Modern Analysis II

Victor Ortega
Columbia University

October 28, 2025

Contents

Chapter 1

Power Series

1.1 Preamble

These are notes from the Fall 2025 Intro to Modern Analysis II class from Dr. Jeanne Boursier at Columbia University. The textbook for this course was Analysis II by Terence Tao.

1.2 Series of Functions

Definition 1.1 (Metric Space). Let X be a non-empty set. Let $d : X \times X \to \mathbb{R} : \cup : 0$ be a function. We say that d is a metric or distance on $X \Leftrightarrow d$ satisfies the following properties

(d_1) $\forall\, x, y \in X \Rightarrow d(x, y) \geqslant 0$

(d_2) $d(x, y) = 0 \Leftrightarrow x = y$

(d_3) $\forall\, x, y \in X \Rightarrow d(x, y) = d(y, x)$

(d_4) $\forall\, x, y, z \in X \Rightarrow d(x, z) \leqslant d(x, y) + d(y, z)$

A metric space is an ordered pair (X, d) where X is non-empty and d is a metric on X.

Definition 1.2. Let (X, d) be a metric space, $x \in X$, and $r \in \mathbb{R}^+$. The open ball of center x and radius r is defined as

$$B(x, r) = \{y \in X \mid d(x, y) < r\}$$

The closed ball of center x and radius r is defined as

$$B[x, r] = \{y \in X \mid d(x, y) \leqslant r\}$$

Definition 1.3. Let (X, d) be a metric space and $A \subseteq X$. A point $x \in X$ is called an adherent point of A if for every $\varepsilon > 0 \Rightarrow$ the open ball $B(x, \varepsilon)$ intersects A.

$$B(x, \varepsilon) \ \cap \ A \neq \varnothing$$

Definition 1.4. Let (X, d) and (Y, ρ) be metric spaces $\Rightarrow$ the set of continuous functions of X in Y is defined as

$$\mathcal{C}^0(X, Y) = \{f : X \to Y \mid f \text{ is continuous } \}$$

Notation. In this text, we adopt the following convention for arrows

$\Rightarrow$ is the colloquial word **then** $\Longrightarrow$ is the formal logical **implies**

Definition 1.5 (Limiting Value). Let (X, d) and (Y, ρ) be metric spaces. Let $E \subseteq X$, and let $f : E \to Y$ be a function. If $x_0 \in X$ is an adherent point of E and $L \in Y$ we say

$$\lim_{x \in E \to x_0} f(x) = L$$

and say $f(x)$ converges to L in Y as x converges to x_0 in E if

$$\forall \varepsilon > 0 \, \exists \, \delta > 0 \, \forall \, x \in E \Rightarrow \ 0 < d(x, x_0) < \delta \Longrightarrow \rho(f(x), L) < \varepsilon$$

Intuition. We are working our way to limits of sequences of functions to explore the concept of power series. We will now define two notions of convergence: uniform and pointwise.

Definition 1.6 (Uniform Convergence). Let X be a non-empty set and (Y, ρ) a metric space. We say that a sequence of functions

$$\langle f_n : X \to Y \mid n \in \mathbb{N} \rangle \qquad \text{or} \qquad (f_n : X \to Y)_{n \in \mathbb{N}}$$

converges uniformly to a function $f : X \to Y \Leftrightarrow$

$$\forall \varepsilon > 0 \, \exists \, N \in \mathbb{N} \, \forall \, n \geqslant N \, \forall \, x \in X \Rightarrow \rho(f_n(x), f(x)) < \varepsilon$$

Notation. In this case we write $f_n \rightrightarrows f$, where f is the uniform limit of the sequence.

Definition 1.7 (Pointwise Convergence). Let X be any non-empty set and (Y, ρ) a metric space. We say that a sequence of functions

$$\langle f_n : X \to Y \mid n \in \mathbb{N} \rangle \qquad \text{or} \qquad (f_n : X \to Y)_{n \in \mathbb{N}}$$

converges pointwise to the function $f : X \to Y \Leftrightarrow \forall \, x \in X \Rightarrow$

$$\lim_{n \to \infty} f_n(x) = f(x)$$

in the space (Y, ρ)

Notation. We say $f_n \to f$, where f is the pointwise limit of the sequence $(f_n)_{n \in \mathbb{N}}$

Definition 1.8 (Series). Let (X, d) be a metric space. Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions $f_n : X \to \mathbb{R}$, and let $f : X \to \mathbb{R}$. If the partial sums

$$S_N(x) = \sum_{k=1}^{N} f_k(x)$$

converge pointwise to $f(x)$ as $N \to \infty$, we say that the series

$$\sum_{n=1}^{\infty} f_n(x)$$

converges pointwise on X to f. For converging uniformly it is very similar.

Definition 1.9. Let $\sum f_n$ be a series of functions defined on a set $A \subseteq \mathbb{R}$. It is said to be absolutely convergent if for every $x \in A$, the series $\sum |f_n(x)|$ converges.

Remark. If $\sum f_n$ converges absolutely $\implies \sum f_n$ converges,

Theorem 1.1. Let f_n be differentiable. Suppose $\exists\, x_o$ s.t. $f_n(x_0)$ converges and $f_n' \rightrightarrows g \implies f_n \to f$ differentiable with $f' = g$

Theorem 1.2 (Weierstrass M-test). Let (X,d) be a metric space. Let $(f_n)_{n=1}^{\infty}$ be a sequence of bounded continuous functions $f_n : X \to \mathbb{R}$ such that

$$\sum_{n=1}^{\infty} \|f_n\|_\infty < \infty$$

$\Rightarrow \sum_{n=1}^{\infty} f_n$ converges uniformly to a function $f : X \to \mathbb{R}$, and f is continuous on X

Proof. Fix $x \in X$. Note that

$$\big|f_n(x)\big| \leqslant \sup_{y \in X}$$

Hence $\sum \big|f_n(x)\big|$ converges $\implies \sum f_n(x)$ converges pointwise

$$f(x) = \sum_{n=0}^{\infty} f_n(x) \implies \left| f(x) - \sum_{n=0}^{N} f_n(x) \right|$$

$$= \left| \sum_{n=N+1}^{\infty} f_n(x) \right| \leqslant \sum_{n=N+1}^{\infty} \|f_n\|_\infty$$

Which implies

$$\left\| f - \sum_{n=0}^{\infty} f_n \right\|_\infty \leqslant \sum_{n=N+1}^{\infty} \|f_n\|_\infty \rightrightarrows 0$$

as $N \to \infty$ □

Theorem 1.3 (Root Test). Let $\sum_{n=1}^{\infty} a_n$ be a series of real or complex numbers and set

$$\ell := \limsup_{n\to\infty} \sqrt[n]{|a_n|}$$

If $\ell < 1 \implies \sum_{n=1}^{\infty} a_n$ converges absolutely and converges. If $\ell > 1 \implies \sum_{n=1}^{\infty} a_n$ diverges. If $\ell = 1 \implies$ the series may be divergent, conditionally convergent, or absolutely convergent.

