

MA259 Multivariable calculus :: Lecture notes

Lecturer: Mario Micalef

October 16, 2023

Contents

1	Convergence of a sequence	1
2	Continuity	2
2.1	Separate continuity	2
2.2	Properties of continuous functions	3
2.3	(USEFUL) Construction of continuous functions of several variable from real-valued functions of a single real variable (1st year analysis)	3
2.4	Caution with taking limits	4
3	Rudiments of topology and continuity	5
3.1	Continuity and topology	7
3.2	Bolzano-Weierstrass theorem	8
3.3	Continuity and sequential compactness	8
4	Linear maps and matrices	9
4.1	Norms on linear maps/matrices	9
4.1.1	Compare $\ A\ $ with $\ A\ _F$	11
4.1.2	Properties of norm	11
4.2	Convergence and continuity in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) / \mathbb{R}^{k \times n}$	11
4.2.1	Continuous matrix valued functions	11
4.3	General linear group $GL(n, \mathbb{R}) := \{A \in \mathbb{R}^n : A \text{ is invertible}\}$	12
5	Derivative of $f : U \rightarrow \mathbb{R}^k$ where $U \subset \mathbb{R}^n$ is open	13
5.1	Directional derivative	13
5.2	The (Fréchet) derivative as affine linear approximation	13
5.3	Partial derivative, Jacobian matrix, gradient	15
5.4	Summary of different types of derivative [just a recap]	16
5.4.1	Directional derivative	16
5.4.2	Partial derivative	16
5.5	Gradient of $f : U \rightarrow \mathbb{R}$, $U \subset \mathbb{R}^n$	18

5.6	Differentiation of matrix-valued functions	18
5.7	Chain rule	19
5.7.1	Common special cases	20
5.7.2	Idea of proof of chain rule (seen in first year analysis)	20
5.8	Continuity of partial derivatives implies differentiability	21
5.9	Mean value inequality (MVI)	23
6	Vector fields, curves, line integrals	24
6.1	Components of a vector	25
6.2	Curves and their parameterisation	25
6.3	Line integral	26
6.4	Flux and divergence in \mathbb{R}^2	29
6.5	Flux and divergence in \mathbb{R}^3	31
6.6	Gradient fields	32
6.7	Conservative fields	33
6.8	Scalar potential of conservative fields	34
6.8.1	Relation of curl and existence of scalar potential for C^1 vector fields . . .	35
6.9	Incompressible and conservative vector fields	36
7	Inverse function theorem	37
7.1	1-d review	37
8	Implicit function theorem	40
8.1	Maximal rank reformulation of Implicit function theorem	42
8.2	Using Implicit function theorem to parameterise (pieces of) level sets	42

Topics

- Continuity and differentiability of $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$
- Higher dimensional version of FTC

Difference between column and row vectors.

- Column vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ represents linear map $T : \mathbb{R} \rightarrow \mathbb{R}^n$, $T(t) = tv$ (line through 0 in direction v)
- Row vector $v = (v_1, \dots, v_n)$ represents $\lambda_v : \mathbb{R}^n \rightarrow \mathbb{R}$, $\lambda_v(x) = v_1x_1 + \dots + v_nx_n$ (linear functional)
 - $\text{Ker } \lambda_v = \{x \in \mathbb{R}^n, \lambda_v(x) = 0, \text{ i.e. } v^T \cdot x = 0\} = H_v$, hyperplane $\perp v^T$ through 0
 - Level set $\Gamma_c = \{x \in \mathbb{R}^n : \lambda_v(x) = c, \text{ i.e. } v^T \cdot x = c\}$ ($c \in \mathbb{R}$)

1 Convergence of a sequence

Recall the definition of convergence from Analysis I.

Definition 1.0.1. x_j is a sequence of vectors $\in \mathbb{R}^n$. We say $x_j \rightarrow x$ if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $j \geq N \Rightarrow$

$$\underbrace{|x_j - x|}_{\text{Euclidean distance} = \sqrt{(x_1 - x_{1j})^2 + \dots + (x_n - x_{nj})^2}} < \varepsilon.$$

Proposition 1.0.2. Limit is unique.

Proposition 1.0.3 (componentwise convergence). A sequence $x_j = (x_{1j} \dots x_{nj})$ converges to $x = (x_1 \dots x_n)$ if and only if $\forall i$, $x_{ij} \rightarrow x_i$.

Proof. “ \Rightarrow ” Suppose $x_j \rightarrow x$. Observe, $\forall i$, $|x_{ij} - x_i| \leq \sqrt{\sum_i (x_{ij} - x_i)^2} = |x_j - x| \rightarrow 0$. By sandwich $|x_{ij} - x_i| \rightarrow 0$ so $x_{ij} \rightarrow x_i$.

“ \Leftarrow ” Suppose $\forall i$, $x_{ij} \rightarrow x_i$. Then given $\varepsilon > 0$, $\exists N_i \in \mathbb{N}$ such that $j \geq N_i \Rightarrow |x_{ij} - x_i| < \varepsilon$. Set $N = \max\{N_i\}$. Then

$$|x_j - x| \leq \sqrt{n\varepsilon^2} = \sqrt{n}\varepsilon \Rightarrow x_j \rightarrow x.$$

□

Remark. That “ $\max\{\}$ ” can be seen as a norm:

- $|x|_\infty := \max\{|x_1|, |x_2|, \dots, |x_n|\}$
- NY taxi cab norm: $|x|_1 := |x_1| + \dots + |x_n|$
- more generally, $|x|_p := (|x_1|^p + \dots + |x_n|^p)^{1/p}$

Remark. The theorem above in Euclidean norm is equivalent to that in $|\cdot|_\infty$.

Lemma 1.0.4. If $x_j \rightarrow x$ then $|x_j| \rightarrow |x|$.

Proof. Given $\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $j \geq N \Rightarrow |x_j - x| < \varepsilon \Rightarrow ||x_j| - |x|| \leq |x_j - x| < \varepsilon$. □

Remark. $|x_j|$ converges $\nRightarrow x_j$ converges.

2 Continuity

Definition 2.0.1 ($\varepsilon - \delta$). $f : U \rightarrow \mathbb{R}^k$ where $U \subset \mathbb{R}^n$ is continuous at $p \in U$ if $\forall \varepsilon > 0, \exists \delta > 0$ such that $|x - p| < \delta \Rightarrow |f(x) - f(p)| < \varepsilon$ and $x \in U$.

Definition 2.0.2 (sequential). $f : U \rightarrow \mathbb{R}^k$ is continuous at p if $\forall x_j \rightarrow p, f(x_j) \rightarrow f(p)$.

Notation. $\mathcal{C}(U, \mathbb{R}^k) = \{f : U \rightarrow \mathbb{R}^k, f \text{ is continuous at all points of } U\}$

Definition 2.0.3. $f : U \rightarrow \mathbb{R}^k$ has a limit at $p \in U$ if $\exists q \in \mathbb{R}^k$ such that $\forall \varepsilon > 0, \exists \delta > 0 : 0 < |x - p| < \delta \Rightarrow |f(x) - q| < \varepsilon$. Then we write

$$\lim_{x \rightarrow p} f(x) = q.$$

2.1 Separate continuity

How do transport what we know about 1-d functions in Analysis to multivariable ones? Vertical slicing of $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. (seen in MA134 when introducing partial derivative)

Given $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ then for each fixed $y \in \mathbb{R}$ we consider $g^y : \mathbb{R} \rightarrow \mathbb{R}$ defined by $g^y(x) := f(x, y)$, which is a familiar function of one variable.

Example 2.1.1. $f(x, y) = x \sin y$. Then $g^1(x) = (\sin 1)x$. We then see f as a family of one-real-variable linear functions, with the parameter y varying them around the origin.

Similarly for each $x \in \mathbb{R}$ we consider $h^x : \mathbb{R} \rightarrow \mathbb{R}$ defined by $h^x(y) := f(x, y)$.

Example 2.1.2. Using the same f above, this is now a family of sine curves with x as a parameter (amplitude), e.g. $h^3(y) = 3 \sin y$.

The graphs (curves) of g^s and h^t are obtained by intersecting the graph of f (a surface) with the vertical plane $y = s$ and $x = t$ respectively. (vertical slicing of the graph of f .)

Definition 2.1.3 (separate continuity). $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is separately continuous at (x_0, y_0) if g^{y_0} is continuous at x_0 and h^{x_0} is continuous at y_0 .

Proposition 2.1.4. Continuity implies separate continuity.

Converse is clearly false. A simple counterexample:

$$f(x, y) = \begin{cases} 0 & \text{if } xy = 0 \\ 1 & \text{if } xy \neq 0 \end{cases},$$

which is separately continuous at $(0, 0)$ since $\forall x, y \ g^0(y) = h^0(x) = 0$, but clearly not continuous at $(0, 0)$.

Remark. This reminds us that partial derivatives only give information on the axis, so in order for the function to be differentiable, it is not enough to consider existence of partial derivatives. (Continuity is needed.)

2.2 Properties of continuous functions

- Continuity of sum of continuous functions
- Continuity of product/quotient [nonzero denominator] of continuous real-valued functions (we cannot multiply vectors yet)
- Continuity of composition of continuous functions

Proposition 2.2.1 (componentwise continuity). $f : U \rightarrow \mathbb{R}^k$, $U \subset \mathbb{R}^n$ is continuous at $p \in U$ iff $f_1(x), \dots, f_k(x)$ where $f_i(x) : U \rightarrow \mathbb{R}$ are continuous at p .

Outline of proof. Either use componentwise convergence of sequence of vectors and sequential definition of continuity OR a ε - δ proof, for which

$$|f_i(x) - f_i(p)| \leq \overbrace{|f(x) - f(p)|}^{\text{distance in Euclidean norm}} \leq \sqrt{k} \max_{1 \leq i \leq k} |f_i(x) - f_i(p)|$$

is needed. □

2.3 (USEFUL) Construction of continuous functions of several variable from real-valued functions of a single real variable (1st year analysis)

Consider $g : E \rightarrow \mathbb{R}$ where $E \subset \mathbb{R}$. For $1 \leq i \leq n$, let

$$U_i = \{(x_1, \dots, x_n) \in \mathbb{R}^n\}, \quad x_i \in E.$$

In \mathbb{R}^2 it would be vertical or horizontal strips. Define $f_i : U_i \rightarrow \mathbb{R}$ by

$$f_i(x_1, \dots, x_n) := g(x_i).$$

Proposition 2.3.1. If g is continuous at $a \in E$ then f_i is continuous on $\{(x_1, \dots, x_n) : x_i = a\}$.

Proof. Continuity of g at a : given $\varepsilon > 0$, $\exists \delta > 0$ such that

$$|a - y_i| < \delta \Rightarrow |g(a) - g(y_i)| < \varepsilon.$$

So if $x, y \in U_u$, $x_i = a$ then

$$\begin{aligned} |x - y| < \delta &\Rightarrow \sqrt{\dots + (a - y_i)^2 + \dots} < \delta \Rightarrow |a - y_i| < \delta \\ &\Rightarrow |g(a) - g(y_i)| < \varepsilon \Rightarrow |f_i(x) - f_i(y)| < \varepsilon. \end{aligned}$$

□

Example 2.3.2. $f(x, y) = \frac{xy}{x^2 + y^2}$, $(x, y) \in \underbrace{\mathbb{R}^2 \setminus \{(0, 0)\}}_{\text{natural domain}}$. Then we claim f is continuous on

this domain. We consider f constructed from single variable functions. Consider $g_1(x) = x$ which is continuous on \mathbb{R} . Then $f_1(x, y) = g_1(x) = x$ and $\tilde{f}_1(x, y) = y = g_1(y)$ are continuous

on \mathbb{R}^2 , U_1 and U_2 respectively. Also consider $g_2(x) = x^2$ which is continuous on \mathbb{R} . Then $f_2(x, y) = g_2(x) = x^2$ and $\tilde{f}_2(x, y) = g_2(y) = y^2$ are continuous on U_1, U_2 . So

$$f(x, y) = \frac{[f_1(x, y)][\tilde{f}_1(x, y)]}{[f_2(x, y) + [\tilde{f}_2(x, y)]],}$$

so by algebraic properties of continuous functions, f is continuous on $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Example 2.3.3.

$$f(x, y, z) := \left(\frac{\log(x+y)}{\sin z}, \arccos y \sqrt{1 + (\cos(xe^2))^2} \right),$$

is continuous on the natural domain $\{(x, y, z) : x + y > 0, z \neq n\pi, n \in \mathbb{Z}, |y| \leq 1\}$.

Example 2.3.4. $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is continuous on \mathbb{R}^{n^2} since it's a polynomial in its variables.

2.4 Caution with taking limits

Example 2.4.1. $f(x, y) = \frac{xy}{x^2 + y^2}$ with natural domain. If we define $f(0, 0) = 0$ then f is separately continuous at $(0, 0)$. Why?

$$g^0(x) := f(x, 0) = \begin{cases} 0 & \text{if } x \neq 0 \text{ by formula} \\ 0 & \text{if } x = 0 \text{ by definition} \end{cases},$$

so $g^0(x) = 0 \forall x$. In particular, g^0 is continuous at 0. Similarly, $h^0(y) = 0 \forall y$, so continuous, so separately continuous. BUT

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y)$$

does not exist.

Consider limit of f along lines through $(0, 0)$ L_θ , parameterised by $r(t) = (t \cos \theta, t \sin \theta)$, $t \in \mathbb{R}$. Then

$$f|_{L_\theta} = f(r(t)) = \frac{(t \cos \theta)(t \sin \theta)}{t^2 \cos^2 \theta + t^2 \sin^2 \theta} = \frac{1}{2} \sin 2\theta,$$

a constant which is independent of t .

Since

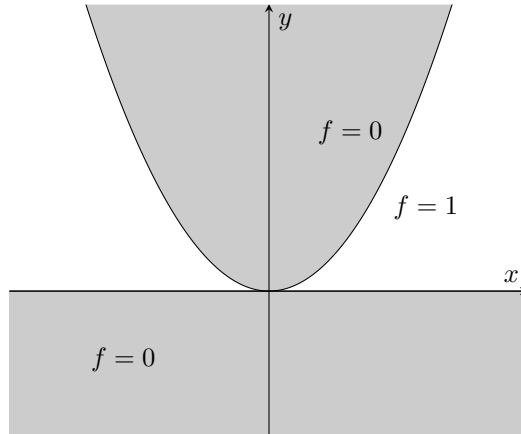
$$\lim_{(x,y) \rightarrow (0,0), (x,y) \in L_\theta} f(x, y)$$

depends on θ , we conclude that $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ does not exist.

Question: if $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is continuous along all lines through $(x_0, y_0) \in \mathbb{R}^2$, is f continuous at (x_0, y_0) ?

Answer: No.

Example 2.4.2 (Counterexample). Consider f visualised by following diagram.



Then we claim f is continuous along all lines (line segments) through $(0,0)$ because for any direction $v = (a, b)$, $\exists \tau > 0$ such that the line segment

$$L_{v,\tau} = \{(ta, tb) : |t| < \tau\}$$

is contained in

$$E = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\} = \{(x, y) \in \mathbb{R}^2 : y \leq 0 \text{ or } y \geq x^2\} = f^{-1}(0).$$

But f is clearly not continuous at $(0,0)$ because

$$\lim_{x \rightarrow 0} f\left(x, \frac{1}{2}x^2\right) = 1 \neq 0 = f(0,0).$$

Definition 2.4.3 (Continuity along lines, or linear continuity in some literature, but it has nothing to do with linearity of function). $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is continuous along lines at $x_0 \in \mathbb{R}^n$ if $\forall v \in \mathbb{R}^n$, $t : \mathbb{R} \rightarrow \mathbb{R}^k \mapsto f(x_0 + tv)$ is continuous at $t = 0$, i.e.

$$\forall v, f(x_0) = \lim_{t \rightarrow 0} f(x_0 + tv).$$

Example 2.4.4. $f(x, y) = \frac{x^2 y}{x^4 + y^2}$ and $f(0,0) = 0$ is continuous on $\mathbb{R}^2 \setminus \{(0,0)\}$ and along lines through $(0,0)$, but $\lim_{(x,y) \rightarrow (0,0)}$ does not exist, because

$$\lim_{x \rightarrow 0} f(x, x^2) = \frac{x^2 x^2}{x^4 + (x^2)^2} = \frac{1}{2} \neq 0 = f(0,0).$$

The whole upshot of this is that continuity is much stronger than continuity along lines, an analogy with differentiability being much stronger than existence of directional derivative, which we are going to see later.

