| \|x\| \leq \|B\| \|A\| \|x\| \quad (1.1)$$

6. Using item 1 and linearity:

$$\begin{aligned} \frac{\|x\|}{\|A^{-1}\|} &= \frac{\|A^{-1}Ax\|}{\|A^{-1}\|} \stackrel{\text{item 1}}{\leq} \frac{\|A^{-1}\|}{\|A^{-1}\|} \|Ax\| = \|Ax\| \\ &= \|(A - B)x + Bx\| \stackrel{\text{item 4}}{\leq} \|A - B\| \|x\| + \|Bx\| \end{aligned}$$

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: $\|BA\| \leq \|B\| \|A\|$.

□

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. □

Lecture 4

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

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

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

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

or

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

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 $df(x_0)$ or $f'(x_0)$.

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

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

$$\begin{aligned} & \|(A_1 - A_2)h\| \\ &= \|A_1h - A_2h\| \\ &= \|f(x_0 + h) - f(x_0) - A_2h - (f(x_0 + h) - f(x_0) - A_1h)\| \\ &\leq \|f(x_0 + h) - f(x_0) - A_2h\| + \|f(x_0 + h) - f(x_0) - A_1h\| \end{aligned}$$

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

$$0 \leq \frac{\|(A_1 - A_2)h\|}{\|h\|} \leq \frac{\|f(x_0 + h) - f(x_0) - A_1h\|}{\|h\|} + \frac{\|f(x_0 + h) - f(x_0) - A_2h\|}{\|h\|}.$$

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

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 $\lim_{t \rightarrow 0} \frac{\|(A_1 - A_2)(th)\|}{\|th\|} = 0$. 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 $A_1 = A_2$. \square

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

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

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 $g : [0, 1] \rightarrow \mathbb{R}^2$ by $g(x) := \begin{pmatrix} c \cos(x) \\ c \sin(x) \end{pmatrix}$, and $\|g\|$ at $x = \frac{1}{2}$: $\lim_{x \rightarrow \frac{1}{2}} \|g(x)\| = 0$, but $\exists \lim_{x \rightarrow \frac{1}{2}} g(x)$.

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 $\mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto Ax + b, \forall b \in \mathbb{R}^m, A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Here the differential is A again.

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

Lecture 5

Proposition 5.34. If any function $f : \mathcal{U} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, 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 $\lim_{x \rightarrow x_0} \frac{\varphi(x - x_0)}{\|x - x_0\|} = 0$. Thus $\lim_{x \rightarrow x_0} \varphi(x - x_0) = 0$, so $\lim_{x \rightarrow x_0} f(x) = f(x_0) + Df(x_0)(x - x_0) = f(x_0)$. \square

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

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

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

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 $f : \mathcal{U} \rightarrow \mathbb{R}^m$, $\mathcal{U} \subset \mathbb{R}^n$ open be differentiable in $x_0 \in \mathcal{U}$ and $g : \mathcal{V} \rightarrow \mathbb{R}^p$, for $\mathcal{V} \subset \mathbb{R}^m$ open with $y_0 := f(x_0) \in \mathcal{V}$. Then $g \circ f : \mathcal{U} \rightarrow \mathbb{R}^p$ 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:

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \varphi(x - x_0), \quad \lim_{x \rightarrow x_0} \frac{\varphi(x - x_0)}{\|x - x_0\|} = 0 \quad (1.2)$$

$$f(y) = f(y_0) + Df(y_0)(y - y_0) + \varphi(y - y_0), \quad \lim_{y \rightarrow y_0} \frac{\varphi(y - y_0)}{\|y - y_0\|} = 0 \quad (1.3)$$

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

$$y - y_0 = f(x) - f(x_0) = Df(x_0)(x - x_0) + \varphi(x - x_0) \quad (1.4)$$

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

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

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

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

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

$$\begin{aligned} & \lim_{x \rightarrow x_0} \frac{\varphi(x - x_0)}{\|x - x_0\|} = 0, \text{ that} \\ & \lim_{x \rightarrow x_0} \frac{Dg(f(x_0))\varphi(x - x_0)}{\|x - x_0\|} = 0 \end{aligned}$$

Secondly, define φ_0 by $\varphi(y-y_0) =: \|y - y_0\| \Psi_0(y-y_0)$, so $\lim_{x \rightarrow x_0} \Psi_0(y-y_0) = 0$ 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 $\lim_{z \rightarrow 0} \Psi_0(z) = 0$ and f is continuous in x_0 , so $\lim_{x \rightarrow x_0} f(x) = f(x_0)$, we obtain $\lim_{x \rightarrow x_0} \Psi_0(f(x) - f(x_0)) = 0$. We can conclude

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

□

Proposition 5.37 (Lipschitz continuity and $Df = 0 \Rightarrow f$ constant). *Let $f : \mathcal{U} \rightarrow \mathbb{R}^m$, $\mathcal{U} \subset \mathbb{R}^n$ 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 the statement that if $f : I \rightarrow \mathbb{R}$ is differentiable with $f' \equiv 0$, then f must be constant.