Proof. Suppose $\ell > 1 \Rightarrow \forall\, N\, \exists\, n \geqslant N$ s.t.

$$|a_n| \geqslant \underbrace{\frac{1+\ell}{2}}_{<\ell} \geqslant \left(\frac{1+\ell}{2}\right)^n > 1$$

But $|a_n| \to +\infty \Rightarrow |a_n| \not\to 0$. Thus $\sum_{n=1}^{\infty} a_n$ diverges. Suppose $\ell < 1 \Rightarrow \exists\, N\, \forall\, n \geqslant N$ s.t.

$$|a_n|^{\frac{1}{n}} < \frac{1+\ell}{2} \implies \underbrace{|a_n|}_{\geqslant 0} < \left(\frac{1+\ell}{2}\right)^n \implies \sum_{n=1}^{\infty} \left(\frac{1+\ell}{2}\right)^n \text{ converges}$$

Thus $\sum_{n=1}^{\infty} a_n$ converges. □

Remark. $\limsup$ always exists $\in \mathbb{R}^+ \cup \{+\infty\}$

Theorem 1.4. If $\forall\, n \in \mathbb{N} \implies f_n$ is $\mathcal{C}^0$ and if $(f_n) \rightrightarrows f \implies f$ is $\mathcal{C}^0$

Theorem 1.5. If $\forall\, n \in \mathbb{N} \implies f_n$ is $\mathcal{C}^0$ and if $(f_n) \rightrightarrows f \implies \forall\, [a,b] \subseteq X \implies$

$$\int_a^b f = \lim_n \int_a^b f_n = \int_a^b \lim_n f_n$$

1.3 Power Series

Definition 1.10 (Formal Power Series)**.** Let $a \in \mathbb{R}$ and let $(c_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \Rightarrow$

$$\sum_{n=0}^{\infty} c_n (x-a)^n$$

is called a formal power series centered at a

Remark. We don´t assume Definition 1.10 converges.

Example 1.1. $\sum x^n a^n$ with $a \in \mathbb{R}^+$ converges $\Leftrightarrow |x| < \frac{1}{a}$

Definition 1.11 (Cauchy)**.** (f_m) with $f_m : X \to \mathbb{R}$ is called a Cauchy sequence if

$$\forall\, \varepsilon > 0\, \exists\, N \in \mathbb{N} \text{ s.t. } m, n \geqslant N \Rightarrow |f_m(x) - f_n(x)| < \varepsilon$$

Remark. This notion can be generalized to a metric space (X, d) and a sequence $(x_n)_{n\in\mathbb{N}}$

$$\forall\, n, m\, \exists\, N \in \mathbb{N} \text{ s.t. } \forall\, n, m \geqslant N \implies d(x_n, x_m) < \varepsilon$$

Theorem 1.6. If (X, d) is a metric space, every convergent sequence in (X, d) is Cauchy.

Proof. Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence in X that converges to an element $x_0 \in X$. Let $\varepsilon > 0$ be arbitrary. By convergence, $\exists\, N \in \mathbb{N}$ such that $\forall\, n \geqslant N$ such that $d(x_n, x_0) < \frac{\varepsilon}{2}$. Hence, $\forall\, n, m \geqslant N$ we have

$$d(x_n, x_m) \leqslant d(x_n, x_0) + d(x_0, x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

$\therefore (x_n)_{n\in\mathbb{N}}$ is Cauchy. □

Definition 1.12. We say that a space (X, d) is a complete metric space $\Leftrightarrow$ every Cauchy sequence in (X, d) converges to an element in (X, d).

Definition 1.13 (Radius)**.** Let $\sum_{n=0}^{\infty} c_n (x-a)^n$ satisfy Definition 1.10 $\Rightarrow$

$$R := \frac{1}{\limsup_{n\to\infty} |c_n|^{\frac{1}{n}}}$$

is defined as the radius of convergence of said series

Theorem 1.7. Let $\sum_n^\infty a_n x^n$ be a formal power series with radius of convergence $R \Rightarrow$

$$R = \sup\{\rho \geqslant 0 \mid (a_n\rho^n)_{n\in\mathbb{N}} \text{ is bounded}\}.$$

Proof. Let $r \in \mathbb{R}^+$ and recall Definition 1.13 $\Rightarrow$ taking $\sum_{n=0}^\infty a_n r^n$ means by Theorem 1.10 that $a_n r^n \to 0$ and as such $\sum_{n=0}^\infty a_n r^n$ is bounded $\Rightarrow$

$$\{\rho \geqslant 0 \mid (a_n\rho^n) \text{ is bounded}\} \supseteq \{r \geqslant 0 \mid \sum_{n=0}^{\infty} a_n r^n \text{ converges in } \mathbb{R}\}$$

Conversely, if $(a_n\rho^n)$ is bounded $\exists\, M \in \mathbb{R}^+$ s.t. $\forall\, n \in \mathbb{N} \Rightarrow |a_n\rho^n| \leqslant M$

$$\Rightarrow\ \forall\, r < \rho \Rightarrow |a_n r^n| = |a_n\rho^n| \left(\frac{r}{\rho}\right)^n \leqslant M\left(\frac{r}{\rho}\right)^n$$

By the comparison test $\sum_{n=0}^\infty a_n r^n$ converges in $\mathbb{R} \Rightarrow$

$$\{\rho \geqslant 0 \mid (a_n\rho^n) \text{ is bounded}\} \subseteq \{r \geqslant 0 \mid \sum_{n=0}^{\infty} a_n r^n \text{ converges in } \mathbb{R}\}$$

$\therefore$ Theorem 1.7 is true since we have shown the sets are equal. ☐

Theorem 1.8. Let $\sum_{n\geqslant 0}^\infty c_n(x-a)^n$ with radius of convergence $R \in \mathbb{R}$

(C_1) If $|x-a| < R \implies \sum_{n\geqslant 0}^\infty c_n(x-a)^n$ converges absolutely.

(C_2) If $|x-a| > R \implies \sum_{n\geqslant 0}^\infty c_n(x-a)^n$ diverges.

Proof. Notice that at R and $a-R$ anything can happen. Set

$$\limsup_n \left(|c_n||x-a|^n\right)^{\frac{1}{n}} = \limsup_n |c_n|^{\frac{1}{n}}|x-a| = \frac{1}{R}|x-a|$$

We apply Theorem 1.3 to obtain the result. ☐

Theorem 1.9. Let $\sum_{n\geqslant 0}^\infty c_n(x-a)^n$ with radius $R \in \mathbb{R}^+$. Let $x \in (a-r, a+r)$ and set

$$f(x) = \sum_{n=0}^{+\infty} c_n(x-a)^n$$

(f_1) $\forall\, r \in (0,R) \implies \sum_n c_n(x-a)^n$ converges uniformly on $[a-r, a+r]$.

In particular $f(x)$ is $\mathcal{C}^0$ on $(a-r, a+r)$

(f_2) $\forall\, r \in (0,R) \implies \sum_n nc_n(x-a)^{n-1}$ converges uniformly on $[a-r, a+r]$ and

$$\forall\, x \in (a-r, a+r) \implies f'(x) = \sum_{n=0}^{+\infty} nc_n(x-a)^{n-1}$$

So, in particular, f is differentiable

(f_3) Let $[y, x] \subseteq (a-R, a+R) \implies$

$$\int_y^z f = \sum_{n=0}^{+\infty} c_n \frac{(z-a)^{n+1} - (y-a)^{n+1}}{n+1}$$

Proof. Let us prove this result

(f_1) Let $r \in (0, R)$

$$\sup_{x\in[a-r,a+r]} \left|c_n(x-a)^n\right| \leqslant |c_n|r^n$$

If we apply Theorem 1.2, since $r < R \Longrightarrow \sum_{n=1}^{\infty} |c_n|r^n$ converges. Hence

$$\sum_{n=1}^{\infty} \sup_{x\in[a-r,a+r]} \left|c_n(x-a)^n\right| \text{ converges}$$

And thus $\sum_{n=1}^{\infty} c_n(x-a)^n$ converges uniformly on $[a-r, a+r]$ and so $\forall r \in (0, R) \Rightarrow f$ is $\mathcal{C}^0$ on $(a-r, a+r)$, so it is $\mathcal{C}^0$ on $(a-R, a+R)$