3 Rudiments of topology and continuity

Definition 3.0.1 (Closed set). $X \subset \mathbb{R}^n$ is *closed* if x_j is a sequence converging to $x \in \mathbb{R}^n \Rightarrow x \in X$, i.e. X is closed under taking limits. e.g. \mathbb{R} is closed but \mathbb{Q} is not.

Definition 3.0.2 (Open set). $U \subset \mathbb{R}^n$ is *open* if $\forall x \in U, \exists \varepsilon > 0$ such that $y \in \mathbb{R}^n$ and $|y - x| < \varepsilon \Rightarrow y \in U$, i.e. all points in U have a little room to move around (can be perturbed a little) and stay in U .

Remark (Convention). Empty set \emptyset and \mathbb{R}^n are both open and closed.

Proposition 3.0.3. $U \subset \mathbb{R}^n$ is open if and only if U^c is closed.

Proof. Let $U \subset \mathbb{R}^n$ be closed and U^c is nonempty. Suppose $\exists v \in U^c : \forall \varepsilon > 0, |u - v| < \varepsilon \forall u \in U$. We can then construct a sequence $u_j \subset U$ such that $\forall \varepsilon > 0, \exists N \in \mathbb{N} : j \geq N \Rightarrow |u_j - v| < \varepsilon$, i.e. $u_j \rightarrow v$. This contradicts the hypothesis that U is closed. Hence $\forall v \in U^c, \exists \varepsilon > 0 : |u - v| \geq \varepsilon \forall u \in U$, which means $\forall u \notin U, |u - v| < \varepsilon$, i.e. U^c is open.

Now suppose $U \subset \mathbb{R}^n$ is open and let $u_j \subset U^c$ be a sequence converging to $x \in \mathbb{R}^n$, i.e. $\forall \varepsilon > 0, \exists N \in \mathbb{N} : j \geq N \Rightarrow |u_j - x| < \varepsilon$. Suppose $x \in U$. But $\forall x \in U, \exists \varepsilon > 0 : |u - x| < \varepsilon \Rightarrow u \in U$, contradicting the hypothesis $u_j \subset U^c$. Hence $x \notin U$, i.e. U^c is closed. \square

Definition 3.0.4 (Open (Euclidean) ball). Open ball of radius $r > 0$, centred at $a \in \mathbb{R}^n$, denoted by $\mathbb{B}(a, r)$ or $\mathbb{B}_r(a)$, is $\{x \in \mathbb{R}^n : |x - a| < r\}$. Also \mathbb{B}_r means $\mathbb{B}(0, r)$ and \mathbb{B} is just \mathbb{B}_1 , the unit ball.

We can therefore rewrite Definition 3.0.2 as: $U \subset \mathbb{R}^n$ is *open* if $\forall x \in U, \exists \varepsilon > 0 : \mathbb{B}(x, \varepsilon) \subset U$.

Proposition 3.0.5. An open ball is open.

Proof. Pick $y \in \mathbb{B}(a, r)$ and set $\delta = r - |a - y| > 0$. Then we claim $\mathbb{B}(y, \delta) \subset \mathbb{B}(a, r)$, since $x \in \mathbb{B}(y, \delta)$ and by triangle inequality, $|x - a| \leq |x - y| + |y - a| < \delta + |y - a| = r$. \square

Definition 3.0.6 (Closed ball). $\overline{\mathbb{B}(a, r)} = \{x \in \mathbb{R}^n : |x - a| \leq r\}$.

Proposition 3.0.7. A closed ball is closed.

Proof. Let $x \in \overline{\mathbb{B}(a, r)}^c = \{x \in \mathbb{R}^n : |x - a| > r\}$ and let $r_0 = |x - a| - r > 0$. We claim $\mathbb{B}(x, r_0) \subset \overline{\mathbb{B}(a, r)}^c$, since for any $y \in \mathbb{B}(x, r_0)$, $|x - a| \leq |x - y| + |y - a| \Rightarrow |y - a| \geq |x - a| - |x - y| > r$. Therefore $\overline{\mathbb{B}(a, r)}^c$ is open, hence $\overline{\mathbb{B}(a, r)}$ is closed. \square

Proposition 3.0.8. Let $\{U_\lambda\}$ be a collection of open sets where $\lambda \in \Lambda$, an indexing set e.g. $\mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{R}^n$, etc. Then $\bigcup_{\lambda \in \Lambda} U_\lambda$ is open.

One-liner proof! (that takes two lines to write). If $p \in \bigcup_{\lambda \in \Lambda} U_\lambda$, then $\exists \lambda^* : p \in U_{\lambda^*}$. But U_{λ^*} is open, therefore $\exists \varepsilon > 0 : \mathbb{B}(p, \varepsilon) \subset U_{\lambda^*} \subset \bigcup_{\lambda \in \Lambda} U_\lambda$. \square

See Definition 2.1.10 and Example 2.1.11 in the typewritten notes for some fun stuff that I believe is not examinable.

Are intersections of open sets open? Let $U_n = (-\frac{1}{n}, \frac{1}{n})$, then $\bigcap_{n \in \mathbb{N}} U_n = \{0\}$, which is not open.

But

Proposition 3.0.9. If the finitely many sets U_1, \dots, U_n are open then so is $\bigcap_{j=1}^n U_j$.

Proof. Pick $p \in \bigcap_{j=1}^n U_j$. Then $p \in U_i \Rightarrow \exists \varepsilon_i > 0 : \mathbb{B}(p, \varepsilon_i) \subset U_i$ for all $i \in \{1, \dots, n\}$. Select $\varepsilon^* := \min\{\varepsilon_1, \dots, \varepsilon_n\}$ which is positive. [This is why you need it to be finite!] Then $\mathbb{B}(p, \varepsilon^*) \subset U_j$ $\forall j \in \{1, \dots, n\}$, i.e. $\mathbb{B}(p, \varepsilon^*) \subset \bigcap_{j=1}^n U_j$. \square

Corollary 3.0.10. Arbitrary intersection of closed sets is closed; finite union of closed sets is closed.

Proof. If $\{U_\lambda\}$ are closed then $\{U_\lambda^c\}$ are open and $\bigcup_\lambda U_\lambda^c$ is open, so $\left(\bigcup_\lambda U_\lambda^c\right)^c = \bigcap_\lambda U_\lambda$ is closed.

If U_1, \dots, U_n are closed then U_1^c, \dots, U_n^c are open and $\bigcap_{j=1}^n U_j^c$ is open, so $\left(\bigcap_{j=1}^n U_j^c\right)^c = \bigcup_{j=1}^n U_j$ is closed. \square

A collection of open sets gives us what is known as a *topology*.

3.1 Continuity and topology

Recall definition of continuity. We can translate that into ball language: $f : U \rightarrow \mathbb{R}^k$ is continuous at $p \in U$ if

$$\forall \varepsilon > 0, \exists \delta > 0 : x \in \mathbb{B}(p, \delta) \cap U \Rightarrow f(x) \in \mathbb{B}(f(p), \varepsilon).$$

or equivalently

$$f^{-1}(\mathbb{B}(f(p), \varepsilon)) \supset \mathbb{B}(p, \delta) \cap U.$$

Consider $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ (so we don't have to fuss about $\cap U$). We have

Theorem 3.1.1 (Continuity via open and closed sets). The following statements are equivalent:

1. f is ε - δ continuous at all points of \mathbb{R}^n .
2. For all open sets $V \subset \mathbb{R}^k$, $f^{-1}(V)$ is open in \mathbb{R}^n .
3. For all closed sets $\mathcal{F} \subset \mathbb{R}^k$, $f^{-1}(\mathcal{F})$ is closed in \mathbb{R}^n .

Proof. $1 \Rightarrow 2$: Let $V \subset \mathbb{R}^k$ be open and let $p \in f^{-1}(V)$, i.e. $f(p) \in V$. Since V is open, $\exists \varepsilon > 0 : \mathbb{B}(f(p), \varepsilon) \subset V$. Continuity of f at p implies that $\exists \delta > 0 : \mathbb{B}(p, \delta) \subset f^{-1}(\underbrace{\mathbb{B}(f(p), \varepsilon)}_{\subset V}) \subset$

$f^{-1}(V)$, i.e. $f^{-1}(V)$ is open.

$2 \Rightarrow 1$: Let $p \in \mathbb{R}^n$ and $\varepsilon > 0$. Set $V := \mathbb{B}(f(p), \varepsilon)$ which is open, hence $f^{-1}(V)$ is open. But $p \in f^{-1}(V)$, therefore $\exists \delta > 0 : \mathbb{B}(p, \delta) \subset f^{-1}(V) = f^{-1}(\mathbb{B}(f(p), \varepsilon))$ which is exactly the ε - δ definition of continuity.

2 and 3 are obviously equivalent.

□

Application: to show openness of a set.

Example 3.1.2.

$$f(x, y) = xy \left(\sin \frac{1}{x} \right) \left(\cos \frac{1}{y} \right)$$

is clearly continuous, therefore $\{(x, y) \in \mathbb{R}^2 : f(x, y) > 0\}$ is open since it equals $f^{-1}((0, \infty))$ where $(0, \infty)$ is open.

Now a flashback:

3.2 Bolzano-Weierstrass theorem

We will not go through the formal proof of B-W, which is in the typewritten notes. But an example will be demonstrated to help us understand the proof.

Example 3.2.1 (Review of bounded sequence in \mathbb{R}).

$$a_j := \sin \left(\frac{\pi j}{2} \right)$$

so we have $a_1 = 1, a_2 = 0, a_3 = -1, a_4 = 0$ and so on. This is bounded so it has convergent subsequences, e.g. $a_{4k-3} \rightarrow 1, a_{2k} \rightarrow 0, a_{4k-1} \rightarrow -1$. Let $a_{2k} = a_{j_k}$, i.e. $j_k = 2k$ which is an increasing sequence $\in \mathbb{N}$; this leads us to

Definition 3.2.2 (Subsequence of a_j). Take j_k to be an increasing sequence of natural number and then a_{j_k} is a subsequence of a_j .

Example 3.2.3 (Bounded sequence in \mathbb{R}^2 (a_j, b_j)).

$$(a_j, b_j) = \left(\sin \frac{\pi j}{2}, \sin \frac{\pi j}{3} \right)$$

is bounded since $|(a_j, b_j)| \leq \sqrt{1^2 + 1^2} = \sqrt{2}$. We know a_{2k} converges but b_{2k} does not. But we know b_{6l} converges, so we set $k_l = 3l$ then $b_{j_{k_l}} = b_{j_{3l}} = b_{6l} \rightarrow 0$. Now a_{6l} also converges since it's a subsequence of the convergent subsequence $a_{j_k} = a_{2k}$. So

$$\lim_{l \rightarrow \infty} (a_{j_{k_l}}, b_{j_{k_l}}) = (0, 0).$$

The proof (or the actual computational process) goes a bit like this: you get convergent subsequences of all the components, and take a subsequence which is a subsequence of all the convergent subsequences.

3.3 Continuity and sequential compactness

Definition 3.3.1 (Sequential compactness subset). $K \subset \mathbb{R}^n$ is *sequentially compact* if every sequence $x_j \in K$ has a convergent subsequence x_{j_k} whose limit is in K .

Definition 3.3.2. $X \subset \mathbb{R}^n$ is bounded if $\exists M > 0 : |x| \leq M \forall x \in X$.

Theorem 3.3.3 (Heine–Borel). $K \subset \mathbb{R}^n$ is sequentially compact if and only if K is closed and bounded.

This is a generalisation of closed bounded.

Proof. See typewritten notes. □

Example 3.3.4 (Application).

$$S^{n-1}(a, r) = \{x \in \mathbb{R}^n : |x - a| = r, r > 0\},$$

the $n - 1$ dimensional sphere centred at a of radius r , is sequentially compact because it is closed: $f : \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = |x - a|$ is continuous by reverse triangle inequality, then $S^{n-1}(a, r) = f^{-1}(\{r\})$ where $\{r\}$ is closed; and bounded by $|a| + r$.

Theorem 3.3.5 (Continuity preserves sequential compactness). If $f : K \rightarrow \mathbb{R}^k$ is continuous and $K \subset \mathbb{R}^n$ is sequentially compact then $f(K)$ is also sequentially compact.

Proof. Straightforward. See typewritten notes. □

Theorem 3.3.6 (Extreme Value Theorem). Let $K \subset \mathbb{R}^n$ be sequentially compact and let $f : K \rightarrow \underbrace{\mathbb{R}}_{\text{not } \mathbb{R}^k}$ be continuous. Then $\exists x_*, x^* \in K : f(x_*) \leq f(x) \leq f(x^*) \forall x \in K$.

Proof. $f(K)$ is sequentially compact subset of \mathbb{R} . This implies $f(K)$ is closed and bounded by Theorem 3.3.3. Boundedness gives $M := \sup f(K)$ and $m := \inf f(K)$ are finite. By definition $\exists a_j, b_j \in f(K) : \lim_{j \rightarrow \infty} a_j = M, \lim_{j \rightarrow \infty} b_j = m$. Closeness gives $M, m \in f(K)$, i.e. $\exists x^*, x_* : f(x^*) = M \geq f(x) \geq m = f(x_*)$. □

4 Linear maps and matrices

Read notation on page 16 of typewritten notes.

4.1 Norms on linear maps/matrices

Reminder: $\mathbb{R}^{k \times n} \cong \mathbb{R}^{kn}$. Wait. So what do we gain when we write $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix}$ instead

of simply $(a_{11}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{k1}, \dots, a_{kn})$? Answer is certain operations are simply form in matrix form, e.g. multiplication, taking inverses.

We then define a norm on this space as the Euclidean norm of the latter vector:

Definition 4.1.1 (Frobenius norm).

$$\|(a_{ij})\|_F := \sqrt{\sum_{i=1}^k \sum_{j=1}^n a_{ij}^2}.$$

This can also be viewed as a norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ via the identification

$$A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) \rightarrow \mu(A) \in \mathbb{R}^{k \times n},$$

where $\mu(A)$ is the matrix representation of A with respect to the standard bases on $\mathbb{R}^k, \mathbb{R}^n$. Note

$$\mu : \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) \rightarrow \mathbb{R}^{k \times n}, \quad Av_j = \sum_{i=1}^k a_{ij} w_i$$

is a linear isomorphism.

Remark (Warning). Sometimes we write A when we mean $\mu(A)$. Context is everything.

Question is given $A \in \mathbb{R}^{k \times n}$, how large can $|Ax|$ get relative to $|x|$ as x ranges over \mathbb{R}^n ? Let $x = (x_1, \dots, x_n)$. Then

$$|Ax|^2 = \sum_{i=1}^k \left(\underbrace{\sum_{j=1}^n a_{ij} x_j}_{\text{dot product of the } i\text{th row with } x} \right)^2$$

and by Cauchy-Schwarz

$$|Ax|^2 \leq \sum_{i=1}^k \left(\sum_{j=1}^n a_{ij}^2 \right) \left(\sum_{j=1}^n x_j^2 \right) = \left(\sum_{i=1}^k \sum_{j=1}^n a_{ij}^2 \right) |x|^2 = \|A\|_F^2 |x|^2.$$

So if $x \neq 0$,

$$\frac{|Ax|}{|x|} \leq \|A\|_F,$$

which gives rise to

Definition 4.1.2 (Operator norm).

$$\|A\| := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|Ax|}{|x|}.$$

If $x \neq 0$,

$$\frac{|Ax|}{|x|} = \left| \frac{Ax}{|x|} \right| = \left| A \left(\frac{x}{|x|} \right) \right|,$$

but $\left| \frac{x}{|x|} \right| = 1$ since it's a unit vector. Therefore we can also write

$$\|A\| := \sup_{x \in S^{n-1}} |Ax|$$

where S^{n-1} = set of unit vectors $\{x \in \mathbb{R}^n : |x| = 1\}$ which is sequentially compact. So since norm is a continuous function, it attains maximum and minimum. [See example sheet 3 for how it's useful.]