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

Example 5.39. $f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(x, y) := \begin{cases} \frac{xy}{x^2 + y^2} & (x, y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$. 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.

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

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

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

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

Example 5.43. 1. $f : \mathbb{R}^2 \rightarrow \mathbb{R}, f((x, y)) := x^2 + y^2$, is continuously partially differentiable on \mathbb{R}^2 , since both $\mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \rightarrow 2x, (x, y) \rightarrow 2y$ 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 (Jacobian). $f : \mathcal{U} \rightarrow \mathbb{R}^m$, $\mathcal{U} \subset \mathbb{R}^n$ 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 $J_f(x_0) := \left(\frac{\partial f_i}{\partial x_j} \right)_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathcal{M}^{m \times n}(\mathbb{R})$ is called the Jacobian matrix of f in x_0 .

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

Example 5.45. $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2, f((x, y, z)) := \begin{pmatrix} x^2 + y^2 + z^2 \\ xyz \end{pmatrix}$. f is continuously partially differentiable everywhere and the Jacobian 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 $f : \mathcal{U} \rightarrow \mathbb{R}^m, \mathcal{U}$ 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 $\lim_{h \rightarrow 0} \frac{\varphi(h)}{h} = 0$. Define $[Df(x_0)] := A(x_0) = (a_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$. Note we

$$\text{can write } Df(x_0)(h) = Ah = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1,j} h_j \\ \vdots \\ \sum_{j=1}^n a_{m,j} h_j \end{pmatrix}.$$

Written componentwise, this means for $i \in \{1, 2, \dots, m\} : f_i(x_0 + h) = f_i(x_0) +$

$$\sum_{j=1}^n a_{i,j} h_j + \varphi_i(h), \text{ where } \varphi \left(\begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{pmatrix} \right) \text{ and } \lim_{h \rightarrow 0} \frac{\varphi_i(h)}{||h||} = 0. \text{ Now putting}$$

$h = te_j, t \in \mathbb{R} \Rightarrow x_0 + te_j \in \mathcal{U}, \forall j \in \{1, 2, \dots, n\}$. Thus, $f_j(x_0 + te_j) = f_j(x_0) + \sum_{l=1}^n a_{jl} th_l + \varphi_j(te_j) = f_j(x_0) + ta_{ij} + \varphi_j(te_j)$. 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 $\frac{\partial f_i}{\partial x_j}(x_0) = \lim_{t \rightarrow 0} \frac{f_i(x_0 + te_j) - f_i(x_0)}{t} = \lim_{t \rightarrow 0} \frac{ta_{ij} + \varphi_i(te_j)}{t} = a_{ij} + \lim_{t \rightarrow 0} \frac{\varphi_i(te_j)}{t} = 0$, since $\lim_{h \rightarrow 0} \frac{\varphi_i(h)}{\|h\|} = 0$.

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. $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2, f((x, y, z)) = \begin{pmatrix} x^2 + y^2 + z^2 \\ xyz \end{pmatrix}, J_f((x_0, y_0, z_0)) = \begin{pmatrix} 2x_0 & 2y_0 & 2z_0 \\ y_0 z_0 & x_0 z_0 & x_0 y_0 \end{pmatrix} \in \mathcal{M}^{2 \times 3}(\mathbb{R})$. Therefore, $Df((x_0, y_0, z_0)) \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ is defined as $Df((x_0, y_0, z_0))(x, y, z) = \begin{pmatrix} 2(x_0 x + y_0 y + z_0 z) \\ x y_0 z_0 + x_0 y z_0 + x_0 y_0 z \end{pmatrix}$.

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

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

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

Lecture 7

Theorem 7.49. Let $f : \mathcal{U} \rightarrow \mathbb{R}^m, \mathcal{U} \subset \mathbb{R}^n$ open be any function and let its partial derivatives $\frac{\partial f}{\partial x_i} : \mathcal{U} \rightarrow \mathbb{R}^m$ $i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\}$ exist and are continuous (f is continuously differentiable). Then f is totally differentiable 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 $\lim_{h \rightarrow 0} \frac{\varphi(h)}{\|h\|} = 0, \forall h \in \mathbb{R}^n$, put $h^0 := 0 \in \mathbb{R}^n, \dots, h^j := \sum_{l=1}^j h_l e_l = (h_1 e_1, \dots, h_j e_j, 0, \dots, 0), \dots, h^n = h$, where $h = \sum_{l=1}^n h_l e_l$ with $h_j \in \mathbb{R}$ and $\{e_1, e_2, \dots, e_n\}$ the standard basis in \mathbb{R}^n . We can write

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

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

Applying the mean value theorem to f restricted to the segments $[x_0 + h^{j-1}, x_0 + h^j]$ 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 $[x, y] := \{(1 - t)x + ty : t \in [0, 1]\}$ $\forall x, y \in \mathbb{R}^n$.