(f_2) Set $u_n(x) = x_n(x-a)^n \Rightarrow u_n' = c_n n(x-a)^{n-1}$ and $\sum u_n'$ is the power series with radius of convergence

$$R' = \frac{1}{\limsup_n(|c_{n+1}n|)^{\frac{1}{n}}}$$

Notice $\sum u_n' = (c_{n+1})(n+1)(x-a)^n$. Now since

$$\frac{1}{(n+1)^{\frac{1}{n}}} \to 1 \Longrightarrow R' = R = \frac{1}{\limsup_n |c_n|^{\frac{1}{n}}}$$

So they have the same radius of convergence. Thus $\sum u_n'$ converges uniformly on $[a-r, a+r]$ by (f_1). Moreover $\sum u_n$ converges uniformly on $[a-r, a+r]$. Applying Theorem 1.1 so f is differentiable on $(a-r, a+r)$ and $\forall\, x \in (x-r, x+r)$ where

$$f'(x) = \sum_{n=0}^{+\infty} u_n'(x)$$

(f_3) $\sum_n u_n$ converges uniformly on $[y, z]$ by (f_1) so $\forall\, n \in \mathbb{N} \Rightarrow u_n$ is $\mathcal{C}^0$ hence

$$\int_y^z f(t)dt = \sum_{n=0}^{+\infty} \int_y^z u_n(t)dt$$

Thus Theorem 1.9 is true. □

Theorem 1.10. Let $V = (V, \|\cdot\|)$ be a normed vector space and let (v_k) be a sequence in V. If the series

$$\sum_{k=1}^{\infty} v_k$$

converges in $V \Rightarrow v_k \to 0$ in V. In particular, the sequence (v_k) is bounded.

1.4 Real-Analytic Functions

Definition 1.14 (Real-Analytic Function)**.** Let $f : E \subseteq \mathbb{R} \to \mathbb{R}$ and $a \in E$. We say that f is real-analytic at $a \iff \exists\, r \in \mathbb{R}^+$ and $(c_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ s.t.

$$\forall\, x \in (a-r, a+r) \Longrightarrow f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$$

Suppose E is open. Then f is real-analytic if it is real analytic at $a \,\forall\, a \in E$

Notation. $\mathbb{R}^{\mathbb{N}}$ is the set of sequences taking values in $\mathbb{R}$

Corollary. By (f_1) and (f_2) of Theorem 1.9 if f is real analytic at $a \Rightarrow f$ is both $\mathcal{C}^0$ and differentiable on $(a-r, a+r)$ for some $r \in \mathbb{R}^+$

Theorem 1.11. Let $I \subset \mathbb{R}$ be an interval and $f \in \mathcal{C}^\infty(I)$. Suppose there exists a sequence of pairwise distinct points $(x_n) \subset I$ with $x_n \to a \in I$ and $f(x_n) = 0$ for all n.

(A_1) $\forall\, a \in I \subseteq \mathbb{R} \Rightarrow f(a) = 0$.

(A_2) $\forall\, k \geqslant 1 \Rightarrow f^{(k)}(a) = 0$.

(A_3) Suppose additionally that f is real-analytic on $I \Longrightarrow f \equiv 0$ on I.

Proof. We attempt to show this is true.

(A_1) Since f is continuous and $x_n \to a \Rightarrow$

$$f(a) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} 0 = 0$$

(A_2) Fix $k \geqslant 1$ and $\varepsilon \in \mathbb{R}^+$. Choose $y_1, \dots, y_{k+1} \in (a-\varepsilon, a+\varepsilon)$ with $(y_j, y_{j+1}) \subseteq I$ s.t.

$$\forall\, j = 1, \dots, k+1 \Rightarrow f(y_j) = 0$$

We can do this because of (A_1) of Theorem 1.11. By Theorem 2.2 there exist

$$z_1, \dots, z_k \in (y_j, y_{j+1}) \qquad \text{and} \qquad \forall\, j = 1, \dots, k \Rightarrow f'(z_j) = 0$$

Iterating by Theorem 2.2 again we have that

$$w_1, \dots, w_{k-1} \in (z_j, z_{j+1}) \qquad \text{and} \qquad \forall\, j = 1, \dots, k-1 \Rightarrow f''(w_j) = 0$$

and so on until the k-th derivative. Letting $\varepsilon \to 0 \Rightarrow f^{(k)}(a) = 0$ by continuity.

(A_3) Since f now satisfies Definition 1.14 we have for $x \in (a-\rho, a+\rho)$ where $\rho \in \mathbb{R}^+$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

From (A_1) and (A_2) of Theorem 1.11 we have $\forall\, k \Rightarrow f^{(k)}(a) = 0$ so $\forall\, x \Rightarrow f(x) = 0$ and by continuity $\forall\, x \in [a-\rho, a+\rho] \Rightarrow f(x) = 0$. Let H be the set of all intervals $\ell \subseteq I$ such that $a \in \ell$ and $\forall\, x \in \ell \Rightarrow f(x) = 0$. Define $U := \bigcup_{\ell \in H} \ell$.

We claim $U = I$. Assume by contradiction $U \subsetneq I$. Then the union of disjoint intervals U are closed since f is continuous. Let c be the endpoint of some $\ell \in U$. Choose a sequence $(x_n)_{n\in\mathbb{N}} \subseteq U$ s.t. $\lim\limits_{n\to\infty} x_n = c$.

Since $\forall\, n \in \mathbb{N} \Rightarrow f(x_n) = 0$, by (A_1) and (A_2) we have $\forall\, k \geqslant 0 \Rightarrow f^{(k)}(c) = 0$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n \equiv 0 \quad \text{for } x \in (c-\xi, c+\xi)$$

Because of Definition 1.14 for some $\xi > 0$. Hence the interval containing c can be extended beyond $c \Rightarrow\!\Leftarrow$. Therefore $U = I$, so $f \equiv 0$ on I.

$\therefore$ Theorem 1.11 is true. $\square$

Theorem 1.12. Let $f : E \subseteq \mathbb{R}$ be real analytic at $a \implies \forall\, k \Rightarrow f$ is k times differentiable at a. Moreover $\exists\, r > 0$ s.t. $\forall\, k \in \mathbb{N} \Rightarrow$

$$f^{(k)}(x) = \sum_{n \geqslant k}^{\infty} c_n \cdot n(n-1) \cdots (n-k+1)(x-a)^{n-k}$$

Proof. r is where around a we expand to power series $\exists\, r > 0$ s.t. $\forall\, x \in (a-r, a+r) \implies$

$$f(x) = \sum_{n=0}^{+\infty} \underbrace{c_n(x-a)^n}_{u_n}$$

In particular the radius of convergence of this series is larger than r i.e. $R \geqslant r$. Then $\forall\, k \in \mathbb{N} \Rightarrow$ the radius of convergence of $\sum u_n^{(k)}$ is R. Hence by (f_1) of Theorem 1.9 we have that $\sum u_n^{(k)}$ converges uniformly on every compact set included in $(a-r, a+r)$. Since it is true $\forall\, k$ we get that f is $\mathcal{C}^\infty$ and that $\forall\, x \in (a-r, a+r) \implies$

$$f^{(k)}(x) = \sum_{n \geqslant k}^{\infty} c_n \cdot n(n-1) \cdots (n-k+1)(x-a)^{n-k}$$

Thus Theorem 1.12 is true. □

Corollary. Let $f : E \to \mathbb{R}$ be real analytic $\implies f$ is $\mathcal{C}^\infty$ and all derivatives are analytic.