4.1.1 Compare $\|A\|$ with $\|A\|_F$

We have

$$\|A\|_F^2 = \sum_{j=1}^n \sum_{i=1}^k a_{ij}^2$$

where a_{ij} is the j -fixed-components of Av_j , or i th column of A . Then

$$\sum_{i=1}^k a_{ij}^2 = |Av_j|^2 \leq \|A\| |v_j| = \|A\|.$$

So

$$\|A\|_F^2 \leq n \|A\|^2,$$

i.e.

$$\frac{1}{\sqrt{n}} \|A\|_F \leq \|A\| \leq \|A\|_F.$$

This equivalence of norms helps us deal with convergence of matrices.

4.1.2 Properties of norm

Read proposition 3.1.2 in typewritten notes.

Proposition 4.1.3 (Composition bound - useful). $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$, $B \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^m)$. Then $BA \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $\|BA\| \leq \|B\| \|A\|$.

One-liner proof! $|BAx| = |B(Ax)| \leq \|B\| |Ax| \leq \|B\| \|A\| |x|$. So

$$\|BA\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|BAx|}{|x|} \leq \|B\| \|A\|.$$

□

4.2 Convergence and continuity in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ / $\mathbb{R}^{k \times n}$

Definition 4.2.1. Sequence $A_j \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$, then

$$\lim_{j \rightarrow \infty} A_j = A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$$

if $\forall \varepsilon > 0, \exists N \in \mathbb{N} :$

$$j \geq N \Rightarrow \|A_j - A\| < \varepsilon \quad \text{or, of course, } \|A_j - A\|_F < \varepsilon.$$

Definition 4.2.2 (Ball in \mathcal{L} (??)). $\mathbb{B}(A, r) = \{B \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) : \|B - A\| < r\}$.

4.2.1 Continuous matrix valued functions

$f : \mathbb{R}^l \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ is continuous at $x \in \mathbb{R}^l$ if $\forall \varepsilon > 0, \exists \delta > 0 : |y - x| < \delta \Rightarrow \underbrace{\|f(y) - f(x)\|}_{\text{Operator norm. Could use Frobenius too}} < \varepsilon$.

ε .

Okay, but what if we replace $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ with $\mathbb{R}^{k \times n} \leftrightarrow \mathbb{R}^{nk}$? Then

Proposition 4.2.3.

$$f(x) = \begin{pmatrix} a_{11}(x) & \cdots & a_{1n}(x) \\ \vdots & \ddots & \vdots \\ a_{k1}(x) & \cdots & a_{kn}(x) \end{pmatrix}$$

is continuous if and only if each $a_{ij}(x)$ is continuous by equivalence of continuity and componentwise continuity.

Corollary 4.2.4. Determinant function $\Delta : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is continuous since each term of $\Delta(A)$ is a polynomial in the entries of A .

$$\Delta(A) = a_{11} \cdots a_{nn} + \cdots$$

in the n^2 variables, each of whose terms is linear in the entries a_{ij} of A . This is an example of multilinear map.

4.3 General linear group $GL(n, \mathbb{R}) := \{A \in \mathbb{R}^n : A \text{ is invertible}\}$

which $= \{A \in (\mathbb{R}^n) : \det A \neq 0\} = \Delta^{-1}(\mathbb{R} \setminus \{0\})$ where $\mathbb{R} \setminus \{0\} = (-\infty, 0) \cup (0, \infty)$ is open, therefore $GL(n, \mathbb{R})$ is an open subset of $\mathbb{R}^{n \times n}$. It essentially means if you jiggle an invertible matrix, it remains invertible. (Compare with singular matrices $\Delta^{-1}(0)$ which is unstable.)

Question: given $A \in GL(n, \mathbb{R})$, how large can we take $r > 0$ such that $\mathbb{B}(A, r) \subset GL(n, \mathbb{R})$?

Answer:

Proposition 4.3.1. Let $r := \frac{1}{\|A^{-1}\|}$. Then

$$\mathbb{B}(A, r) \subset GL(n, \mathbb{R}).$$

Furthermore, if $B \in \mathbb{B}(A, r)$, then

$$\|B^{-1}\| \leq \frac{1}{r - \|B - A\|}.$$

Proof. Let $B \in \mathbb{B}(A, r)$. We know $x = A^{-1}(Ax)$. Then $|x| \leq \|A^{-1}\| |Ax|$, so $|Ax| \geq r|x|$. [Side note: This is a measure of injectivity of A , i.e. since $Ax = 0 \Rightarrow x = 0$, it is a measure of how far A takes a vector away from 0.] We would like to show B is invertible by showing $|Bx| \geq \delta|x|$ for some $\delta > 0$. We write

$$|Bx| = |Bx - Ax + Ax| \geq |Ax| - |(B - A)x| \geq \underbrace{(r - \|B - A\|)}_{\delta} |x|. \quad (*)$$

So if $B \in \mathbb{B}(A, r)$ $\|B - A\| < r$, then $Bx = 0 \Rightarrow x = 0$, i.e. $\text{Ker} B = \{0\}$ i.e. image has same dimension n by rank-nullity theorem i.e. invertible.

To estimate $\|B^{-1}\|$ we use $x = B(B^{-1})x$. By $(*)$, $|x| \geq (r - \|B - A\|)|B^{-1}x|$, so if $x \neq 0$, $\frac{1}{r - \|B - A\|} \geq \frac{|B^{-1}x|}{|x|}$, then take sup on right-hand side and by definition of operator norm, we have desired result. \square

Remark. By taking $A = I_{n \times n}$, the value $r = \frac{1}{\|A^{-1}\|}$ is optimal.

Remark (Linear maps are continuous). $A \in \mathbb{R}^{k \times n}, x \mapsto Ax : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is continuous. Since $|Ax - Ay| = |A(x - y)| \leq \|A\| |x - y|$. So given $\varepsilon > 0$, can take $\delta = \frac{\varepsilon}{\|A\|}$ so $|x - y| < \delta \Rightarrow \|A\| |x - y| < \varepsilon \Rightarrow |Ax - Ay| < \varepsilon$. So linear maps are also uniformly continuous since δ is independent of x .

Definition 4.3.2 (Lipschitz). $f : U \rightarrow \mathbb{R}^k$ is Lipschitz if $\exists M > 0$ such that $|f(x) - f(y)| < M|x - y| \forall x, y \in U$.

5 Derivative of $f : U \rightarrow \mathbb{R}^k$ where $U \subset \mathbb{R}^n$ is open

Rate of change of a function of 2 or more variables depends on direction, not just the point. Simplest example: $f(x, y) = x$ changes in x direction but not in y . We therefore introduce directional derivative, i.e. rate of change along lines. [Compare with continuity along lines.]

5.1 Directional derivative

Fix $x \in U$ and a direction vector $v \neq 0$ (line would be $L_{x,v}$). In principle $|v| = 1$ but we'll not insist that. Parameterising $L_{x,v}$ we have $r(t) := x + tv$, and define $g_{x,v}(t) := f(x + tv) = f|_{L_{x,v}}$. Then

Definition 5.1.1.

$$\partial_v f(x) = \left. \frac{d}{dt} g_{x,v}(t) \right|_{t=0} = \left. \frac{d}{dt} f(x + tv) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} \in \mathbb{R}^k.$$

In G&M, ∂_v is denoted D_v .

Remark (Warning). If $\partial_v f(x)$ exists $\forall v \in \mathbb{R}^n \setminus \{0\}$, then f may not be continuous at x .

Example 5.1.2. Refer to Example 2.4.2. We have then $\partial_v f(0, 0) = 0$ since f is zero for any line through $(0, 0)$ but is not continuous at $(0, 0)$. We need a better definition of derivative such that if it is differentiable then it is continuous!

5.2 The (Fréchet) derivative as affine linear approximation

Definition 5.2.1. $\Delta : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is affine linear if $\exists A \in \mathbb{R}^{k \times n}, b \in \mathbb{R}^k : \Delta x = Ax + b$. So $b = 0 \Leftrightarrow \Delta$ is linear.

Recall 1-d: $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$ is not possible for $h \in \mathbb{R}^n$ as we cannot divide by vectors. But through magic of complex numbers we can in \mathbb{R}^2 . So we have two types of derivatives in \mathbb{R}^2 : complex derivative, and the one we are introducing now. Of course they are related.

For which n 's do we have a notion for multiplication in \mathbb{R}^n ? 1 and 2 of course. You may suggest cross product for 3, but bad news is $V \times V = 0$. We don't want zero for nonzero vectors multiplied by themselves. But later Hamilton described quaternion \mathbb{H} , which is valid multiplication on \mathbb{R}^4 , making cross product a special case. Bad news in this case is it's not commutative. Then people realised you can do multiplication in \mathbb{R}^8 , i.e. Cayley numbers or octonion \mathbb{O}And you stop there. Nothing more.

Back to derivative. We can rewrite the definition as

$$\lim_{h \rightarrow 0} \frac{|f(x + h) - f(x) - hf'(x)|}{|h|} = 0.$$

The trick is to regard $f'(x)$ as a 1×1 matrix and then $h \mapsto f'(x)h + f(x)$ is a affine linear map, which is a approximation of nonlinear map $h \mapsto f(x + h)$ fixing x . Then it's not hard to write the following:

Definition 5.2.2. $f : U \rightarrow \mathbb{R}^k$ is differentiable at $x \in U$ if $\exists A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ such that

$$\lim_{h \rightarrow 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0. \quad (1)$$

This is of course equivalent to that,

$$\forall \varepsilon > 0, \exists \delta > 0 : |h| < \delta \Rightarrow |f(x+h) - f(x) - Ah| \leq \varepsilon |h|. \quad (2)$$

Recall for $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$, $|Th| \leq \|T\||h|$. So given nonzero T , we can choose ε in (2) so that $0 < \varepsilon < \|T\|$ and (2) imply that the difference between the map $h \mapsto f(x+h)$ and the affine linear map $h \mapsto Ah + f(x)$ is smaller than $|Th|$ for any linear map T .

We can use this to prove that linear map A that satisfies the definition must be unique, if it exists at all. In other words, when A satisfying definition does exist, we denote it by $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$, Fréchet derivative of f at x . We can write

$$\forall \varepsilon > 0, \exists \delta > 0 : |h| < \delta \Rightarrow |f(x+h) - f(x) - Df(x)h| \leq \varepsilon |h|. \quad (3)$$

which essentially means that $h \mapsto Df(x)h + f(x)$ is the best affine linear approximation of $h \mapsto f(x+h)$ where x is fixed.

Proposition 5.2.3. Differentiability at x implies continuity at x .

Proof. By (3) we write

$$|f(x+h) - f(x)| \leq |f(x+h) - f(x) - Df(x)h| + |Df(x)h| \leq (\varepsilon + \|Df(x)\|)|h|.$$

So take

$$\delta^* = \min \left\{ \delta, \frac{\varepsilon}{\varepsilon + \|Df(x)\|} \right\},$$

where δ is again from (3), we then have $|h| < \delta^* \Rightarrow |f(x+h) - f(x)| \leq \varepsilon$. \square

Proposition 5.2.4 (Componentwise differentiability). $f : U \rightarrow \mathbb{R}^k$ is differentiable at x if $f_1 : U \rightarrow \mathbb{R}, \dots, f_k : U \rightarrow \mathbb{R}$ are all differentiable at x where $f(x) = (f_1(x), \dots, f_k(x))$.

Proposition 5.2.5 (Relation between derivative and directional derivative). If $Df(x)$ exists then $\partial_v f(x)$ exists $\forall v \in \mathbb{R}^n$ and

$$\partial_v f(x) = Df(x)v.$$

In particular, if f is differentiable at x then $\partial_v f(x)$ is linear in v .

Proof. Let $v \neq 0$. Set $h = tv$ in (1). So

$$\lim_{t \rightarrow 0} \frac{|f(x+tv) - f(x) - Df(x)(tv)|}{|t||v|} = 0.$$

Multiplying by length of $|v|$ we get

$$\lim_{t \rightarrow 0} \frac{f(x+tv) - f(x) - tDf(x)v}{t} = 0.$$

We can remove the norm since limit is zero, so limit of norm and limit of what's inside the norm is equivalent by definition. Therefore

$$\lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} = \lim_{t \rightarrow 0} \frac{tDf(x)v}{t} = Df(x)v$$

where the left hand side is exactly the definition of $\partial_v f(x)$. □

Linearity of $\partial_v f(x)$ means for $a, b \in \mathbb{R}$, $v, w \in \mathbb{R}^n$,

$$\begin{aligned} \delta_{av+bw} f(x) &= Df(x)(av + bw) = aDf(x)v + bDf(x)w \\ &= a\partial_v f(x) + b\partial_w f(x). \end{aligned}$$

5.3 Partial derivative, Jacobian matrix, gradient

Definition 5.3.1. For $1 \leq i \leq n$, $\partial_{v_i} f(x)$ is called *ith partial derivative* of f , or $\partial_i f(x) \in \mathbb{R}^k$, or $f_{x_i}(x)$, where v_i is the *ith* vector in standard basis.

$$\partial_i f(x) = \begin{pmatrix} \partial_i f_1(x) \\ \vdots \\ \partial_i f_k(x) \end{pmatrix}$$

where

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}.$$

Remark. Notation like $\frac{df}{dx}$ causes trouble with chain rule. Let $f(x) = \sin x$ and $g(x) = x^2$.

Example 5.3.2 (Compare partial derivatives with separate continuity). Let

$$f(x, y) = \begin{cases} 1 & xy \neq 0 \\ 0 & xy = 0 \end{cases}$$

(0 on axis and 1 everywhere else) Then $\partial_1 f(x, 0) = 0$ and $\partial_2 f(0, y) = 0$. But if $v \neq (a, 0), (0, b)$ then $\partial_v f(0, 0)$ does not exist.

Definition 5.3.3. $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^k$. The $k \times n$ *Jacobian matrix* is defined

$$\partial f(x) := \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \cdots & \partial_n f_k(x) \end{pmatrix}$$

where

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}.$$

So column-wise the matrix is

$$(\partial_1 f(x) \quad \cdots \quad \partial_n f(x)),$$

list of partial derivatives of f , and row-wise it is

$$\begin{pmatrix} \partial f_1(x) \\ \vdots \\ \partial f_k(x) \end{pmatrix},$$

list of Jacobians of the components.

Proposition 5.3.4. If $Df(x)$ exists then $\partial f(x)$ exists and

$$\underbrace{Df(x)}_{\in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)} h = \underbrace{\partial f(x)}_{\in \mathbb{R}^{k \times n}} h.$$

Remark. $\partial f(x)$ may exist even if $Df(x)$ does not. This is why we need this proposition.

Proof. For $h \in \mathbb{R}^n$, $h = h_1 v_1 + \dots + h_n v_n$. So

$$\begin{aligned} Df(x)h &= Df(x) \left(\sum_{i=1}^n h_i v_i \right) \\ &= \sum_{i=1}^n h_i Df(x)v_i && \text{linearity} \\ &= \sum_{i=1}^n h_i \partial_i f(x) && \text{Proposition 5.2.5} \\ &= \partial f(x)h. && \text{column-wise definition of Jacobian} \end{aligned}$$

□

5.4 Summary of different types of derivative [just a recap]

$f : U \rightarrow \mathbb{R}^k$, $U \subset \mathbb{R}^n$ open.