Substituting these equalities into equation 1.5 leads to:

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

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

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

By construction, $(x_0 + h^{j-1} + \xi^j h_j e_j) \xrightarrow{h \rightarrow 0} x_0$. Therefore, by continuity of $\frac{\partial f}{\partial x_j} \Rightarrow (x_0 + h^{j-1} + \xi^j h_j e_j) \rightarrow \frac{\partial f}{\partial x_j} (x_0)$ and hence $\lim_{h \rightarrow 0} \varphi(h) = 0$. 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 $\Psi : h \rightarrow \sum_{j=1}^n \Psi_j(h) e_j$, 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,

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

□

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

2.

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

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

Definition 7.52. Let $f : \mathcal{U} \rightarrow \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ open, partially differentiable in $x_0 \in \mathcal{U}$. Then, the Jacobian of f in x_0 is also called the gradient of f in x_0 and 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 $f : \mathcal{U} \rightarrow \mathbb{R}, \mathcal{U} \subset \mathbb{R}^n$ 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 $\varphi : (-\epsilon, \epsilon) \rightarrow \mathcal{U}; \varphi(t) := x + tv$ and consider $F := f \circ \varphi : (-\epsilon, \epsilon) \rightarrow \mathbb{R}$. Applying the chain rule in this situation yields $J_f(t) = J_F(\varphi(t)) \cdot J_g(t) = \nabla f(\varphi(t)) \cdot$

$$\begin{pmatrix} \varphi'_1(t) \\ \vdots \\ \varphi'_n(t) \end{pmatrix} \stackrel{\text{definition}}{=} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \langle \nabla f(\varphi(t)) | v \rangle. \text{ In particular for } t = 0, J_F(0) = \langle \nabla f(x) | v \rangle. \text{ By definition of the directional derivative, we have that } D_v f(x) = \lim_{t \rightarrow 0} \frac{f(x+tv) - f(x)}{t} = \lim_{t \rightarrow 0} \frac{F(t) - F(0)}{t} = F'(0). \quad \square$$

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 $v \perp \nabla f(x) \Leftrightarrow \theta = \frac{\pi}{2} \Leftrightarrow \cos(\theta) = 0 = \langle \nabla f(x) | v \rangle$. \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: $f : \mathbb{R}^2 \rightarrow \mathbb{R}, f((x, y)) = x^2 + y^2$. 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. $f : \mathbb{R}^2 \rightarrow \mathbb{R}, f((x, y)) := x^2 + e^{xy} \sin(y)$. Some question; Find the direction $\frac{v}{\|v\|}$ and magnitude $\|v\|$ of maximal increase v of f in $(0, 1)$.

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

Lecture 8

The inverse function theorem

Motivation In analysis of functions of one variable we had that

Theorem 8.56. *Given*

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

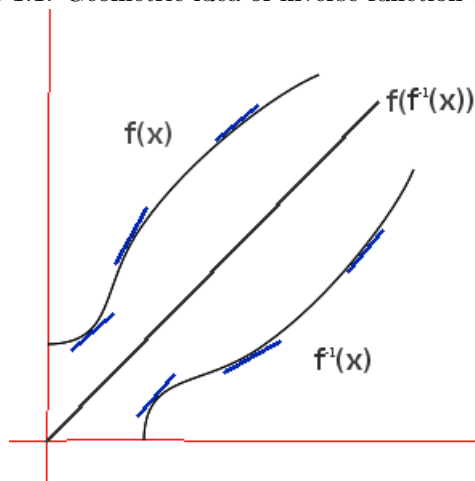
Then f is invertible (bijective), its inverse f^{-1} is also continuously differentiable and

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

Remark. Can be derived as

$$f^{-1}(f(x)) = x \xrightarrow{\text{differentiate}} (f^{-1})'(f(x)) \cdot f'(x) = 1 \iff (f^{-1})'(f(x)) = \frac{1}{f'(x)}$$

Figure 1.1: Geometric idea of inverse function theorem



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

Local definition 8.56.1. We call $f|_u$ a restriction of f at u ; we only consider f on the sub-domain u and its image from u .