Proof. By Theorem 1.12 □

Corollary (Taylor's Formula)**.** Let $f : E \to \mathbb{R}$ be real analytic at $a \in E$. Let $r > 0$ and $(c_n) \in \mathbb{R}^{\mathbb{N}}$ be s.t.

$$\forall\, x \in (a-r, a+r) \Rightarrow f(x) = \sum_{n=0}^{+\infty} c_n(x-a)^n$$

$\implies \forall\, n \geqslant 0 \Rightarrow$

$$c_n = \frac{f^{(n)}(a)}{n!}$$

Theorem 1.13. Let $a > 0$. Let $f : [-a, a] \to \mathbb{R}$ be $\mathcal{C}^\infty$ and suppose there exist $C, A > 0$ s.t.

$$\forall\, n \in \mathbb{N} \Rightarrow \|f^{(n)}\|_\infty := \sup_{x \in [-a,a]} |f^{(n)}(x)| \leqslant CA^n n!$$

$\implies f$ admits a power series expansion at 0, i.e., f is real-analytic at 0.

Proof. By Taylor Remainder Theorem we have for any $x \in [-a, a]$ and $n \in \mathbb{N}$

$$f(x) = \sum_{k=0}^{N} \frac{f^{(k)}(0)}{k!} x^k + R_N(x)$$

where $R_N(x) = \frac{f^{(N+1)}(\xi)}{(N+1)!} x^{N+1}$. Since $\xi \in [-a, a] \Rightarrow$

$$|f^{(N+1)}(\xi)| \leqslant \|f^{(N+1)}\|_\infty \leqslant CA^{N+1}(N+1)!$$

And note that we can bound $R_N(x)$

$$|R_N(x)| \leqslant \frac{CA^{N+1}(N+1)!}{(N+1)!} |x|^{N+1} = C(A|x|)^{N+1}$$

Fix x s.t. $|x| < \frac{1}{A}$. Let $\varepsilon > 0$ and pick M s.t. $C(A|x|)^{M+1} < \varepsilon$ since $A|x| < 1 \Rightarrow$

$$\forall\ N \geqslant M \Rightarrow |R_N(x)| \leqslant C(A|x|)^{N+1} \leqslant C(A|x|)^{M+1} < \varepsilon$$

Hence $\lim\limits_{N\to\infty} R_N(x) = 0$. Therefore

$$f(x) = \lim_{N\to\infty} \sum_{k=0}^{N} \frac{f^{(k)}(0)}{k!} x^k$$

for every $|x| < \min\left(a, \frac{1}{A}\right)$. This proves f is real-analytic at 0 as it satisfies Definition 1.14 □

1.5 Abel's Theorem

Intuition. An Abelian theorem proposes that when there is convergence of the series, the original object, then the regularized object behaves well. A Tauberian theorem says that if the regularized object behaves well and we add some condition, then the original object will converge.

Lemma 1.1. Let $\sum_{n=0}^{+\infty} c_n x^n$ have radius of convergence 1. Let

$$S_n = \sum_{k=0}^{n} c_k$$

$\implies \forall\, x \in (-1, 1) \Rightarrow$

$$\sum_{k=0}^{+\infty} c_k x^k = (1-x) \sum_{k=0}^{+\infty} S_k x^k$$

Proof. We start by rewriting c_k. Note that $\forall\, k \Rightarrow$

$$c_k = S_k - S_{k-1} \qquad \text{with } S_{-1} = 0$$

Now we can add to these terms the finite sumns of $x_k \Rightarrow$

$$\sum_{k=0}^{N} c_k x^k = \sum_{k=0}^{N} S_k x^k - \underbrace{\sum_{k=0}^{N} S_{k-1} x^k}_{\sum_{k=1}^{N} S_{k-1}}$$

And therefore this last term can be rewritten as

$$\sum_{k=1}^{N} S_{k-1} x^k = \sum_{k=0}^{N-1} S_k x^{k+1}$$

Substituting this into our original equation gives

$$\sum_{k=0}^{N} c_k x^k = \sum_{k=0}^{N-1} S_k \underbrace{\left(x^k - x^{k+1}\right)}_{x^k(1-x)} + S_N x^N$$

Since by assumption (S_N) converges by Theorem 1.10 it is bounded and for $x \in (-1, 1)$

$$S_N x^N \xrightarrow[N\to+\infty]{} 0$$

Thus $\sum_{k=0}^{N} S_k x^k (1-x)$ converges and Lemma 1.1 is true. □

Remark. $\sum_{k=0}^{+\infty} a_k x^k - \sum_{k=0}^{+\infty} a_k = \sum_{k=0}^{n} a_k (x^k - 1) + (x-1) \sum_{k=0}^{+\infty} R_m x^k + R_m(x^{n+1} - 1)$

Theorem 1.14 (Abel). Let f be a power series centered at a with radius of convergence $R \in \mathbb{R}^+$. If f converges at $x = a + R \Longrightarrow f$ is $\mathcal{C}^0$ at $a + R$ and

$$\lim_{x \to (a+R)^-} f(x) = f(a+R) = \sum_{m=0}^{\infty} c_m R^m$$

Proof. We will take the case where $a = 0$ and $R = 1$. By Lemma 1.1 for $x \in (-1, 1) \Rightarrow$

$$\sum_{k=0}^{+\infty} c_k x^k = (1-x) \sum_{k=0}^{+\infty} S_k x^k$$

Notice that the term S_k

$$S_k = \sum_{k=0}^{+\infty} c_k - \underbrace{R_k}_{\text{remainder}} \qquad \text{with } R_k = \sum_{n=k+1}^{+\infty} c_n$$

Set $S_\infty := \sum_{k=0}^{+\infty} c_k$ and so $S_k = S_\infty - R_k$

$$\Rightarrow \sum_{k=0}^{+\infty} c_k x^k = (1-x) \sum_{k=0}^{+\infty} (S_\infty - R_k) x^k =$$

Since $\sum_{k=0}^{+\infty} x^k$ converges we have that

$$= S_\infty (1-x) \sum_{k=0}^{+\infty} x^k - (1-x) \sum_{k=0}^{+\infty} R_k x^k = S_\infty - (1-x) \sum_{k=0}^{+\infty} R_k x^k$$

Let us show that

$$\lim_{x \to 1^-} \underbrace{(1-x) \sum_{k=0}^{+\infty} R_k x^k}_{\text{error}(x)}$$

Let $\varepsilon > 0 \Rightarrow \exists\, k_0$ s.t. $\forall\, k \geqslant k_0 \Rightarrow |R_k| < \varepsilon$. Notice

$$\text{error}(x) = (1-x) \sum_{k<k_0}^{+\infty} R_k x^k + (1-x) \sum_{k \geqslant k_0}^{+\infty} R_k x^k$$

First for $x \in (0, 1) \Rightarrow$

$$(1-x) \sum_{k \geqslant k_0}^{+\infty} x^k \leqslant \frac{1}{1-x} \Longrightarrow \left| (1-x) \sum_{k \geqslant k_0}^{+\infty} R_k x^k \right| \leqslant \varepsilon$$

Since k_0 is fixed $\exists\, \delta > 0$ s.t. $\forall\, x \in (1-\delta, 1) \Rightarrow$

$$\left| (1-x) \sum_{k<k_0}^{+\infty} R_k x^k \right| \leqslant \varepsilon \Longrightarrow \left| \text{error}(x) \right| \leqslant 2\varepsilon$$