5.4.1 Directional derivative

$\partial_h f(x)$, $h \in \mathbb{R}^n \setminus \{0\}$ where h is the ‘direction’ (a vector in \mathbb{R}^k not a real number) is defined

$$\partial_h f(x) := \lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t} = \left. \frac{d}{dt} f(x + th) \right|_{t=0}.$$

Geometric interpretation: consider $g = f|_{L_{x,h}}$, i.e. $g_{x,h}(t) := f(x + th)$, $g_{x,h} : (-\tau, \tau) \rightarrow \mathbb{R}^k$ which parameterises a curve $\gamma \in \mathbb{R}^k$ through $f(x) = g_{x,h}(0)$. So $\partial_h f(x) = g'_{x,h}(0) =$ tangent (velocity) vector to γ at $f(x)$

5.4.2 Partial derivative

It’s just a special case of directional derivative.

Let v_i ’s be the standard basis of \mathbb{R}^n , then the i th partial derivative $\partial_i f(x) := \partial_{v_i} f(x)$, e.g. $\partial_1 f(x) = \partial_{(1,0,\dots,0)} f(x)$.

This is also written as $\frac{\partial f}{\partial x_i}(x)$ (because when computing, $\frac{d}{dt}f(x + tv_i)$ only changes the i th component of x) or $f_{x_i}(x)$.

Then we see given $h = (h_1, \dots, h_n)$, we can write it as linear combination of standard basis $h_1v_1 + \dots + h_nv_n$, so

$$\partial_h f(x) = \partial_{h_1v_1 + \dots + h_nv_n} f(x) \underset{\text{under condition below}}{=} h_1\partial_1 f(x) + \dots + h_n\partial_n f(x).$$

If $Df(x)$ exists then $Df(x)h = \partial_h f(x)$, and so, since $Df(x)$ is linear, then

$$\begin{aligned} \partial_h f(x) &= Df(x)h = Df(x)(h_1v_1 + \dots + h_nv_n) \\ &= h_1Df(x)v_1 + \dots + h_nDf(x)v_n \\ &= h_1\partial_1 f(x) + \dots + h_n\partial_n f(x). \end{aligned}$$

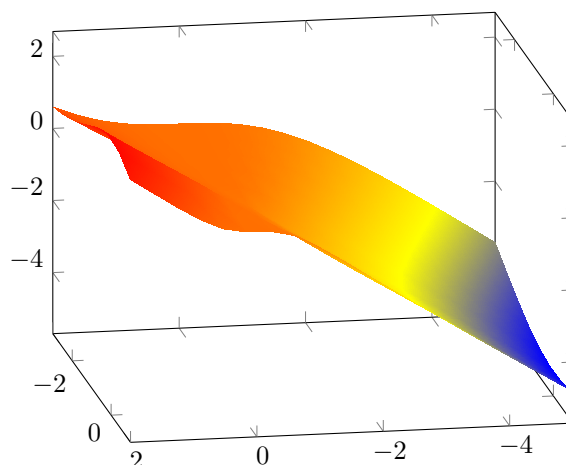
In plain English, if f is (Fréchet) differentiable, then $\partial_h f(x)$ is linear in h .

(In that case, we can know all directional derivatives just by calculating the n partial derivatives.)

Reverse is clearly not true, recall the naughty function defined in Example 2.4.2.

Example 5.4.1. $f : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$f(x, y) = \begin{cases} \frac{x^3}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$



1. Calculate $\partial_v f(0, 0)$.

Set $v = (a, b)$. Then

$$\partial_v f(0, 0) = \lim_{t \rightarrow 0} \frac{f(ta, tb) - f(0, 0)}{t} = \lim_{t \rightarrow 0} \frac{\frac{t^3 a^3}{t^2 a^2 + t^2 b^2} - 0}{t} = \frac{a^3}{a^2 + b^2}.$$

2. Calculate $\partial f(0, 0)$.

This is

$$(\partial_1 f(0, 0) \quad \partial_2 f(0, 0)) = (\partial_{(1,0)} f(0, 0) \quad \partial_{(0,1)} f(0, 0)) = (1 \quad 0).$$

What you do not do is differentiate f with respect to x and set $x = y = 0$, since it's not defined there.

3. Calculate $\partial f(0,0)h$.

This is $\begin{pmatrix} 1 & 0 \\ a & b \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = a \neq \partial_h f(0,0)$ unless $b = 0$, i.e. $\partial_h f(0,0)$ not linear in h . Therefore $Df(0,0)$ does not exist.

4. Calculate $f_x(x,y)$, $f_y(x,y)$ where $(x,y) \neq (0,0)$.

We get $f_x(x,y) = \frac{x^2(x^2+3y^2)}{(x^2+y^2)^2}$, $f_y(x,y) = -\frac{2x^3y}{(x^2+y^2)^2}$. Are they continuous at $(0,0)$? We write

$$\lim_{y \rightarrow 0} f_x(0,y) = 0 \neq f_x(0,0) = 1,$$

and similarly

$$\lim_{x \rightarrow 0} f_y(x,0) = -\frac{1}{2} \neq f_y(0,0) = 0,$$

so neither is continuous.

5.5 Gradient of $f : U \rightarrow \mathbb{R}$, $U \subset \mathbb{R}^n$

f is real-valued not vector-valued!

We then have Jacobian matrix as a row matrix:

$$\partial f(x) = (\partial_1 f(x), \dots, \partial_n f(x))$$

which is a linear functional, i.e. a map $\mathbb{R}^n \rightarrow \mathbb{R}$. Then

$$\partial f(x)h = \underbrace{(\partial_1 f(x), \dots, \partial_n f(x)) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}}_{\text{matrix multiplication}} = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(x).$$

We then define the *gradient* of f at x , $\nabla f(x) = (\partial f(x))^T$, a column vector. Note that $\partial f(x)h = \underbrace{\nabla f(x) \cdot h}_{\text{dot product}}$.

5.6 Differentiation of matrix-valued functions

Remember $\mathbb{R}^{k \times n} \leftrightarrow \mathbb{R}^{kn}$ so there's nothing scary.

Example 5.6.1. Define $f : \mathcal{L}(\mathbb{R}^n) \rightarrow \mathcal{L}(\mathbb{R}^n)$ (or $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$) by $f(T) = T^2$.

We look at the special case $n = 2$, so $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$f(T) = T^2 = \begin{pmatrix} a^2 + bc & ab + bd \\ ca + dc & cb + d^2 \end{pmatrix}.$$

As a map $f : \mathbb{R}^4 \rightarrow \mathbb{R}^4$, this is nothing but

$$f(a,b,c,d) = (a^2 + bc, ab + bd, ca + dc, cb + d^2).$$

And as you can see derivative is not $2T$. Let's use definition to differentiate f . Consider

$$f(T+H) - f(T) = Df(T)H$$

and we need to guess what $Df(T)$ is. The left hand side is equal to

$$(T + H)(T + H) - T^2 = T^2 + TH + HT + H^2 - T^2 = TH + HT + H^2.$$

Just to remind you, $Df(T)$ is an element of $\mathcal{L}(\mathcal{L}(\mathbb{R}^n), \mathcal{L}(\mathbb{R}^n))$, so it has to be linear. We see that $TH + HT$ is linear in H . We then guess

$$Df(T)H = TH + HT.$$

We now verify the guess:

$$\begin{aligned} \lim_{H \rightarrow 0} \frac{\|f(T + H) - f(T) - Df(T)H\|}{\|H\|} &= \lim_{H \rightarrow 0} \frac{\|H^2\|}{\|H\|} \\ &\leq \lim_{H \rightarrow 0} \frac{\|H\|\|H\|}{\|H\|} \\ &= \lim_{H \rightarrow 0} \|H\| = 0. \end{aligned}$$

So we win! Recall if $f(t) = t^2$ then $f'(t) = 2t$. So we get an analogous result, different though only because matrix multiplication is not commutative.

Since we're having so much fun, why don't we calculate directional derivative $\partial_H f(T)$? We need to restrict f' to the line $L_{T,H}$ through T in the direction of H . Let $r(t) := T + tH$ and we write

$$g_{T,H} = f|_{L_{T,H}} = g(t) = f(r(t)) = f(T + tH) = (T + tH)^2 = T^2 + t(TH + HT) + t^2 H^2.$$

Then

$$\partial_H f(T) = \frac{d}{dt} f(T + tH)|_{t=0} = TH + HT + 2tH^2|_{t=0} = TH + HT = Df(T)H.$$

We just proved that if derivative exists directional derivative is just derivative operator acting on direction, as expected.

We can also write its Jacobian matrix:

$$\partial f(a, b, c, d) = \begin{pmatrix} 2a & c & b & 0 \\ b & a + d & 0 & b \\ c & 0 & a + d & 0 \\ 0 & c & b & 2d \end{pmatrix}$$

5.7 Chain rule

Theorem 5.7.1. Suppose

- $U \subset \mathbb{R}^n$ open, $V \subset \mathbb{R}^k$ open
- $f : U \rightarrow \mathbb{R}^k$ differentiable at $x \in U$ and $f(x) \in V$
- $g : V \rightarrow \mathbb{R}^m$ differentiable at $f(x)$

Conclusion: $g \circ f : U \rightarrow \mathbb{R}^m$ is differentiable at x and

$$\underbrace{(Dg \circ f)(x)}_{\in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)} = \underbrace{(Dg(f(x)))}_{\in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^m)} \circ \underbrace{Df(x)}_{\in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)}.$$

So we also have Jacobian form of chain rule:

$$\begin{aligned}\partial g \circ f(x) &= \partial g(f(x)) \cdot \partial f(x) \\ &= \begin{pmatrix} \partial_1 g_1(f(x)) & \cdots & \partial_k g_1(f(x)) \\ \vdots & & \vdots \\ \partial_1 g_m(f(x)) & \cdots & \partial_k g_m(f(x)) \end{pmatrix} \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \cdots & \partial_n f_k(x) \end{pmatrix}.\end{aligned}$$

5.7.1 Common special cases

Case 1: $k = m = 1$. In this case $\partial f(x) = (\partial_1 f(x), \dots, \partial_n f(x))$, a row vector, and $\partial g(x) = g'(x)$, a number. Then $\partial g \circ f(x) = g'(f(x))(\partial_1 f(x), \dots, \partial_n f(x))$.

Example 5.7.2. Calculate $\nabla(|x|)$ where $x \in \mathbb{R}^n \setminus \{0\}$.

Apply chain rule to $f(x) = |x|^2 = x_1^2 + \cdots + x_n^2$, $g(t) = \sqrt{t}$, $t > 0$. We write

$$\partial f(x) = (2x_1, 2x_2, \dots, 2x_n), \text{ i.e. } \nabla f(x) = 2x.$$

and $g'(t) = \frac{1}{2\sqrt{t}}$. Then

$$\partial|x| = \frac{1}{2\sqrt{|x|^2}}(2x_1, \dots, 2x_n) = \frac{1}{|x|}(x_1, \dots, x_n), \text{ i.e. } \nabla|x| = \frac{x}{|x|}.$$

Case 2: $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $r : \mathbb{R} \rightarrow \mathbb{R}^n$ (**parameterisation of a curve in \mathbb{R}^n**). We can write

$$r(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} \text{ and } f \circ r : \mathbb{R} \rightarrow \mathbb{R}. \text{ Then}$$

$$\begin{aligned}\partial f \circ r(t) &= (f \circ r)'(t) = \partial f(r(t))r'(t) \\ &= (\partial_1 f(r(t)), \dots, \partial_n f(r(t))) \begin{pmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{pmatrix} \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(r(t)) \frac{dx_i}{dt} \\ &= \nabla f(r(t)) \cdot \frac{dr}{dt} = \nabla(f(r(t))) \cdot r'(t)\end{aligned}$$

Example 5.7.3. Fix $x \in \mathbb{R}^n$ and define $r : \mathbb{R} \rightarrow \mathbb{R}^n$ by $r(t) = tx$ (a line through origin in direction of x).

Given $f : \mathbb{R}^n \rightarrow \mathbb{R}$, calculate $(f \circ r)'(t)$.

What is $r'(t)$? It's x , of course. Then

$$(f \circ r)'(t) = \nabla f(tx) \cdot x.$$

5.7.2 Idea of proof of chain rule (seen in first year analysis)

We need a reformulation of definition of derivative.

Lemma 5.7.4. Given $f : U \rightarrow \mathbb{R}^k$ and $x \in U$. Since U is open, $\exists r > 0 : \mathbb{B}(x, r) \subset U$. Also we are given a linear map $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$. Then define

$$\Delta_{x,A}f(h) = \begin{cases} \frac{f(x+h) - f(x) - Ah}{|h|} & h \neq 0 \\ 0 & h = 0 \end{cases}$$

Then f is differentiable at x with $Df(x) = A$ if and only if $\Delta_{x,A}f$ is continuous at 0.

Proof. $\Delta_{x,A}f$ is continuous at 0 $\Leftrightarrow \lim_{h \rightarrow 0} \Delta_{x,A}f(h) = \Delta_{x,A}f(0) = 0 \Leftrightarrow Df(x)$ exists and equals A by definition of derivative. \square

Notation. If $Df(x)$ exists, abbreviate $\Delta_{x,Df(x)}f$ to $\Delta_x f$.

Note that

$$f \text{ differentiable at } x \Leftrightarrow f(x+h) = f(x) + Df(x)h + (\Delta_x f(x))|h|$$

and similarly

$$g \text{ differentiable at } y \Leftrightarrow g(y+k) = g(y) + Dg(y)k + (\Delta_y g(y))|k|.$$

We need to show

$$\begin{aligned} g \circ f(x+h) &= g(f(x+h)) \\ &= g(\underbrace{f(x)}_y + \underbrace{Df(x)h + (\Delta_x f(x))|h|}_k) \\ &= g(f(x)) + Dg(f(x))(Df(x)h + \Delta_x f(x)|h|) \\ &\quad + (\Delta_{f(x)}g(Df(x)h + \Delta_x f(x)|h|))|Df(x)h + \Delta_x f(x)|h||. \\ &= g \circ f(x) + Dg(f(x))(Df(x)h) + \text{the rest which goes to zero} \end{aligned}$$

Details in the typewritten notes.

5.8 Continuity of partial derivatives implies differentiability

Theorem 5.8.1. Given

- $f : U \rightarrow \mathbb{R}^k$, $U \subset \mathbb{R}^n$ open
- $x \in U$, so $\exists r > 0 : \mathbb{B}(x, r) \subset U$
- Jacobian matrix $\partial f(y)$ exists $\forall y \in U$
- ∂f is continuous at x

then f is differentiable at x and

$$Df(x)h = \partial f(x)h \quad \forall h \in \mathbb{R}^n.$$

Remark. ∂f is continuous at x if and only if $\partial_1 f_1, \dots, \partial_n f_1, \dots, \partial_1 f_k, \dots, \partial_n f_l$ are all continuous at x . So all the pathology concerning $\partial f(x)h \neq \partial_h f(x)$ can only arise if at least one of the partial derivatives is not continuous at x .

Basic ideas in the proof. We'll demonstrate this by fixing $n = 2$, $k = 1$, i.e. $f(x_1, x_2) \in \mathbb{R}$. Full proof in appendix D in typewritten notes.

We know that if $Df(x)$ exists it satisfies

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x) - Df(x)h}{|h|} = 0$$

so we just need to guess what the linear map $Df(x)$ is. In particular, it follows that

$$Df(x_1, x_2) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = (\partial_1 f(x_1, x_2))h_1 + (\partial_2 f(x_1, x_2))h_2 = (\partial f(x_1, x_2)) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

So we consider

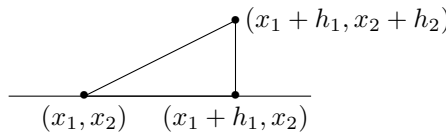
$$\Delta f(h_1, h_2) := \underbrace{f(x_1 + h_1, x_2 + h_2)}_{f(x+h)} - \underbrace{f(x_1, x_2)}_{f(x)} - \underbrace{(h_1 \partial_1 f(x_1, x_2) + h_2 \partial_2 f(x_1, x_2))}_{\text{our guess for } Df(x)h}$$

and we need to show

$$\lim_{(h_1, h_2) \rightarrow (0,0)} \frac{|\Delta f(h_1, h_2)|}{|(h_1, h_2)|} = 0.$$

If we succeed, we would manage to have two things done: prove that f is differentiable, and verify that the derivative is what it has to be.