Theorem 8.57 (Inverse function theorem for functions of several variables).
Given

1. $f : u \rightarrow \mathbb{R}^n$. $u \subset \mathbb{R}^n$ (u is an open n -ball)

2. f is differentiable on u , or, equivalently, $f|_u$ is differentiable
3. $Df : u \rightarrow L(\mathbb{R}^n)$ is continuous.
4. $Df(x_0)$ is invertible $x_0 \in u$, or, equivalently, that the jacobian $J_f(x_0)$ is invertible in $M^{n \times n}$

Then \exists neighborhood $u_0 \subset u$ 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:

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

II v_0 is open

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

Definition 8.58. A function defined from metric space (\mathcal{X}, d) to itself $\kappa : \mathcal{X} \rightarrow \mathcal{X}$ is called a contraction if

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

Remark. Intuitive picture of a contraction κ involves a function where the preimage always changes faster than the image; graph would be 'flatter' the smaller c is.

Local definition 8.58.1. A complete metric space is a metric space where all cauchy sequences converge. i.e.

$$((x_p) \in \mathcal{X} \implies \forall \epsilon > 0. \exists N_\epsilon \in \mathbb{N} : d(x_p, x_q) < \epsilon. \forall p, q \geq N_\epsilon) \implies x_p \xrightarrow{p \rightarrow \infty} x \in \mathcal{X}$$

Theorem 8.59 (Banach fixed point theorem). Let (\mathcal{X}, d) be a complete metric space. If $\kappa : \mathcal{X} \rightarrow \mathcal{X}$ is a contraction mapping, then κ has a unique fixed point $x \in \mathcal{X}$, i.e. a point $x \in \mathcal{X}$ where $\kappa(x) = x$.

Remark. Intuitive picture of Banach fixed point theorem: because κ is a contraction (the preimage x is always changing faster than the image $\kappa(x)$), you know that somewhere as you're moving towards either its maximal point (e.g: $+\infty$) or minimal point (e.g: $-\infty$) the value of x will 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

$$d(x_{p+1}, x_p) \leq c^{p-1}d(x_2, x_1). p \in \mathbb{N}. c \in [0, 1) \quad *$$

where c is the contraction constant for κ .

4. Proving above inequality by induction.

Base case ($p = k = 1$):

$$d(x_{1+1}, x_1) \leq c^{1-1} d(x_2, x_1)$$

$$d(x_2, x_1) \leq (c^0) d(x_2, x_1) \implies d(x_2, x_1) = d(x_2, x_1)$$

Inductive step ($p = k + 1$)

$$d(x_{k+2}, x_{k+1}) = d(\kappa(x_{k+1}), \kappa(x_k))$$

Since κ is a contraction

$$= d(\kappa(x_{k+1}), \kappa(x_k)) \leq c d(x_{k+1}, x_k)$$

Activate induction

$$\leq c c^{k-1} d(x_2, x_1) = c^k d(x_2, x_1)$$

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 (*)

$$\leq c^{m-2} d(x_2, x_1) + c^{m-3} d(x_2, x_1) + \cdots + c^{n-1} d(x_2, x_1)$$

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

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

So finally

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

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

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 \implies Lipschitz (with constant < 1) \iff Lipschitz continuous. Thus we can have

$$\kappa\left(\lim_{p \rightarrow \infty} x_p\right) = \lim_{p \rightarrow \infty} x_{p+1}$$

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

So x^* is a fixed point of κ

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

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

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. □

The inverse function theorem proof. :

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

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

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

(b) The differential map $Df : u \rightarrow L(\mathbb{R}^n)$ 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

$$\|Df(x) - Df(x_0)\| < \lambda \forall x \in u_0 \quad (1)$$

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

$$\kappa(x) = x + A^{-1}(y - f(x)) \quad (2)$$

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) is

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

(f) By proposition, since the differential is bounded \implies map is Lipschitz. We have that $\|\kappa(u) - \kappa(v)\| \leq \frac{1}{2}\|u - v\|$, $\forall u, v \in u_0 \implies \kappa$ 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 $\kappa = \kappa_y \implies f(x) = y$. In particular, $\forall y \in v_0 = f(u_0)$ there exists exactly one ¹ $x \in u_0$ with $f(x) = y \implies f|_{u_0}$ is injective $\implies f|_{u_0}$ 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

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

- (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

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

- (e) But this means that $\kappa(\overline{B}) \subset \overline{B} \implies \kappa$ is a contraction of a complete metric space \overline{B} with the euclidean metric.
 (f) By Banach fixed point theorem, $\exists! x \in \overline{B} : \kappa(x) = x \iff f(x) = y \iff y \in v_0 \implies v_0$ is open.

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

□

¹Banach 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 .