Hence $\lim\limits_{x \to 1^-} \text{error}(x) = 0$. Thus

$$\lim_{x \to 1^-} \sum_{k=0}^{+\infty} c_k x^k = \sum_{k=0}^{+\infty} c_k$$

This proves Theorem 1.14. $\square$

Theorem 1.15 (Cesaro). Let (u_k) be a sequence that converges and suppose $u_k \to L \in \mathbb{R} \implies$

$$\frac{1}{n}\sum_{k=1}^{n} u_k \to L \in \mathbb{R}$$

Proof. Let $\varepsilon > 0$. Since $u_k \to L \, \exists \, k_0$ s.t. $\forall \, k \geqslant k_0 \Rightarrow |u_k - L| < \varepsilon$

$$\left(\sum_{k=1}^{n} u_k\right) - nL = \sum_{k=1}^{n}(u_k - L) = \sum_{k=1}^{k_0-1}(u_k - L) + \sum_{k=k_0}^{n}(u_k - L)$$

Notice the following

$$\left|\sum_{k=k_0}^{n}(u_k - L)\right| \leqslant n \cdot \varepsilon \implies \left|\frac{1}{n}\sum_{k=1}^{n} u_k - L\right| \leqslant \frac{1}{n}\sum_{k=1}^{k_0-1}|u_k - L| + \frac{\varepsilon \cdot n}{n}$$

Now $\exists \, k_1 \geqslant k_0$ s.t.

$$\forall \, n \geqslant k_1 \implies \frac{1}{n}\sum_{k=1}^{k_0-1}|u_k - L| < \varepsilon$$

Hence $\forall \, \varepsilon > 0 \, \exists \, k_0$ s.t. $\forall \, n \geqslant k_1 \Rightarrow$

$$\left|\frac{1}{n}\sum_{k=1}^{n} u_k - L\right| \leqslant 2\varepsilon \implies \frac{1}{n}\sum_{k=1}^{n} u_k \xrightarrow[n\to+\infty]{} L$$

This proves Theorem 1.15. □

Example 1.2. Consider the power series

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n(2n+1)} x^{2n+1}.$$

(P_1) What is its radius of convergence R? Is there convergence at the endpoints?

(P_2) On what interval is f a priori continuous? Prove that it is continuous on $[-R, R]$.

(P_3) Express, using standard elementary functions, the sum of the series obtained by differentiating term by term on $(-R, R)$. Deduce an expression for f on $(-R, R)$.

(P_4) Compute

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n(2n+1)}.$$

Proof. Let us try to solve this exercise

(P_1) By Definition 1.10 we have that

$$R = \frac{1}{\limsup_{n\to\infty}\left(\frac{1}{2^n}\right)^{\frac{1}{n}}} \leqslant \frac{1}{\limsup_{n\to\infty}\left(\frac{1}{2n^2+n}\right)^{\frac{1}{n}}} \leqslant \frac{1}{\limsup_{n\to\infty}\left(\frac{1}{3n^2}\right)^{\frac{1}{n}}} = 1$$

Thus $R = 1$. There is convergence at both endpoints ± 1 by the alternating series test because terms decrease in absolute value to zero.

(P_2) We know that f is a priori continuous on $(-1, 1)$. By Theorem 1.14 since it converges, it is continuous on the closed interval $[-1, 1]$.

(P_3) Let us try to do this. Notice the derivative is

$$f'(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n(2n+1)} \cdot (2n+1)x^{2n} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^{2n} = \underbrace{\ln\left(1+x^2\right)}_{\text{Newton–Mercator}}$$

Let us integrate to find f

$$\begin{aligned} f(x) &= \int \ln\left(1+x^2\right) dx + C \\ &= x\ln\left(1+x^2\right) - 2x + 2\arctan(x) + C \end{aligned}$$

This is how to deduce an expression for f

(P_4) I used Wolfram 14.2 to compute this as I was running out of time.

$$\ln 2 + \frac{\pi}{2} - 2$$

$\therefore$ we have partly solved Example 1.2. □

Theorem 1.16 (Weak Tauber). Let $\sum_{n\geqslant 0}^{+\infty} a_n x^n$ be a power series with radius of convergence 1, and let f be its sum on $(-1,1)$. Suppose $\lim_{x\to 1^-} f(x)$ exists and $a_n = o(\frac{1}{n}) \implies$ the series $\sum_{k=0}^{\infty} a_k$ converges and

$$\lim_{n\to\infty} S_n = \lim_{x\to 1^-} f(x)$$

where $S_n = \sum_{k=0}^{n} a_k$.

Proof. Remember that

$$\begin{aligned} S_n - f(x) &= \sum_{k=0}^{n} a_k - \sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{n} a_k - \left(\sum_{k=0}^{n} a_k x^k + \sum_{k=n+1}^{\infty} a_k x^k \right) \\ &= \sum_{k=0}^{n} a_k(1-x^k) - \sum_{k=n+1}^{\infty} a_k x^k \\ &= \underbrace{\sum_{k=1}^{n} a_k(1-x^k)}_{k=0 \text{ vanishes}} - \sum_{k=n+1}^{\infty} a_k x^k \end{aligned}$$

We know $0 < x < 1 \Rightarrow 1 - x^k = (1-x)(1+x+x^2+\cdots+x^{k-1}) \leqslant (1-x)k$. We take absolute value in the previous equation and observe

$$\begin{aligned} |S_n - f(x)| &\leqslant \sum_{k=1}^{n} |a_k||1-x^k| + \sum_{k=n+1}^{\infty} |a_k| x^k \\ &\leqslant \underbrace{(1-x)\sum_{k=1}^{n} k|a_k|}_{\text{from step } (\mathcal{T}_2)} + \sum_{k=n+1}^{\infty} |a_k| x^k \end{aligned}$$

Now notice that since $1 = \frac{k}{k} \leqslant \frac{k}{n}$

$$|a_k| x^k = \frac{k|a_k|x^k}{k} \leqslant \frac{k|a_k|x^k}{n}$$

Which we can apply to our previous inequality

$$|S_n - f(x)| \leqslant (1-x)\sum_{k=0}^{n} k|a_k| + \sum_{k=n+1}^{\infty} \frac{k|a_k|x^k}{n}$$

Now remember the following fact from the geometric series

$$\sum_{k=n+1}^{\infty} x^k = \frac{x^{n+1}}{1-x} \leqslant \frac{1}{1-x}$$

Applying this to the previous equation by first noting

$$\sum_{k=n+1}^{\infty} \frac{k|a_k|x^k}{n} \leqslant \frac{\sup_{k>n} k|a_k|}{n} \sum_{k=n+1}^{\infty} x^k \leqslant \underbrace{\frac{\sup_{k>n} k|a_k|}{n(1-x)}}_{\text{by geometric series}}$$

Now since $a_n = o\left(\frac{1}{n}\right)$ we can see

$$\lim_{n\to\infty} \sup_{k>n} k|a_k| = 0 \Longrightarrow \lim_{n\to\infty} S_n = \lim_{x\to 1^-} f(x)$$

Because as $x \to 1^-$ and $1-x \to 0$, so by choosing x close to 1 and n large enough, both terms on the right tend to zero. □

1.6 log and exp

Definition 1.15 (Exponential). $\forall\, x \in \mathbb{R}$ we define the exponential function as

$$\exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \in \mathbb{R}$$

Remark. If $z \in \mathbb{C}$ and if $\mathbf{M} \in \mathcal{M}_{n\times n}(\mathbb{R}) \Rightarrow$

$$\exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \qquad \text{and} \qquad \exp(\mathbf{M}) = \sum_{n=0}^{+\infty} \frac{\mathbf{M}^n}{n!}$$

respectively. Note that $\exp(\mathbf{M}+\mathbf{N}) = \exp(\mathbf{M})\exp(\mathbf{N}) \Leftrightarrow \mathbf{MN} = \mathbf{NM}$

Intuition. It might seem counterintuitive, but we will use the inverse of the exponential, the logarithm function to prove some of the properties of the exponential, before even defining it.