Partial derivatives provide information only along lines parallel to axes. So we go from (x_1, x_2) to $(x_1 + h_1, x_2 + h_2)$ via $(x_1 + h_1, x_2)$.



Computationally this means

$$\begin{aligned} & f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) \\ &= \underbrace{f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2)}_{\text{II}} + \underbrace{f(x_1 + h_1, x_2) - f(x_1, x_2)}_{\text{I}} \\ &= h_2 \partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2) + h_1 \partial_1 f(x_1 + \theta_1 h_1, x_2) \quad \text{for some } \theta_1, \theta_2 \in (0, 1) \text{ by MVT} \end{aligned}$$

So

$$\begin{aligned} \Delta f(h_1, h_2) &= h_1 [\partial_1 f(x_1 + \theta_1 h_1, x_2) - \partial_1 f(x_1, x_2)] \\ &\quad + h_2 [\partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2) - \partial_2 f(x_1, x_2)] \end{aligned}$$

and by continuity of $\partial f(x_1, x_2)$, given $\varepsilon > 0$,

$$\exists \delta > 0 : 0 < |(\tilde{h}_1, \tilde{h}_2)| < \delta \Rightarrow |\partial f(x_1 + \tilde{h}_1, x_2 + \tilde{h}_2) - \partial f(x_1, x_2)| < \varepsilon.$$

And we know that if $|(h_1, h_2)| < \delta$ then we also have

$$|(x_1 + \theta_1 h_1, x_2)| = \theta_1 |h_1| \leq \theta_1 |(h_1, h_2)| < \theta_1 \delta < \delta$$

and

$$|(x_1 + h_1, x_2 + \theta_2 h_2) - (x_1, x_2)| = \sqrt{(h_1)^2 + (\theta_2 h_2)^2} \leq |(h_1, h_2)| < \delta.$$

so we have

$$|\partial_1 f(x_1 + \theta_1 h_1, x_2) - \partial_1 f(x_1, x_2)| \leq |\partial f(x_1 + \theta_1 h_1, x_2) - \partial f(x_1, x_2)| < \varepsilon$$

and similarly

$$|\partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2) - \partial_2 f(x_1, x_2)| < \varepsilon.$$

So as long as $0 < |(h_1, h_2)| < \delta$ we have

$$\frac{|\Delta f(h_1, h_2)|}{|(h_1, h_2)|} \leq \frac{\varepsilon(|h_1| + |h_2|)}{\sqrt{h_1^2 + h_2^2}} < 2\varepsilon.$$

□

Proposition 5.8.2 (Very useful). $f : U \rightarrow \mathbb{R}^k$ is continuously differentiable at all points of U , i.e.

$$x \mapsto Df(x) : U \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) \text{ is continuous } \forall x \in U$$

if and only if

$$x \mapsto \partial f(x) : U \rightarrow \mathbb{R}^{k \times n} \text{ is continuous } \forall x \in U.$$

Remark. Recall that by equivalence of norms, a function $F : U \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ is continuous (with respect to operator norm) if and only if $\mu(F) : U \rightarrow \mathbb{R}^{k \times n}$ is continuous (with respect to Frobenius norm).

Proof. Suppose $Df(x)$ is continuous at all points $x \in U$. Then $\partial f(x) = \mu(Df(x))$ is continuous at all points $x \in U$. This follows immediately from the remark.

Now suppose $\partial f(x)$ is continuous at all points $x \in U$. This is slightly trickier because we don't know if $Df(x)$ exists or not, so we cannot directly apply the remark. But fortunately by Theorem 5.8.1, it does. So $Df(x)$ is continuous at all points $x \in U$ by continuity of ∂f . □

We can now check existence and continuity of Df simply by checking those of ∂f , which is just A-level stuff.

Notation. $C^1(U, \mathbb{R}^k) = \{f : U \rightarrow \mathbb{R}^k \mid \partial f : U \rightarrow \mathbb{R}^k \text{ is continuous}\}.$

$C^1(U, \mathbb{R})$ is abbreviated to $C^1(U)$.

5.9 Mean value inequality (MVI)

Recall 1-d MVT: $\exists c \in (x, y) : \frac{f(y) - f(x)}{y - x} = f'(c)$. It's more useful in this form: if $|f'(u)| \leq M \forall u \in (x, y)$ then $\underbrace{|f(y) - f(x)| \leq M|y - x|}_{\text{MVI}}$, i.e. Lipschitz.

Proposition 5.9.1 (Generalised MVI). Given $x, y \in U$ open, suppose \exists a continuously differentiable path $r : [a, b] \rightarrow U$ such that $r(a) = x$, $r(b) = y$. Suppose $f \in C^1(U, \mathbb{R}^k)$ and that $\exists M \geq 0$ such that $\|\partial f(u)\| \leq M \forall u \in U$. Then $|f(x) - f(y)| \leq M \text{ length}(C_{xy})$ where $C_{xy} = r([a, b])$.

Proof. We have

$$f(y) - f(x) = f(r(b)) - f(r(a)) \stackrel{\text{FTC}}{=} \int_a^b \frac{d}{dt} f(r(t)) dt$$

where $\int_a^b g(t) dt := \left(\int_a^b g_1(t) dt, \dots, \int_a^b g_k(t) dt \right)$. How do we differentiate that? Chain rule of course:

$$= \int_a^b \partial f(r(t)) \cdot r'(t) dt$$

So

$$\begin{aligned} |f(y) - f(x)| &\leq \int_a^b |\partial f(r(t)) \cdot r'(t)| dt \leq \int_a^b \|\partial f(r(t))\| |r'(t)| dt \\ &\leq M \int_a^b |r'(t)| dt \quad \text{by assumption} \\ &= M \text{length}(C_{xy}) \quad \text{by definition of length} \end{aligned}$$

□

Definition 5.9.2 (Path connectedness). $U \subset \mathbb{R}^n$ is (differentiably) path connected if $\forall p, q \in U$, \exists a continuously differentiable path $r : [a, b] \rightarrow U$ such that $r(a) = p$, $r(b) = q$.

It's really a notion in topology. Differentiability doesn't make much sense in those things since it implies some kind of linear structure. This is why differentiability is with parentheses. We'll see a lot more about this in Norms, Metrics and Topologies.

Corollary 5.9.3 (Vanishing derivative). Suppose $U \subset \mathbb{R}^n$ is differentiably path connected and $f : U \rightarrow \mathbb{R}^k$ satisfies $\partial f(x) = 0 \forall x \in U$. Then f is constant on U .

Proof. Fix $y \in U$. Given another $x \in U$, $\exists r : [a, b] \rightarrow U : r(a) = x$, $r(b) = y$. Then $f(y) - f(x) = \underbrace{\int_a^b \partial f(r(t)) r'(t) dt}_0 = 0$, i.e. $f(x) = f(y) \forall x \in U$. □

Definition 5.9.4. $U \subset \mathbb{R}^n$ is convex if $\forall x, y \in U$, the straight line L_{xy} joining x to y is contained in U .

Corollary 5.9.5 (Mean value inequality). $U \subset \mathbb{R}^n$ is convex. If $f \in C^1(U, \mathbb{R}^k)$ and $\|\partial f(x)\| \leq M \forall x \in U$. Then $|f(x) - f(y)| \leq M|x - y|$.

Remark. f being Lipschitz does not imply that $\|\partial f(x)\| \leq M$, nor does it imply f is differentiable. e.g. $f(x) = |x|$ or $\frac{x^3}{x^2+y^2}$.

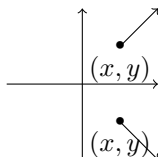
6 Vector fields, curves, line integrals

This is mainly review from MA134.

Definition 6.0.1. A vector field \underline{v} on $U \subset \mathbb{R}^n$ is a function $\underline{v} : U \rightarrow \mathbb{R}^n$, $\underline{v}(x) = (v_1(x_1, \dots, x_n), \dots, v_n(x_1, \dots, x_n))$.

Why is \underline{v} underlined? To view $\underline{v}(x)$ as “attached” at x .

Example 6.0.2. Consider $\underline{v}(x, y) = (x, y)$. You would view this as a map, in fact the identity map. But we want to view it as a whole bunch of vectors attached to every vector in \mathbb{R}^2 .



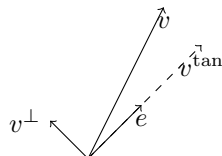
Also consider $\underline{v}(x, y) = \frac{1}{x^2+y^2}(x, y)$, so the length of the vector attached to (x, y) is $\frac{1}{\sqrt{x^2+y^2}}$, i.e. it decays.

6.1 Components of a vector

Let $e \in \mathbb{R}^n$ be a unit vector. Given $v \in \mathbb{R}^n$ set $v^{\text{tan}} = (v \cdot e)e$. Observe

$$(v - v^{\text{tan}}) \cdot e - v \cdot e - (v \cdot e)(e \cdot e) = 0,$$

i.e. $v - v^{\text{tan}} \perp e$. So we set $v^\perp := v - v^{\text{tan}}$ and $v = v^{\text{tan}} + v^\perp$.



6.2 Curves and their parameterisation

Definition 6.2.1. A *curve* C is a subset of \mathbb{R}^n which can be realised as the image of a continuous map $r : [a, b] \rightarrow \mathbb{R}^n$, which is called a (not unique!) *parameterisation* of C .
could also be $[0, \infty)$ etc.

Example 6.2.2. C = line through $0 \in \mathbb{R}^2 \perp (a, b) = \{(x, y) : ax + by = 0\}$. A natural parameterisation of C would be $r(u) = (-bu, au)$. But we can also take something like $r(u) = -b \sinh u + a \cosh u$.

Parameterisation is a kinematic description of how a point $u \in [a, b]$ moves along C . Definition of curve allows weird examples like space-filling curves. We want to restrict ourselves to non-weird curves:

Definition 6.2.3. A *regular* curve is one which admit *regular* parameterisation:

- r is continuously differentiable
- $r'(u) \neq 0 \forall u \in [a, b]$ (you cannot momentarily stop)

Read appendix F of typewritten notes for demystification of arc length.

Consider regular C with parameterisation $r : [a, b] \rightarrow \mathbb{R}^n$. Length of segment of C between $r(a)$ and $r(u)$ is denoted $s(u)$,

$$s(u) := \int_a^u |r'(\tau)| \, d\tau.$$

But how do we know it does not depend on different parameterisation? By change of variables formula:

$$\frac{ds}{du} \underbrace{=}_{\text{FTC}} |r'(u)| \underbrace{\geq 0}_{\text{regular}}, \quad s(a) = 0$$

this is an initial value ODE problem which has a solution (but typically cannot be solved explicitly).

Let $L = \text{length of } C = s(b)$. Then $s : [a, b] \rightarrow [0, L]$ is clearly bijection by monotonicity. So the inverse exists, and we call it u (since the function is a point in the domain interval). Note that $u(s(u)) = u$, $s(u(s)) = s$. We then can give the following definition:

Definition 6.2.4. Arc length parameterisation ρ of C is defined by

$$\rho(s) = r(u(s)).$$

Note by chain rule

$$\frac{d\rho}{ds} = r'(u(s)) \frac{du}{ds} = \overbrace{r'(u(s))}^{\text{view as function of } u} \frac{ds}{du},$$

i.e. $\rho'(s(u)) = \frac{r'(u)}{|r'(u)|}$, a unit vector, i.e. arc length parameterisation is unit speed, i.e. $\rho'(s) = t(s)$ = unit tangent to C at $\rho(s)$.

6.3 Line integral

Definition 6.3.1. Given $f \in C(\mathbb{R}^n)$ and regular curve $C \subset \mathbb{R}^n$, the *line integral* of f along C is defined

$$\int_C f \, ds := \int_0^L f(\rho(s)) \, ds.$$

But we've been warned that this is in general not computable. So what do we do? Note

$$\int_0^L f(\rho(s)) \, ds \quad \underbrace{=}_{\text{change variable to } u} \int_a^b f(r(u)) \frac{ds}{du} du = \int_a^b f(r(u)) |r'(u)| \, du,$$

which is the computation formula of line integral, and does not depend on choice of r as long as its regular.

Definition 6.3.2. Given continuous vector field $\underline{v}(s)$ and a regular curve $C \subset \mathbb{R}^n$. The *tangential line integral* of \underline{v} along C is defined

$$\int_C (\underline{v} \cdot t) \, ds.$$

Remark. $\underline{v} \cdot t > 0$ means \underline{v} points along t and $\underline{v} \cdot t < 0$ means \underline{v} points against t . Now $\frac{\int_C \underline{v} \cdot t \, ds}{\int_C 1 \, ds}$ is the average of $\underline{v} \cdot t$ along C where the denominator is just length of C . So

$$\int_C \underline{v} \cdot t \, ds = (\text{length of } C) \underbrace{(\text{average of } \underline{v} \cdot t \text{ along } C)}_{\text{how much "on average" } \underline{v} \text{ points along } C}$$

To compute this we change variables again:

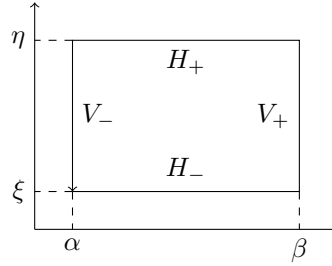
$$= \int_a^b \underline{v}(r(u)) \cdot \frac{dr}{du} du.$$

Definition 6.3.3. If C is a closed curve in \mathbb{R}^n then

$$\oint_C \underline{v} \cdot dr$$

is often called the *circulation* of \underline{v} around C (because this is positive if, on average, \underline{v} flows around C). The arrow indicates curve is oriented.

Circulation around rectangle $R = [\alpha, \beta] \times [\xi, \eta] \subset \mathbb{R}^2$



If we parameterise H_- we have $r(x) = (x, \xi)$, $\alpha \leq x \leq \beta$ and

$$\int_{H_-} \underline{v} \cdot dr = \int_{\alpha}^{\beta} \underline{v}(r(x)) \cdot r'(x) dx = \int_{\alpha}^{\beta} (a(x, \xi), b(x, \xi)) \cdot (1, 0) dx = \int_{\alpha}^{\beta} a(x, \xi) dx.$$

And similarly for V_+ we have $r(y) = (\beta, y)$ and

$$\int_{V_+} \underline{v} \cdot dr = \int_{\xi}^{\eta} b(\beta, y) dy, \quad \int_{H_+} \underline{v} \cdot dr = - \int_{\alpha}^{\beta} a(x, \eta) dx$$

and so for V_-

$$\int_{V_-} \underline{v} \cdot dr = - \int_{\xi}^{\eta} b(\alpha, y) dy$$

and finally

$$\begin{aligned} \oint_R \underline{v} \cdot dr &= \int_{\alpha}^{\beta} (a(x, \xi) - a(x, \eta)) dx + \int_{\xi}^{\eta} (b(\beta, y) - b(\alpha, y)) dy \\ &= \int_{\alpha}^{\beta} \left(- \int_{\xi}^{\eta} \frac{\partial a}{\partial y}(x, y) dy \right) dx + \int_{\xi}^{\eta} \left(\int_{\alpha}^{\beta} \frac{\partial b}{\partial x}(x, y) dx \right) dy \quad \text{by FTC} \\ &= \iint_R \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y} \right) \underbrace{dA_{x,y}}_{\text{in Euclidean it's just } dx dy} \quad \text{by Fubini} \end{aligned}$$

This is Green's theorem in its most specific case. We want to extend this to general regions. First question is what direction should we go? Orientation of boundary is tricky. We formulate these questions with some definitions first.

Definition 6.3.4. A bounded open subset $\Omega \subset \mathbb{R}^n$ is called a *region* if $\exists f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

1. all partial derivatives of f exist and are continuous
2. $\Omega = \{x \in \mathbb{R}^n : f(x) < 0\} = f^{-1}((-\infty, 0))$
3. $\nabla f(p) \neq 0 \forall p \in f^{-1}\{0\}$

f is called a *defining function* of Ω .