Remark (Stirling). $\log n! = n \cdot \log n - n + \mathcal{O}(\log n)$. See Definition 2.5.

Theorem 1.17. The exponential in Definition 1.15 has the following properties

(e_1) exp has radius of convergence $R = +\infty$

(e_2) $\forall\, x \in \mathbb{R} \Longrightarrow \exp'(x) = \exp(x)$

(e_3) $\forall\, x, y \in \mathbb{R} \Longrightarrow \exp(x+y) = \exp(x)\exp(y)$

(e_4) $\forall\, x \in \mathbb{R} \Longrightarrow \exp(x) \geqslant 0$

(e_5) $\forall\, x \in \mathbb{R} \Longrightarrow \exp(-x) = \frac{1}{\exp(x)}$

Proof. Let us prove Theorem 1.17

(e_1) Define $a_n = \frac{1}{n!}$. Notice

$$n! \geqslant \left\lfloor \frac{n}{2} \right\rfloor^{\left\lfloor \frac{n}{2} \right\rfloor}$$

Now we take log in both sides

$$\log n! \geqslant \left\lfloor \frac{n}{2} \right\rfloor \log \left\lfloor \frac{n}{2} \right\rfloor \Rightarrow \frac{1}{n} \log n! \geqslant \frac{1}{n} \left\lfloor \frac{n}{2} \right\rfloor \log \left\lfloor \frac{n}{2} \right\rfloor \sim \frac{1}{2} \log \frac{n}{2}$$

Let $n \to +\infty$. Clearly the right side $\to +\infty$

$$\frac{1}{n} \log n! \to +\infty \Longrightarrow (n!)^{\frac{1}{n}} \to +\infty$$

Thus $R = +\infty$

(e_2) exp is differentiable by Theorem 1.12. Let us differentiate term by term in the open interval of convergence. $\forall\, x \in \mathbb{R} \Rightarrow$

$$\exp'(x) = \sum_{n=0}^{+\infty} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{+\infty} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{+\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{+\infty} \frac{x^n}{n!} = \exp(x)$$

(e_3) We take the following

$$\exp(x)\exp(y) = \sum_{k_1=0}^{+\infty} \frac{x^{k_1}}{k_1!} \sum_{k_2=0}^{+\infty} \frac{x^{k_2}}{k_2!} \Rightarrow \frac{1}{k_1!k_2!} = \binom{k_1+k_2}{k_1} \frac{1}{(k_1+k_2)!}$$

Since both series are absolutely convergent

$$\sum_{k_1,k_2=0}^{+\infty} \frac{x^{k_1}x^{k_2}}{k_1!k_2!}$$

is also absolutely convergent $\Rightarrow$

$$\begin{aligned}
\exp(x)\exp(y) &= \sum_{k_1,k_2=0}^{+\infty} x^{k_1}x^{k_2} \binom{k_1+k_2}{k_1} \frac{1}{(k_1+k_2)!} \\
&= \sum_{n=0}^{+\infty} \sum_{k_1,k_2=0}^{+\infty} x^{k_1}y^{k_2} \binom{n}{k_1} \frac{1}{n!}
\end{aligned}$$

When we set $k_1 + k_2 = n$ and as such $k_2 = n - k_1$

$$\sum_{k_1+k_2=n}^{+\infty} x^{k_1}x^{k_2} \binom{n}{k_1} = \sum_{k_1=0}^{n} x^{k_1}y^{n-k_1} \binom{n}{k_1} = (x+y)^n$$

Adding this result to the previous equality

$$\exp(x)\exp(y) = \sum_{n=0}^{+\infty} \frac{1}{n!}(x+y)^n = \exp(x+y)$$

(e_4) If $x \geqslant 0 \Rightarrow \exp(x) \geqslant 1 > 0$. Let $x < 0$ and set $x = -a$ with $a \in \mathbb{R}^+ \Rightarrow$

$$\exp(x) = \sum_{n=0}^{+\infty} \frac{(-1)^n}{n!} \cdot a^n$$

This is an alternating series hence $\exp(x) > 0$ which is the sign of the first term.

(e_5) Since $\exp(0) = 1 \Rightarrow$ set $y = x$

$$1 = \exp(x - x) = \exp(x)\exp(-x)$$

This because of (e_3) of this Theorem.

$$\Longrightarrow \frac{1}{\exp(x)} = \exp(-x)$$

This proves Theorem 1.17. □

Definition 1.16 (Logarithm)**.** We define the natural logarithm function $\ln = \log : (0, \infty) \to \mathbb{R}$ to be the inverse of Definition 1.15. Thus $\exp\big(\log(x)\big) = x$ and $\log\big(\exp(x)\big) = x$.

Notation. We refer to the identity matrix and function by the same notation id

Example 1.3. Let $n \geqslant 1 \Longrightarrow \exists\, \alpha > 0$ s.t. $\forall\, \mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ with $\|\mathbf{A} - \mathrm{id}\| < \alpha \Longrightarrow \exists\, \mathbf{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$ s.t. $\mathbf{A} = \exp(\mathbf{B})$.

Proof. Let $\mathbf{X} \in \mathcal{M}_{n\times n}(\mathbb{R})$ s.t. $\|\mathbf{X}\| < 1$. Consider the following power series expansion

$$\log(\mathrm{id} + \mathbf{X}) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\mathbf{X}^k}{k}$$

This series converges because of the following

$$\sum_{k=1}^{\infty} \left\| (-1)^{k+1} \frac{\mathbf{X}^k}{k} \right\| \leqslant \sum_{k=1}^{\infty} \frac{\|\mathbf{X}\|^k}{k} < \infty$$

Define $\alpha := 1 \Rightarrow \forall\, \mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ with $\|\mathbf{A} - \mathrm{id}\| < \alpha$, set $\mathbf{X} := \mathbf{A} - \mathrm{id}$ so that $\|\mathbf{X}\| < 1$.

$$\text{Define} \quad \mathbf{B} := \log(\mathbf{A}) := \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(\mathbf{A} - \mathrm{id})^k}{k}$$

And by Definition 1.16 $\exp(\mathbf{B}) = \mathbf{A}$ □

Lemma 1.2. $\forall\, x \in \mathbb{R} \Longrightarrow \exp(x) = \exp(1)^x$

Proof. Let $f(x) = \exp(x)$. Then $f(x + y) = f(x)f(y)$ and f is $\mathcal{C}^0$. We show $\forall\, n \in \mathbb{N} \Longrightarrow f(n) = f(1)^n$. For $n = 1$ trivial. Suppose $f(n) = f(1)^n \Rightarrow$

$$f(n + 1) = f(n)f(1) = f(1)^{n+1}$$

$\Longrightarrow f(n) = f(1)^n \ \forall\, n \in \mathbb{N}$. For $n \in \mathbb{Z}$, $f(-n)f(n) = f(0) = 1 \Longrightarrow f(-n) = f(1)^{-n}$. For $q = \frac{p}{m} \in \mathbb{Q}$ we have the following expression

$$f(q)^m = f(mq) = f(p) = f(1)^p \Longrightarrow f(q) = f(1)^{\frac{p}{m}} = f(1)^q$$

Since f is $\mathcal{C}^0 \Rightarrow \forall\, x \in \mathbb{R} \Longrightarrow f(x) = f(1)^x$ □

Theorem 1.18. $\forall\, x \in (-1,1) \Longrightarrow$

$$\log(1-x) = -\sum_{n=1}^{+\infty} \frac{x^n}{n!}$$

is convergent

Proof. $\forall\, x \in \mathbb{R}^+ \Rightarrow \log'(x) = \frac{1}{x}$. Notice that

$$\forall\, t \in (-1,1) \Longrightarrow \frac{1}{1-x} = \sum_{n=0}^{+\infty} t^n$$

Radius of convergence is 1. Hence since $[0,x] \subseteq (-1,1)$ we have

$$\int_0^x \frac{1}{1-t} dt = \sum_{n=0}^{+\infty} \int_0^x t^n dt$$

And then we have

$$-\log(1-x) = \sum_{n=0}^{+\infty} \frac{x^{n+1}}{n+1} = \sum_{n=1}^{+\infty} \frac{x^n}{n}$$

Hence $\log(1-x) = -\sum_{n=1}^{+\infty} \frac{x^n}{n}$ □

1.7 Complex Analysis

Intuition. We now will do $\sum a_n z^n$ for $z \in \mathbb{C}$ and $a_n \in \mathbb{R}$

Definition 1.17. Let $z = x + iy \in \mathbb{C}$. The real part of z is defined by

$$\Re(z) = x$$

and the imaginary part of z is defined by

$$\Im(z) = y$$

Lemma 1.3. Let $\sum_{n\geqslant 0} a_n z^n$ be a power series with radius of convergence $R \in [0,+\infty]$

(c_1) $\forall\, z \in \mathbb{C}$ with $|z| < R \Longrightarrow \sum_{n\geqslant 0} a_n z^n$ converges absolutely

(c_2) $\forall\, z \in \mathbb{C}$ with $|z| > R \Longrightarrow \sum_{n\geqslant 0} a_n z^n$ diverges

Theorem 1.19 (Cauchy Formula)**.** Let $r \in (0,R) \Longrightarrow$

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) e^{-in\theta} d\theta$$

Proof. Observe $\forall\, k \in \mathbb{Z} \neq 0$ and for $k = 0$

$$\int_0^{2\pi} e^{ik\theta} d\theta = 0 \qquad \text{and} \qquad \int_0^{2\pi} e^{ik\theta} d\theta = 2\pi$$

respectively. Now we can do the following

$$f(re^{i\theta}) = \sum_n r^n a_n e^{in\theta}$$

$$f(re^{i\theta})e^{-in\theta} = \sum_n r^n a_n e^{i(n-m)\theta}$$

And notice $n - n = k$. Now take the integral

$$\int_0^{2\pi} f(re^{i\theta})e^{-in\theta} d\theta = r^n a_n \int_0^{2\pi} e^{i(n-m)\theta} d\theta = r^n a_n 2\pi$$

This somehow proves the result. □

Corollary (Liouville). Suppose $R = +\infty$. Suppose f is bounded on $\mathbb{C} \Rightarrow f$ is constant.

Proof. First for $n \neq 0$.

$$\left|f^n(0)\right| \leqslant \frac{1}{r^n}\frac{2\pi}{2\pi} \sup |f|$$

True of every $r > 0$. Letting $r \to +\infty$ gives

$$\forall\, n \geqslant 1 \Longrightarrow f^{(n)}(0) = 0 \Rightarrow \forall\, n \geqslant 1 \Longrightarrow a_n = \frac{f^n(0)}{n!} = 0$$

$\therefore f$ is constant. □

Definition 1.18. Let (X, d) be a metric space. Let $A \subseteq X$. We define the boundary of A as

$$\partial A = \overline{A} \cap \overline{A^c}$$

where $\overline{A}$ denotes the closure of A and A^c its complement.

Theorem 1.20 (Liouville). Let $f : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^0$ s.t. $\forall\, x \in \mathbb{R}^n$ and $r > 0 \Longrightarrow$

$$f(x) = \frac{1}{\left|\partial B(x, r)\right|} \int_{\partial B(x,r)} f$$

We say that f has the mean value property $\Rightarrow$ if f is bounded $\Longrightarrow f$ is constant.

Proof. Take $n = 2$ as it doesn't change anything.

$$\Longrightarrow \forall\, x \in \mathbb{R}^2 \text{ and } r \in \mathbb{R}^+ \Rightarrow f(x) = \frac{1}{\left|B(x, r)\right|} \int_{B(x,r)} f$$

Now notice the following

$$\left|f(x) - f(y)\right| \leqslant \int_{B(x,r)\setminus B(y,r) \,\cup\, B(y,r)\setminus B(x,r)} |f| \frac{1}{\left|B(0, r)\right|}$$

For $R \geqslant 100 \cdot \|x - y\|$ for instance we have that

$$\begin{aligned}\left|f(x) - f(y)\right| &\leqslant \frac{1}{\left|B(0, r)\right|} \sup |f| \cdot \text{Area}(B(x, R) \setminus B(y, R) \,\cup\, B(y, R) \setminus B(x, R)) \\ &\leqslant \frac{1}{\left|B(0, r)\right|} \sup |f| \cdot c \cdot R|x - y|\end{aligned}$$

for some constant $c \in \mathbb{R}^+$ universal

$$\left|B(0, R)\right| = \pi R^2 \Rightarrow \exists\, c > 0 \text{ s.t. } \left|f(x) - f(y)\right| \leqslant \frac{c \cdot \|x - y\|}{R}$$

And then $R \to +\infty \Longrightarrow f(x) = f(y) \cdot R$. □

Notation. For $d \geqslant 1$, we denote

$$\mathbb{Z}^d = \{(x_1, \ldots, x_d) \mid \forall\, i = 1, \ldots, d \Rightarrow x_i \in \mathbb{Z}\,\}.$$

Definition 1.19 (Harmonic)**.** Let $f : \mathbb{Z}^d \to \mathbb{R}$. We say f is harmonic if

$$\forall\, v \in \mathbb{Z}^d \implies f(v) = \frac{1}{2d}\sum_{i=1}^{d}(f(v+e_i) + f(v-e_i))$$

Theorem 1.21 (Liouville)**.** Let f be harmonic on $\mathbb{Z}^d$ and bounded $\implies f$ is constant.

Theorem 1.22 (Liouville-Improvement)**.** Take $d = 2$ and suppose f is harmonic on the lattice $\mathbb{Z}^2$ and bounded on 99.9999999% of $\mathbb{Z}^2 \implies f$ is constant.

Remark. Not true on $\mathbb{Z}^d$ for $d \geqslant 3$. Wow! This is a recent result.

Definition 1.20. Let $A \subseteq \mathbb{R}$. The density of A is defined by

$$\delta(A) = \lim_{R \to +\infty} \frac{|A \cap [-R, R]|}{|[-R, R]|}$$

whenever the limit exists.