Condition 3 implies for $p \in f^{-1}\{0\}$ we can set $n_+(p) = \frac{\nabla f(p)}{|\nabla f(p)|}$ which is perpendicular to $f^{-1}\{0\}$ by proof we've seen in MA134. Then $\exists \delta > 0 : f(p + tn_+(p)) > 0 \ \forall t \in (0, \delta)$ and $f(p + tn_+(p)) < 0 \ \forall t \in (-\delta, 0)$. (To see this see Q5 of Ex 6). Geometrically, $n_+(p)$ is outward unit normal and $f^{-1}\{0\}$ is *boundary* of Ω , denoted $\partial\Omega$.

Example 6.3.5. $f(x, y, z) = x^2 + y^2 + z^2 - 4$. Then $\Omega = f^{-1}((-\infty, 0)) = \{x^2 + y^2 + z^2 < 4\} =$ ball centred at $(0, 0)$ of radius 2, and $\partial\Omega = \{0\} = S^2(2)$. Also $\nabla f = (2x, 2y, 2z) \neq (0, 0, 0)$ when $(x, y, z) \in \partial\Omega$, and $n_+(x, y, z) = \frac{1}{2\sqrt{x^2+y^2+z^2}}(2x, 2y, 2z) = \frac{1}{2}(x, y, z)$.

If we are in \mathbb{R}^2 , $f^{-1}\{0\}$ would be a regular curve and in \mathbb{R}^3 it's a regular surface. This requires implicit function theorem which will be proved near the end of term. Now we focus in \mathbb{R}^2 .

Definition 6.3.6. Given region $\Omega \in \mathbb{R}^2$, $p \in \Omega$ and $n_+(p) = (h(p), k(p))$. Rotate n_+ by $\frac{\pi}{2}$ anticlockwise to get *positively oriented tangent* $t_+(p) := (-k(p), h(p))$.

Theorem 6.3.7 (Green's theorem in plane). Denote $\Omega \cup \partial\Omega$ as $\bar{\Omega}$ (called *closure* of Ω), which is a subset of $U \subset \mathbb{R}^2$ open. Given vector field $\underline{v} \in C^1(U, \mathbb{R}^2)$, $\underline{v}(x, y) = (a(x, y), b(x, y))$, define

$$\text{curl } \underline{v}(x, y) := \frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}.$$

Then

$$\iint_{\Omega} \text{curl } \underline{v} \, dA = \oint_{\partial\Omega} \underline{v} \cdot t_+ \, ds \quad \underbrace{=}_{\text{computationally}} \quad \oint_{\partial\Omega} \underline{v} \cdot d\mathbf{r},$$

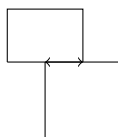
where s is an arc length parameterisation of $\partial\Omega$ and \mathbf{r}' is a positively oriented parameterisation of $\partial\Omega$, i.e. $\frac{\mathbf{r}'}{|\mathbf{r}'|} = t_+$.

Wait... the specific rectangle case is not a region by definition. So we extend the Ω in Green's theorem to *piecewise region*. (Similarly in analysis, if a function is only finitely different than a uniformly continuous function it's still integrable.)

Strategy of proof. Approximate Ω by a rectangle mesh R_ε such that area of $(\Omega \setminus R_\varepsilon) < \varepsilon$, then

$$\iint_{R_\varepsilon} \text{curl } \underline{v} \, dA \rightarrow \iint_{\Omega} \text{curl } \underline{v} \, dA$$

since their difference is at most $\max_{\Omega} \{\text{curl } \underline{v}\} \varepsilon$, which can be attained since $\text{curl } \underline{v}$ is continuous. After this we need to see Green's theorem holds for R_ε due to cancellation:



Finally we need to show

$$\int_{\partial R_\varepsilon} \underline{v} \cdot t \, ds \rightarrow \int_{\partial\Omega} \underline{v} \cdot t \, ds, \quad (*)$$

which is not at all obvious! We all know why the proof of $\pi = 4$ is invalid, right? We cannot simply approximate curves by stairs, i.e. sum of length of stairs is not C^1 convergence. So we need to proceed with more caution.

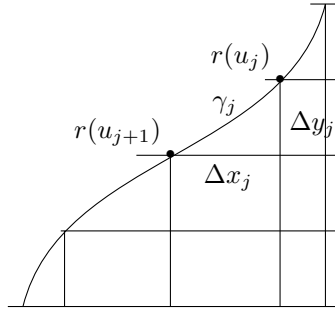
Parameterise $\partial\Omega$ by $r : [\alpha, \beta] \rightarrow \mathbb{R}^2$, partition $[\alpha, \beta]$ into

$$[\underset{=u_0}{\alpha}, u_1], [u_1, u_2], \dots, [u_{l-1}, \underset{=u_l}{\beta}].$$

and set $\gamma_j = r([u_{j-1}, u_j])$, $1 \leq j \leq l$. Then

$$\begin{aligned} \int_{\partial\Omega} \underline{v} \cdot dr &= \sum_{j=1}^l \int_{\gamma_j} \underline{v} \cdot dr \approx \sum_{j=1}^l \underline{v}(r(u_j)) \cdot (r(u_{j+1}) - r(u_j)) \\ &= \sum_{j=1}^l (a(r(u_j)), b(r(u_j))) \cdot (\Delta x_j, \Delta y_j) \\ &= \sum_{j=1}^l (a(r(u_j)), b(r(u_j))) \cdot (\Delta x_j, 0) + \sum_{j=1}^l (a(r(u_j)), b(r(u_j))) \cdot (0, \Delta y_j) \\ &\approx \int_{\Gamma_j} \underline{v} \cdot dr \end{aligned}$$

where Γ_j denotes the relevant part of boundary of R_ε .



Now we can show that $\varepsilon \rightarrow 0$ we have (*).

The modern argument uses something cleverer called partition of unity which again reduces Green's theorem to a rectangle, done in manifolds. (Breaking up functions to simpler ones is easier than breaking up domains.) \square

6.4 Flux and divergence in \mathbb{R}^2

If $v = (a, b)$, define $v^\perp := (b, -a)$ = clockwise rotation of v by $\frac{\pi}{2}$, we have

- $v \cdot w = v^\perp \cdot w^\perp$
- $(v^\perp)^\perp = -v$
- if n_+ is outward unit normal then $t_+ = -n_+^\perp$

So clearly this \perp has some nice properties. Natural question is what's Green's theorem applied to v^\perp ? If $\underline{v}(x, y) = (a(x, y), b(x, y))$ then

$$\text{curl } \underline{v}^\perp = \text{curl } (b, -a) = \frac{\partial}{\partial x}(-a) - \frac{\partial}{\partial y}b = -\left(\frac{\partial a}{\partial x} + \frac{\partial b}{\partial y}\right).$$

This leads to

Definition 6.4.1. The *divergence* of \underline{v} , denoted $\nabla \cdot \underline{v}$ or $\text{div } \underline{v}$, is

$$\frac{\partial a}{\partial x} + \frac{\partial b}{\partial y} \quad \text{or written as} \quad \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) \cdot (a, b).$$

So

Theorem 6.4.2 (Divergence (Gauss's) theorem (and its proof)).

$$\begin{aligned} \iint_{\Omega} \nabla \cdot \underline{v} \, dA &= - \iint_{\Omega} \text{curl } \underline{v}^{\perp} \, dA \\ &= - \oint_{\partial\Omega} \underline{v}^{\perp} \cdot \underline{t}_+ \, ds \quad \text{by Green's} \\ &= - \oint_{\partial\Omega} \underline{v}^{\perp} \cdot (-\underline{n}_+^{\perp}) \, ds \\ &= \underbrace{\oint_{\partial\Omega} \underline{v} \cdot \underline{n}_+ \, ds}_{\text{flux}}. \end{aligned}$$

Example 6.4.3. We have 2 vector fields $\underline{w}(x, y) = (x, y)$ and $\underline{v}(x, y) = \frac{1}{x^2 + y^2}(x, y)$. Then

$$\begin{aligned} \nabla \cdot \underline{w} &= \frac{\partial}{\partial x} x + \frac{\partial}{\partial y} y = 2 \\ \nabla \cdot \underline{v} &= \frac{\partial}{\partial x} \frac{x}{x^2 + y^2} + \frac{\partial}{\partial y} \frac{y}{x^2 + y^2} = 0. \end{aligned} \tag{*}$$

Now let C_1 be unit circle. We want to calculate flux of \underline{v} across C_1 . Clearly $\underline{n}_+(x, y) = (x, y)$, so on C_1 ,

$$\underline{v}(x, y) \cdot \underline{n}_+(x, y) = \frac{1}{x^2 + y^2}(x, y) \cdot (x, y) = 1.$$

Therefore the flux is

$$\int_{C_1} 1 \, ds = \text{length of } C_1 = 2\pi.$$

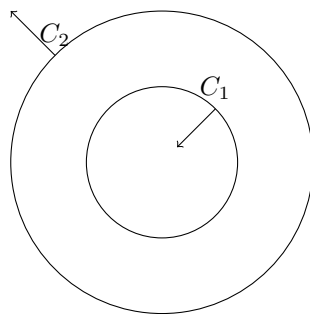
This is not a contradiction of the above theorem. It's tempting to say by (*), $\iint_{\text{disk}} \nabla \cdot \underline{v} \, dA$ is 0, but in fact this integral doesn't exist at all since the field is not bounded.

What's the flux across C_2 the circle of radius 2? Clearly $\underline{n}_+(x, y) = \frac{1}{2}(x, y)$, so the flux is

$$\int_{C_2} \underline{v} \cdot \underline{n}_+ \, ds = \int_{C_2} \frac{1}{4}(x, y) \cdot \frac{1}{2}(x, y) \, ds = \int_{C_2} \frac{1}{2} \, ds = \text{half of length of } C_2 = 2\pi.$$

Let's apply divergence theorem to the annulus $\Omega = \{(x, y) : 1 \leq x^2 + y^2 \leq 2\}$ where \underline{v} is continuously differentiable to see why the two fluxes we calculated are the same:

$$\begin{aligned} \iint_{\Omega} \nabla \cdot \underline{v} \, dA &= \int_{\partial\Omega} \underline{v} \cdot \underline{n}_+ \, ds \\ &= \int_{C_2} \underline{v} \cdot \underline{n}_+ \, ds + \int_{C_1} \underline{v} \cdot \overbrace{\underline{n}_+}^{\text{opposite to the previous one}} \, ds \\ &= 0. \end{aligned}$$



6.5 Flux and divergence in \mathbb{R}^3

We need notion of element of area A of a parameterised regular surface (we want a tangent plane, or more simply its normal vector everywhere).

Definition 6.5.1. Consider $r : \mathcal{O} \rightarrow \mathbb{R}^3$ where \mathcal{O} is an open subset. $r(u, v) = (x(u, v), y(u, v), z(u, v))$ is a *regular* parameterisation of surface $S = r(\mathcal{O})$ if r is continuously differentiable and

$$\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}(u, v) \neq 0 \quad \forall (u, v) \in \mathcal{O}.$$

i.e. $\frac{\partial r}{\partial u}, \frac{\partial r}{\partial v}$ are linearly independent and tangent plane $T_{r(u, v)}S$ of S at $r(u, v)$ is spanned by $\frac{\partial r}{\partial u}, \frac{\partial r}{\partial v}$.

Definition 6.5.2. *Element of area* $dA_{u, v}$ on S is

$$\left| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}(u, v) \right| du dv,$$

which is independent of parameterisation by change of variables.

The intuition behind this the area can be written as

$$\Delta S(u, v) = r(\Delta R_{u, v})$$

where area of $\Delta R_{u, v}$ is just $\Delta u \times \Delta v$. Then area of $\Delta S(u, v)$ is approximately

$$\left| \frac{\partial r}{\partial u} \Delta u \times \frac{\partial r}{\partial v} \Delta v \right| = \left| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right| \Delta u \Delta v.$$

We can't simply use polyhedral to approximate surfaces similar to what we've done for lines: see Schwarz lantern.

Now given continuous $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, we can define surface integral

$$\iint_S f dA := \iint_{\mathcal{O}} f(r(u, v)) \left| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}(u, v) \right| du dv.$$

And we can do flux.

Definition 6.5.3. Given vector field $\underline{v}(x, y, z)$ across surface S , *flux* is defined by

$$\iint_S \underline{v} \cdot \underline{n} \, dA = \iint_{\mathcal{O}} \underline{v}(r(u, v)) \cdot \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \, du \, dv$$

where \underline{n} is normal to S . The left hand side notation emphasises independence of parameterisation, and since we can write

$$\underline{n} = \frac{\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}}{\left| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right|}, \quad dA = \left| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right| \, du \, dv,$$

cancellation of these leads to the right hand side computational formula and proves independence of parameterisation.

How to interpret this? Think of \underline{v} as velocity with which piece of surface $\Delta S_{u,v}$ moves. We look at how a small area, approximated by a parallelogram $\Delta\sigma(u, v)$, moves in time to $\Delta\tilde{\sigma}(u, v)$. It sweeps out a volume enclosed by the two surfaces and the “walls” parallel to $\underline{v}\Delta t$. So amount of volume that flows across S in time Δt is just

$$\left(\iint_S \underline{v} \cdot \underline{n} \, \text{Area}(\Delta\sigma(u, v)) \right) \Delta t$$

and rate is

$$\iint_S \underline{v} \cdot \underline{n} \, \text{Area}(\Delta\sigma(u, v)) = \iint_S \underline{v} \cdot \underline{n} \, dA.$$

Theorem 6.5.4 (Divergence theorem for regions in \mathbb{R}^3). Given solid region $\Omega \subset \mathbb{R}^3$ and open subset $U \subset \mathbb{R}^3$ such that $\bar{\Omega} = \Omega \cup \partial\Omega \subset U$, and $\underline{v} : U \rightarrow \mathbb{R}^3$ is C^1 vector field, then

$$\iiint_{\Omega} \nabla \cdot \underline{v} \, dV = \iint_{\partial\Omega} \underline{v} \cdot \underline{n}_+ \, dA = \iint_{\mathcal{O}} \underline{v}(r(u, v)) \cdot \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \, du \, dv.$$

where $\nabla \cdot \underline{v}$ is similarly defined to Definition 6.4.1, \underline{n}_+ is outward unit normal to Ω and r is a positively oriented parameterisation of $\partial\Omega$ in sense that $\frac{\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}}{\left| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right|} = \underline{n}_+$.

See example in notes and example sheets.

6.6 Gradient fields

Definition 6.6.1. $\underline{v} : U \rightarrow \mathbb{R}^n$ is a *gradient field* if $\exists f \in C^1(U)$:

$$\underline{v}(x) = \nabla f(x) \quad \forall x \in U,$$

so in components

$$v_1(x) = \frac{\partial f}{\partial x_1}(x), \dots, v_n(x) = \frac{\partial f}{\partial x_n}(x),$$

where $\underline{v}(x) = (v_1(x), \dots, v_n(x))$.

Theorem 6.6.2 (FTC for gradient fields). Given $f \in C^1(U)$, curve $C_{pq} \subset U$ parameterised by a continuously differentiable $r : [a, b] \rightarrow U : r(a) = p, r(b) = q$. Then

$$\int_{C_{pq}} \nabla f \cdot dr = f(q) - f(p).$$

Remark. The usual FTC is

$$\int_p^q f'(t) \, dt = f(q) - f(p).$$

Proof.

$$\begin{aligned} \int_{C_{pq}} \nabla f \cdot dr &= \int_a^b \nabla f(r(u)) \cdot \frac{dr}{du} \, du \\ &= \int_a^b \frac{d}{du} f(r(u)) \, du && \text{chain rule} \\ &= f(r(b)) - f(r(a)) && \text{FTC} \\ &= f(q) - f(p). \end{aligned}$$

□

Corollary 6.6.3 (Independence of path; stronger than that of parameterisation). Given 2 arbitrary different paths C_{pq} and \tilde{C}_{pq} between p, q , we have

$$\int_{C_{pq}} \nabla f \cdot dr = \int_{\tilde{C}_{pq}} \nabla f \cdot dr.$$

Corollary 6.6.4.

$$\oint_C \nabla f \cdot dr = 0$$

for all closed curves $C \subset U$ (does not have to be embedded).

Take any point on the curve as both p and q and we can see this.

6.7 Conservative fields

Definition 6.7.1. $\underline{v} : U \rightarrow \mathbb{R}^n$ is *conservative* if

$$\oint_C \nabla f \cdot dr = 0$$

for all closed curves $C \subset U$.

So all gradient fields are conservative. What about the converse?

Proposition 6.7.2. \underline{v} is conservative if and only if $\int_{C_{pq}} \underline{v} \cdot dr$ is independent of choice of path in U from p to q .

Proof. \Rightarrow : We need to show

$$\int_{C_{pq}} \nabla f \cdot dr = \int_{\tilde{C}_{pq}} \nabla f \cdot dr$$

for two arbitrary paths. Denote \tilde{C}_{pq} with opposite orientation as $-\tilde{C}_{pq}$. Then $C_{pq} \cup -\tilde{C}_{pq} = C$ is closed (go from p to q along C_{pq} and from q back to p along opposite direction of \tilde{C}_{pq}). So

$$\oint_C \nabla f \cdot dr = \int_{C_{pq}} \underline{v} \cdot dr + \int_{-\tilde{C}_{pq}} \underline{v} \cdot dr = 0.$$

Hence

$$\int_{C_{pq}} \underline{v} \cdot dr = - \int_{-\tilde{C}_{pq}} \underline{v} \cdot dr = \int_{\tilde{C}_{pq}} \underline{v} \cdot dr$$

\Leftarrow : Let C be a closed curve and $p \in C$. What are the paths from p to p ? Along the closed curve, and just don't go anywhere and stay at p .

$$\begin{aligned}\oint_C \underline{v} \cdot d\mathbf{r} &= \int_{C_{pp}} \underline{v} \cdot d\mathbf{r} \\ &= \int_{C_p^*} \underline{v} \cdot d\mathbf{r} \quad \text{by independence of path} \\ &= \int_0^1 \underline{v}(r(u)) \cdot \underbrace{\frac{d\mathbf{r}}{du}}_{=0} du \\ &= 0,\end{aligned}$$

where C_p^* is the constant path: $r(u) = p \ \forall u \in [0, 1]$.

□

6.8 Scalar potential of conservative fields

Given $\underline{v}(x) = (v_1(x), \dots, v_n(x))$, $x = (x_1, \dots, x_n)$, is there $f : U \rightarrow \mathbb{R} : \nabla f(x) = \underline{v}(x)$? i.e. Can we solve the following n -partial differential equations?

$$\frac{\partial f}{\partial x_1}(x) = v_1(x), \dots, \frac{\partial f}{\partial x_n}(x) = v_n(x).$$

We have n equations for 1 unknown function. Too many equations! If $n = 1$, we just have $\frac{df}{dx} = v(x)$ so $f(x) = \int^x v(u) du$.

Theorem 6.8.1. A continuous vector field $\underline{v} : U \rightarrow \mathbb{R}^n$ is a gradient field if and only if \underline{v} is conservative.

Remark. You would say this is not very helpful. Checking if a vector field is conservative is painful: you need to check infinite number of closed curves.

Sketch of proof. Gradients fields are conservative by definition of latter. So we suppose \underline{v} to be conservative and find a way to construct f in Definition 6.6.1. The $n = 1$ case gives us a hint: we do this by integration, in particular line integral along arbitrary paths from p to q in U .

Fix $q \in U$. Assume U is path connected. So given $p \in U$, \exists a (non unique) path $C_{qp} \subset U$. Then define

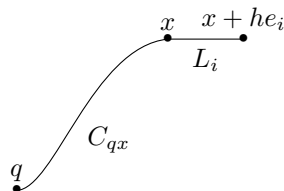
$$f(p) := \int_{C_{qp}} \underline{v} \cdot d\mathbf{r}.$$

\underline{v} being conservative ensures that f is well-defined, in the sense that it does not depend on choice of C . Now we need to check $\nabla f(x) = \underline{v}(x)$, i.e. $\frac{\partial f}{\partial x_i}(x) = v_i(x)$ for all $i = 1, \dots, n$. Since

$$\begin{aligned}f(x + he_i) &= \int_{C_{qx}} \underline{v} \cdot d\mathbf{r} + \int_{L_i} \underline{v} \cdot d\mathbf{r} \\ &= f(x) + \int_0^h \underline{v}(x + te_i) \cdot e_i dt \\ &= f(x) + \int_0^h v_i(x + te_i) dt,\end{aligned}$$

we have

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h \rightarrow 0} \frac{f(x + he_i) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{\int_0^h v_i(x + te_i) dt}{h} = v_i(x)$$



□

Example 6.8.2 (Explicit calculation of scalar potential of a conservative vector field on \mathbb{R}^n). $\underline{v}(x, y, z) = (xye^{xy} + e^{xy} + 1, x^2e^{xy}, 2z)$. We want to find $f : \underline{v} = (f_x, f_y, f_z)$, i.e. solve 3 equations. Starting with the easiest one,

$$\frac{\partial f}{\partial z} = 2z \Rightarrow f(x, y, z) = z^2 \underbrace{+ h(x, y)}_{\text{the "constant"}},$$

then

$$\frac{\partial}{\partial y} (z^2 + h(x, y)) = \frac{\partial h}{\partial y} = x^2 e^{xy} \Rightarrow h(x, y) = x e^{xy} + g(x),$$

finally

$$\begin{aligned} \frac{\partial}{\partial x} (z^2 + x e^{xy} + g(x)) &= e^{xy} + x y e^{xy} + \frac{\partial g}{\partial x} = x y e^{xy} + e^{xy} + 1 \\ &\Rightarrow g = x + C. \end{aligned}$$

We conclude that

$$f(x, y, z) = z^2 + x e^{xy} + x + C.$$

This process exploits the niceness of \mathbb{R}^n . We can of course use the recipe constructed in the proof above to find f , but would run into nasty integrals.

Example 6.8.3 (What goes wrong if \underline{v} is not conservative?). $\underline{v}(x, y) = (0, x)$. We need

$$\frac{\partial f}{\partial x} = 0 \Rightarrow f(x, y) = g(y)$$

and

$$\frac{\partial f}{\partial y} = \frac{dg}{dy} = x$$

which is impossible.

6.8.1 Relation of curl and existence of scalar potential for C^1 vector fields

Suppose $f \in C^2(\Omega)$, $\Omega \in \mathbb{R}^2$. Then $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ and this can be written as

$$\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) - \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = 0,$$

i.e. $\text{curl } \nabla f = 0$. (So if $\text{curl } \underline{v} \neq 0$ then \underline{v} is not conservative, a useful indicator.) Converse is true, only locally, i.e. if $\underline{v} \in C^1(U)$ and $\text{curl } \underline{v} = 0$ on U , then

$$\forall x \in U, \exists p > 0 : \mathbb{B}(x, p) \subset U \text{ and } \exists f \in C^2(\mathbb{B}(x, p)) : \nabla f(y) = \underline{v}(y) \forall y \in \mathbb{B}(x, p).$$

However, there may not exist a global scalar potential $f \in C^2(U)$ such that $\underline{v}(y) = \nabla f(y) \forall y \in U$ (see Q2 in assignment 4) unless the first homology group $H_1(\Omega, \mathbb{R}) = \{0\}$ is trivial. This is something that measures 1-d holes. Ω being simply connected, i.e. $\pi_1(\Omega) = 1$, is sufficient for this.

6.9 Incompressible and conservative vector fields

Consider fluid with density $\rho(x, t)$ and velocity $v(x, t)$ at position x at time t . They satisfy the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \underline{v}) = 0,$$

which can be viewed as infinitesimal conservation of mass. Assume fluid is incompressible (e.g. water at low pressure and constant temperature). Then ρ is constant and by above equation \underline{v} satisfies $\nabla \cdot \underline{v} = 0$.

Definition 6.9.1. Using synecdoche we call such \underline{v} *incompressible* fields, also *solenoidal* fields in context of electromagnetism (also *divergence-free* in context of, well, mathematics).

Suppose \underline{v} conservative, i.e. $\underline{v} = \nabla f$. Then if $\nabla \cdot \underline{v} = 0$ then $\nabla \cdot (\nabla f) = 0$, i.e.

$$\left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right) \cdot \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) = 0,$$

so

$$\frac{\partial^2 f}{\partial x_1^2} + \dots + \frac{\partial^2 f}{\partial x_n^2} =: \Delta f = 0$$

Definition 6.9.2. Δ is called the Laplace operator or *Laplacian*, and such f are called *harmonic*.

Example 6.9.3 (of harmonic functions). • linear functions $f(x) = a \cdot x$ where $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. We have $\nabla f(x) = a$.

- real/imaginary parts of \mathbb{C} -analytic functions, e.g. $f(z) = z^2 = x^2 - y^2 + 2ixy$. Then $x^2 - y^2$ and $2xy$ are both harmonic: consequence of Cauchy–Riemann

We seek radial harmonic functions on $\mathbb{R}^2 \setminus \{0\}$.

Definition 6.9.4. $f : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$ is *radial/spherically symmetric* if $\exists \varphi : \mathbb{R}_{>0} \rightarrow \mathbb{R} : f(x) = \varphi(|x|)$. (i.e. f is constant on circles)

If f is radial, then

$$\nabla f(x) = \varphi'(|x|) \nabla(|x|) = \varphi'(|x|) \underbrace{\frac{x}{|x|}}_{\text{unit radial vector}}$$

Proposition 6.9.5 (Zero flux property of divergence-free vector fields). If $\underline{v} \in C^1(U, \mathbb{R}^3)$ is divergence-free, then the flux

$$\iint_{\partial \Omega} \underline{v} \cdot \underline{n}_+ \, dA = 0$$

for all region $\Omega : \overline{\Omega} \subset U$.

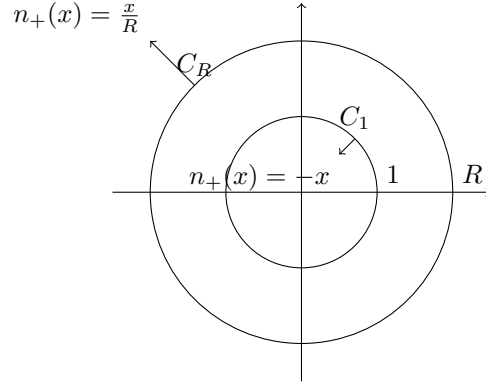
Proof. Immediately from divergence theorem. \square

Now notice that f harmonic is the same as ∇f divergence-free. So we can apply zero flux property with

$$\Omega = A_R = \mathbb{B}_R \setminus \overline{\mathbb{B}_1} = \{x \in \mathbb{R}^2 : 1 < |x| < R\}$$

and

$$\partial A_R = C_R \cup C_1$$



by writing

$$\begin{aligned} 0 &= \int_{\partial A_R} \nabla f \cdot n \, ds = \int_{C_R} \varphi'(R) \frac{x}{R} \cdot \frac{x}{R} \, ds + \int_{C_1} \varphi'(1) x \cdot (-x) \, ds \\ &= \varphi'(R) \int_{C_R} ds - \varphi'(1) \int_{C_1} ds \\ &= \varphi'(R) 2\pi R - \varphi'(1) 2\pi \end{aligned}$$

So φ satisfies the ODE

$$\varphi'(R) = \frac{\varphi'(1)}{R}$$

i.e. $\varphi(R) = \varphi'(1) \log R + C$. So $f(x) = a \log(|x|) + b$.

7 Inverse function theorem

7.1 1-d review

Assume

- $f : (a, b) \rightarrow \mathbb{R}$ is differentiable on (a, b) and $f'(x) > 0 \, \forall x \in (a, b)$

Then

1. f is increasing (MVT)
2. $f : (a, b) \rightarrow (\alpha, \beta)$ is a bijection where $\alpha := \lim_{x \downarrow a+} f(x)$ and $\beta := \lim_{x \uparrow b-} f(x)$ exist because of
 - (i). Surjection from IVT

3. $f^{-1} : (\alpha, \beta) \rightarrow (a, b)$ is differentiable and

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

Alternative formulation of 2. and 3.: set $y = f(x)$ and view this as an equation for x in terms of y , then

2. $y = f(x)$ can be solved for x in terms of y , i.e. $x = f^{-1}(y)$

3.

$$\frac{dx}{dy}(y) = \frac{1}{\frac{dy}{dx}(f^{-1}(y))}$$

Example 7.1.1. $f(x) = e^x = y$. Solve for x and we have $x = \log y = f^{-1}(y)$, and

$$\frac{dx}{dy} = \frac{1}{y} = \frac{1}{e^{\log y}} = \frac{1}{\frac{dy}{dx}(\log y)}.$$

We want to extend these results to higher dimensions. We start with linear cases which are the simplest. A linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a bijection if and only if $\det T \neq 0$. Then we can solve $Tx = y$ for $x : x = T^{-1}y$. Also $DT = T$ since... well if approximate a linear function linearly you get the function itself. This suggests that a continuously differentiable map $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ may be a bijection if $\det \partial\psi(x) \neq 0 \forall x \in \mathbb{R}^n$. Some evidence this may be true:

Recall assignment 2 Q2. We were given

- $f \in C^1(\mathbb{B}, \mathbb{R}^k)$ where \mathbb{B} is the unit ball in \mathbb{R}^n
- $Df(0)$ is injective, i.e. (Example sheet 3 Q5) $\exists \alpha > 0 : |Df(0)h| \geq \alpha|h| \forall h \in \mathbb{R}^n$.

We conclude $\exists \delta > 0 : f$ is injective on \mathbb{B}_δ . Quantitatively $|f(x) - f(y)| \geq \frac{\alpha}{2}|x - y| \forall x, y \in \mathbb{B}_\delta$.

Example 7.1.2. We view the complex function $f(z) = e^z$ as $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $\psi(x, y) = (e^x \cos y, e^x \sin y)$. Then its Jacobian is

$$\partial\psi(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

and $\det \partial\psi(x, y) = e^{2x} > 0 \forall (x, y) \in \mathbb{R}^2$. But ψ is not injective; in fact, it's periodic.

Seems that we arrive at the wrong conclusion. The insight that the above question recalled gives us is that maybe we don't have a global inverse from $\det \partial \neq 0$ yet: we just have a *local inverse*.

Definition 7.1.3. A *neighbourhood* \mathcal{N}_p of p is any open set that contains p .

Definition 7.1.4. $\psi : U \rightarrow V$ is a *local bijection* at $p \in U$ if p has a neighbourhood $\mathcal{N}_p \subset U$ and $q := \psi(p)$ has a neighbourhood $\mathcal{N}_q \subset V : \psi : \mathcal{N}_p \rightarrow \mathcal{N}_q$ is a bijection.

Theorem 7.1.5 (Inverse function). Given

- $\psi \in C^1(U, \mathbb{R}^n)$, $U \subset \mathbb{R}^n$ open
- $D\psi(p)$ invertible at $p \in U$, i.e. $\det \partial\psi(p) \neq 0$

and set $q = \psi(p)$, we have

- \exists a neighbourhood \mathcal{N}_p and $\mathcal{N}_q : p \in \mathcal{N}_p, q \in \mathcal{N}_q \subset \psi(U)$ and $\psi : \mathcal{N}_p \rightarrow \mathcal{N}_q$ is a bijection
- $\psi^{-1} : \mathcal{N}_q \rightarrow \mathcal{N}_p$ is continuously differentiable and

$$D\psi^{-1}(y) = (D\psi(\psi^{-1}(y)))^{-1} \quad \forall y \in \mathcal{N}_q. \quad (*)$$

Informally this means $\psi(x) = y$ can be solved for x and $x = \psi^{-1}(y) \quad \forall y \in \mathcal{N}_q$.

*Proof of *.* We already have ψ^{-1} exists and is continuously differentiable. So differentiate both sides of

$$\psi(\psi^{-1}(y)) = y \quad \forall y \in \mathcal{N}_q$$

to get

$$\underbrace{D\psi(\psi^{-1}(y)) \cdot D\psi^{-1}(y)}_{\text{chain rule}} = I_n$$

which gives what's desired immediately. \square

Example 7.1.6. $\psi(x, y) = (e^x + e^y, y) = (u(x, y), v(x, y))$. To compute ψ^{-1} , we solve $e^x + e^y = u, y = v$ for x, y in terms of u, v : $y(u, v) = v$ and $x(u, v) = \log(u - e^v)$, so

$$\psi^{-1}(u, v) = (\log(u - e^v), v) = (x(u, v), y(u, v)).$$

This is case where the inverse is explicitly calculable. What the inverse function theorem tells us that even if that's not the case, nevertheless a solution exists.

Example 7.1.7. $U = \{(x, y) : |x|, |y| < \frac{\pi}{2}\}$, and $\psi : U \rightarrow \mathbb{R}^2$ is defined by $\psi(x, y) = (xy, \cos x + \cos y)$.

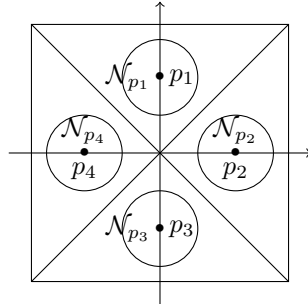
Inverse function theorem tells us to calculate Jacobian first, so:

$$\partial\psi(x, y) = \begin{pmatrix} y & x \\ -\sin x & -\sin y \end{pmatrix}$$

and

$$\det \partial\psi(x, y) = x \sin x - y \sin y = 0 \text{ when } x = y \text{ or } x = -y.$$

In fact $\psi(x, y) = \psi(-x, -y) = \psi(y, x) = \psi(-y, -x)$. Take $p_1 = (0, \frac{\pi}{3})$, and if we set symmetries $p_2 = (\frac{\pi}{3}, 0)$, $p_3 = (0, -\frac{\pi}{3})$, $p_4 = (-\frac{\pi}{3}, 0)$ then $\psi(p_1) = \psi(p_2) = \psi(p_3) = \psi(p_4)$, so ψ is clearly not injective.



But what inverse function theorem tells us that there are local neighbourhoods \mathcal{N}_{p_i} where bijections exist. What we get is 4 branches of inverses of the same function on 4 different domains (like positive and negative square roots as 2 inverses of $x \mapsto x^2$). We can't explicitly write them but we know their derivatives:

$$\partial\psi_1^{-1}\left(0, \frac{3}{2}\right) = \left(\partial\psi\left(0, \frac{\pi}{3}\right)\right)^{-1} = \begin{pmatrix} \frac{\pi}{3} & 0 \\ 0 & -\frac{\sqrt{3}}{2} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{3}{\pi} & 0 \\ 0 & -\frac{2}{\sqrt{3}} \end{pmatrix}$$

and

$$\partial\psi_2^{-1}\left(0, \frac{3}{2}\right) = \left(\partial\psi\left(\frac{\pi}{3}, 0\right)\right)^{-1} = \begin{pmatrix} 0 & \frac{\pi}{3} \\ -\frac{\sqrt{3}}{2} & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & -\frac{2}{\sqrt{3}} \\ \frac{3}{\pi} & 0 \end{pmatrix}$$

and so on.

Definition 7.1.8. $U, V \subset \mathbb{R}^n$ open. $\psi : U \rightarrow V$ is a *diffeomorphism* if

- ψ is a bijection
- ψ and ψ^{-1} are both continuously differentiable.

ψ is a *local* diffeomorphism near $p \in U$ if \exists neighbourhood $\mathcal{N}_p \subset U$ of p and $\mathcal{N}_q \subset V$ of $q = \psi(p) : \psi : \mathcal{N}_p \rightarrow \mathcal{N}_q$ is a diffeomorphism.

So we can reformulate inverse function theorem: if $\psi \in C^1(U, \mathbb{R}^n)$ and $\det \partial\psi(p) \neq 0$ then ψ is a local diffeomorphism near p .

Now...

Strategy of proof of Theorem 7.1.5. 1. $D\psi(p)$ is invertible and continuous at p implies that $\exists \delta > 0 : D\psi(x)$ is injective $\forall x \in \mathbb{B}(p, \delta)$ and $\psi : \mathbb{B}(p, \delta) \rightarrow \mathbb{R}^n$ is injective. We've seen this in assignment 3 Q2. In fact, $\exists \alpha > 0 : |\psi(x) - \psi(z)| \geq \frac{\alpha}{2}|x - z| \forall x, z \in \mathbb{B}(p, \delta)$.

2. Note that this implies $|\psi(x) - \psi(p)| \geq \frac{\alpha}{4}\delta \forall x \in \partial\mathbb{B}(p, \frac{1}{2}\delta)$. Then it's hard. Read appendix of typewritten notes if interested.

□

8 Implicit function theorem

Scenario: The relation $x^2 + y^2 = c > 0$ defines y *implicitly* in terms of x and defines x implicitly in terms of y . The 2 possible solutions, $y_+ = \sqrt{c - x^2}$, $y_- = -\sqrt{c - x^2}$ are called *explicit* determinations of y in terms of x .

General set-up:

- $\mathbb{R}^{n+l} = \mathbb{R}^n \oplus \mathbb{R}^l = \{(x, y) : x \in \mathbb{R}^n, y \in \mathbb{R}^l\}$
- $U \subset \mathbb{R}^{n+l}$ open
- $F : U \rightarrow \mathbb{R}^l$ continuously differentiable
- $F(x, y) = (F_1(x_1, \dots, x_n, y_1, \dots, y_l), \dots, F_l(x_1, \dots, x_n, y_1, \dots, y_l))$

$$\bullet \partial F(x, y) = \begin{pmatrix} \partial_{x_1} F_1 & \cdots & \partial_{x_n} F_1 & \partial_{y_1} F_1 & \cdots & \partial_{y_l} F_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \partial_{x_1} F_l & \cdots & \partial_{x_n} F_l & \partial_{y_1} F_l & \cdots & \partial_{y_l} F_l \end{pmatrix} = \begin{pmatrix} \partial_x F & \partial_y F \end{pmatrix} \text{ where } \partial_y F \text{ is a square matrix}$$

Fix $c \in \mathbb{R}^l$, $c = (c_1, \dots, c_l)$, then the l -equations $F(x, y) = c$, i.e.

$$\begin{aligned} F_1(x_1, \dots, x_n, y_1, \dots, y_l) &= c_1 \\ &\vdots \\ F_l(x_1, \dots, x_n, y_1, \dots, y_l) &= c_l \end{aligned}$$

define y *implicitly* in terms of x . A solution of these equations on a neighbourhood \mathcal{N}_{x_0} of x_0 is an *explicit* function $g : \mathcal{N}_{x_0} \rightarrow \mathbb{R}^l : F(x, g(x)) = c \ \forall x \text{ in } \mathcal{N}_{x_0}$.

In the example scenario, $n = l = 1$, $F(x, y) = x^2 + y^2$ and $g(x) = \sqrt{c - x^2}$ or $-\sqrt{c - x^2}$, $x_0 = 0$, $\mathcal{N}_{x_0} = (-\sqrt{c}, \sqrt{c})$.

Intuition: given l equations in $n + l$ unknowns, we can attempt to eliminate the l -variables y_1, \dots, y_l by writing them as functions of the remaining l -variables x_1, \dots, x_n : $y_1 = g_1(x_1, \dots, x_n), \dots, y_l = g_l(x_1, \dots, x_n)$ which are explicit.

Simplest (linear) case: $F(x, y) = Ax + By$ where $A \in \mathbb{R}^{l \times n}$, $B \in \mathbb{R}^{l \times l}$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} & b_{11} & \cdots & b_{1l} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{l1} & \cdots & a_{ln} & b_{l1} & \cdots & b_{ll} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_l \end{pmatrix}$$

If we attempt to solve for C , matrix B had better be invertible. This again gives us insight into how the theorem would look like: we require $\partial_y F$ to be invertible.

Theorem 8.0.1 (Implicit function theorem). Given

- $U \subset \mathbb{R}^{n+l} = \mathbb{R}^n \oplus \mathbb{R}^l$, U open
- $c \in \mathbb{R}^l$

Assume

- $F \in C^1(U, \mathbb{R}^l)$ and $\exists (x_0, y_0) \in U : F(x_0, y_0) = c$
- $\det \partial_y F(x_0, y_0) \neq 0$

Conclusion

- \exists neighbourhood \mathcal{N}_{x_0} of x_0 in \mathbb{R}^n and $g \in C^1(\mathcal{N}_{x_0}, \mathbb{R}^l) : g(x_0) = y_0$ and $F(x, g(x)) = c \ \forall x \in \mathcal{N}_{x_0}$, i.e. $y = g(x)$ solves $F(x, y) = c$.
- $\partial_y F(x, g(x))$ is invertible $\forall x \in \mathcal{N}_{x_0}$ (not surprising by continuity) and

$$\partial g(x) = - \underbrace{(\partial_y F(x, g(x)))^{-1}}_{l \times l} \underbrace{(\partial_x F(x, g(x)))}_{l \times n}. \quad (*)$$

*Proof of *.* Differentiating $F(x, g(x)) = c$ with respect to x using chain rule, we have

$$\partial_x F(x, g(x)) + \partial_y F(x, g(x)) \cdot \partial g(x) = 0, \quad (**)$$

which immediately leads to what's desired. \square

Back to example $F(x, y) = x^2 + y^2 = 4$. We have $\partial_y F(x, y) = 2y$. Pick (x_0, y_0) so that $\partial_y F(x_0, y_0) \neq 0$, e.g. $y_0 = 1, x_0 = \sqrt{3}$. Then $g(x) = y = \sqrt{4 - x^2}$. Explicit calculation gives $g'(x) = -\frac{x}{\sqrt{4-x^2}}$. Above theorem gives $g'(x) = -\frac{2x}{2y} = -\frac{x}{y}$ which agrees.

Now if we look back at high school implicit differentiation: $x^2 + y^2 = 4$ with respect to x gives $2x + 2y \frac{dy}{dx} = 0$. That's just $(**)$.

Had we picked $(x_0, y_0) = (\sqrt{3}, -1)$, then $g(x) = -\sqrt{4 - x^2}$ and $g'(x)$ is still $-\frac{x}{y}$ but now is $\frac{x}{\sqrt{4-x^2}}$.

8.1 Maximal rank reformulation of Implicit function theorem

If $P \in \mathbb{R}^{l \times (n+l)}$ then $\text{rank } P \leq l$ and the equal sign occurs if and only if after possibly reordering columns of P we can write $P = (A \ B)$ where $A \in \mathbb{R}^{l \times n}$ and B is a square matrix $\in \mathbb{R}^{l \times l}$ with $\det B \neq 0$. (Just put l linearly independent columns of P at the end.)

We had $F : U \rightarrow \mathbb{R}^l$, $U \subset \mathbb{R}^{n+l}$ open and

$$\partial F = (\partial_x F \ \partial_y F). \quad (*)$$

So the condition $\det \partial_y F \neq 0$ can be rephrased as $\partial F(x, y)$ has maximal rank. To achieve form $(*)$ we may have to relabel coordinates $x_1, \dots, x_n, y_1, \dots, y_l$.

8.2 Using Implicit function theorem to parameterise (pieces of) level sets

Definition 8.2.1. Given $F \in C^1(U, \mathbb{R}^l)$, $U \subset \mathbb{R}^{n+l}$ open and $c \in \mathbb{R}^l$, the *level set* Γ_c of F is defined by

$$\Gamma_c = \{z \in \mathbb{R}^{n+l} : F(z) = c\}.$$

Previously $z = (x, y)$.

We know $r \in C^1((a, b), \mathbb{R}^k)$ is regular if $r'(u) \neq 0 \ \forall u \in (a, b)$. The image $r((a, b))$ is an open subset of a regular curve $C \subset \mathbb{R}^k$, called a 1-dimensional manifold.

Similarly, $U \subset \mathbb{R}^2$, $r \in C^1(U, \mathbb{R}^k)$ is *regular* if $\partial r(u, v) = (\frac{\partial r}{\partial u} \ \frac{\partial r}{\partial v})$ has maximal rank 2. The image $r(U)$ is an open subset of a regular surface $S \subset \mathbb{R}^k$, called a 2-dimensional manifold.

If we replace the 2 above by n , we have the definition of n -dimensional manifold.

The tangent space $T_{r(x)}M$ of M at $r(x)$ is spanned by $\partial_1 r(x), \dots, \partial_n r(x)$, i.e. columns of Jacobian $\partial r(x)$.

Back to level sets. If $\partial F(z)$ has maximal rank $l \ \forall z \in \Gamma_c$, then by Implicit function theorem, if $z_0 = (x_0, y_0) \in \Gamma_c$, then $\exists \mathcal{N}_{x_0}$ in \mathbb{R}^n and $g \in C^1(\mathcal{N}_{x_0}, \mathbb{R}^l) : (x, g(x)) \in \Gamma_c \ \forall x \in \mathcal{N}_{x_0}$ because $F(x, g(x)) = c \ \forall x \in \mathcal{N}_{x_0}$ and definition of level sets. Set $r(x) = (x, g(x)) \ \forall x \in \mathcal{N}_{x_0}$, then r is a regular parameterisation of its image $r(\mathcal{N}_{x_0}) \subset \Gamma_c$. Indeed:

$$\partial r(x) = \begin{pmatrix} I_n \\ \partial g(x) \end{pmatrix}$$

with clearly maximal rank n . What is $T_{r(x)}\Gamma_c$? For this, differentiate $F(x, g(x)) = c$ with respect to x :

$$0 = \partial_x F + \partial_y F \cdot \partial g = \begin{pmatrix} \partial_x F & \partial_y F \end{pmatrix} \begin{pmatrix} I_n \\ \partial g \end{pmatrix} = \partial F(x, g(x)) \cdot \partial r(x),$$

which means linearly independent vectors $\partial_1 r, \dots, \partial_n r \in \text{Ker } \partial F(r(x))$, i.e. Nullity $\partial F(r(x)) \geq n$. But we assume $\partial F(r(x))$ to be precisely l in order to apply Implicit function theorem, therefore

$$\text{Ker } \partial F(r(x)) = \text{span}\{\partial_1 r, \dots, \partial_n r\} = T_{r(x)}\Gamma_c.$$

Example 8.2.2. $l = 1$, $F : U \rightarrow \mathbb{R}$, $U \subset \mathbb{R}^{n+1}$, rank $\partial F(z)$ is maximal (1) $\forall z \in \Gamma_c$. This means $\nabla F(z)$, the transpose of $\partial F(z)$, is nonzero $\forall z \in \Gamma_c$. Moreover, $c \in \text{Ker } \partial F(z)$ if and only if

$$0 = \begin{pmatrix} \partial_1 & \cdots & \partial_{n+1} F \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_{n+1} \end{pmatrix} = \nabla F(z) \cdot v$$

i.e. $v \perp \nabla F(z)$. So Implicit function theorem says that near any point $z_0 \in \Gamma_0$ can be parameterised as a graph and $T_{z_0}\Gamma_0 = \text{orthogonal complement of } \nabla F(z_0) = \text{Ker } \partial F(z_0)$.

Example 8.2.3. $\Delta : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ defined by $\Delta(A) = \det A$. The special linear group $SL(n, \mathbb{R}) = \{A \in \mathbb{R}^{n \times n} : \det A = 1\}$ is precisely level set Γ_1 . In Q6 of Example sheet 4, we showed

$$(\partial_H \Delta)(A) = \text{trace } (A^{-1}H)$$

if $\det A \neq 0$. What is rank $\partial \Delta(A)$? We write

$$(\partial_A \Delta)A = \text{trace } (A^{-1}A) = \text{trace } I_n = n \neq 0,$$

which means rank $\partial \Delta(A) \geq 1$. But $\partial \Delta(A) \leq 1$, so $\partial \Delta(A) = 1$. Then by Implicit function theorem,

$$T_A SL(n, \mathbb{R}) = \{H : \text{trace } (A^{-1}H) = 0\}$$

and in particular

$$T_{I_n} SL(n, \mathbb{R}) = \{H : \text{trace } H = 0\},$$

a beautiful interpretation of traceless matrices.