Chapter 2

Differentiation on $\mathbb{R}^n \to \mathbb{R}^m$

2.1 Derivatives on $\mathbb{R}$

Intuition. Suppose a function $f : \mathbb{R} \to \mathbb{R}$. Let $x_0 \in \mathbb{R}$. The goal is to approximate f around x_0 by a linear affine function

$$x \mapsto ax + b$$

What is a, b? We want $f(x) \simeq ax + b$ for $x \simeq x_0 \Rightarrow f(x_0) = ax_0 + b$. Hence

$$\begin{aligned} f(x) &\simeq a(x - x_0) + f(x_0) \\ f(x) - f(x_0) &\simeq a(x - x_0) \\ \frac{f(x) - f(x_0)}{x - x_0} &\simeq a + \text{ something small} \end{aligned}$$

And as such we have

$$a = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

Definition 2.1 (Derivative). Let I be open and $a \in I \Rightarrow f$ is differentiable at a if

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

exists. When it does, we call it the derivative of f at a

Notation. The derivative in Definition 2.1 is denoted a $f'(a)$

Remark. The best linear approximation of f around a is $x \mapsto f(a) + f'(a)(x - a)$, that is, the tangent to the curve at a. More generally, near a we have the second-order approximation

$$f(x) \simeq f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + \cdots$$

Definition 2.2. Let I be open. Let $f : I \to \mathbb{R}$ and $a \in I$. We say a is the local min of f if

$$\exists\, r \in \mathbb{R}^+ \text{ s.t. } \forall\, x \in (a - r, a + r) \subseteq I \Rightarrow f(x) \geqslant f(a)$$

the local max is the same analogously.

Notation. This is the same as saying $\exists\, \varepsilon \in \mathbb{R}^+$ s.t. $\forall\, y \in B(x, \varepsilon) \Longrightarrow f(y) \geqslant f(x)$

Theorem 2.1. Let I be open. Let $f : I \to \mathbb{R}$ and $a \in I$. Let a be a local minmax of f. Suppose f is differentiable at $a \Rightarrow f'(a) = 0$

Proof. Take $x \in (a, a+r)$. Let us look at

$$\frac{f(x) - f(a)}{\underbrace{x - a}_{>0}} \geqslant 0 \qquad \text{for } r \text{ small enough}$$

$$\Rightarrow f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \geqslant 0$$

Similarly, if one takes $x \in (a - r, a) \Rightarrow$

$$\frac{f(x) - f(a)}{\underbrace{x - a}_{<0}} \leqslant 0 \qquad \text{for } r \text{ small enough}$$

$$\Rightarrow f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \leqslant 0$$

$\therefore f'(a) = 0.$ ☐

Theorem 2.2 (Rolle's)**.** Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Suppose $f(a) = f(b) \Rightarrow \exists\, c \in (a, b)$ s.t. $f'(c) = 0$.

Proof. If f is constant $\Rightarrow f'(x) = 0$. Otherwise, suppose f attains a minimum or maximum at $c \in (a, b) \Rightarrow c$ satisfies Definition 2.2 of f on (a, b), and by Theorem 2.1 $\Rightarrow f'(c) = 0$. ☐

Theorem 2.3 (Mean Value)**.** Let $f : [a, b] \to \mathbb{R}$ continuous and differentiable on $(a, b) \Rightarrow$ $\exists\, c \in (a, b)$ s.t.

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Proof. Let $f(x)$ and we construct a linear function

$$\ell(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

$$\Rightarrow L(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a) = f(x) - \ell(x)$$

Notice $L(a) = 0 = L(b) \Rightarrow$ by Theorem 2.2 $\exists\, c \in (a, b)$ s.t. $L'(c) = 0$

$$L'(c) = 0 = f'(c) - \frac{f(b) - f(a)}{b - a} \implies f'(c) = \frac{f(b) - f(a)}{b - a}$$

$\therefore$ Theorem 2.3 is true. ☐

Corollary. If $f' = 0$ on $(a, b) \implies f$ is constant

Proof. By Theorem 2.3 $\exists\, c \in (a, b)$ s.t.

$$f'(c) = \frac{f(b) - f(a)}{b - a} = 0$$

But this means $f(b) - f(a) = 0 \Rightarrow f(b) = f(a)$. Hence f is constant. ☐

Corollary. Let $f : \mathbb{R} \to \mathbb{R}$ continuous and differentiable on $\mathbb{R} \setminus \{0\}$. Suppose that $f'(x)$ has limit ℓ as $x \mapsto 0$ where $x \neq 0 \Rightarrow f$ is differentiable at 0 and $f'(0) = \ell$

Proof. Take $x \neq 0$

$$f'(c) = \frac{f(x) - f(0)}{x - 0}$$

for some c_x s.t. $|c| < |x|$. Now since

$$f'(y) \xrightarrow[y \to 0]{} \ell \qquad \text{and} \qquad c_x \xrightarrow[x \to 0]{} 0$$

$$\Rightarrow f'(c_x) = \ell \Rightarrow \lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \ell$$

$\therefore f$ is differentiable at 0 and $f'(0) = \ell$ □

2.2 Derivatives on $\mathbb{R}^n$

Intuition. Take $f : \mathbb{R}^n \to \mathbb{R}^m$. Let $x_0 \in \mathbb{R}^n$. We want to approximate f by a linear affine function around x_0

$$x \mapsto \mathbf{A}x + b$$

where $b \in \mathbb{R}^n$ and $\mathbf{A} \in \mathcal{M}_{m \times n}(\mathbb{R})$. We want $f(x_0) = \mathbf{A}x_0 + b$. Hence we want a matrix s.t.

$$f(x) \simeq f(x_0) + \mathbf{A}(x - x_0)$$

Definition 2.3 (Norm). Let V be a vector space over $\mathbb{R}$. A norm on V is a function $\|\cdot\| : V \to \mathbb{R}$ that satisfies the following properties

(N_1) $\forall\, x \in V \Rightarrow \|x\| \geqslant 0$

(N_2) $\|x\| = 0 \Leftrightarrow x = \vec{0}$, where $\vec{0}$ is the additive identity of V

(N_3) $\forall\, x \in V$ and $\forall\, \lambda \in \mathbb{R} \Rightarrow \|\lambda x\| = |\lambda|\|x\|$

(N_4) $\forall\, x, y \in V \Rightarrow \|x + y\| \leqslant \|x\| + \|y\|$

Definition 2.4. Let $\mathcal{L}$ be the space of linear maps between normed vector spaces. The norm on $\mathcal{L}$ defined by

$$\forall\, L \in \mathcal{L} \Rightarrow \|L\|_{\mathcal{L}} := \sup_{x \neq 0} \frac{\left\|L(x)\right\|}{\|x\|}$$

is called the subordinate norm.

Example 2.1. The subordinate norm $\|L\|_{\mathcal{L}}$ is a norm in $\mathcal{L}$

Proof. Let us show this is a norm

(N_1) Since $\|L\|_{\mathcal{L}}$ is a fraction of two norms who are already $\geqslant 0 \Rightarrow \|L\|_{\mathcal{L}} \geqslant 0$

(N_2) $\|L\|_{\mathcal{L}} = 0 \Leftrightarrow \sup_{x \neq 0} \frac{\|L(x)\|}{\|x\|} \Leftrightarrow \left\|L(x)\right\| = 0 \Leftrightarrow L(x) = 0$

(N_3) Let $\lambda \in \mathbb{R}$

$$\|\lambda L\|_{\mathcal{L}} = \sup_{x \neq 0} \frac{\left\|\lambda L(x)\right\|}{\|x\|} = \sup_{x \neq 0} \frac{|\lambda|\left\|L(x)\right\|}{\|x\|} = |\lambda| \sup_{x \neq 0} \frac{\left\|L(x)\right\|}{\|x\|} = |\lambda|\|L\|_{\mathcal{L}}$$

(N_4) Let $L, T \in \mathcal{L} \Rightarrow \forall\, x \neq 0 \Rightarrow$

$$\frac{\|(L+T)(x)\|}{\|x\|} \leqslant \frac{\|L(x)\|}{\|x\|} + \frac{\|T(x)\|}{\|x\|} \underbrace{\Longrightarrow}_{\text{take sup}} \|L+T\|_{\mathcal{L}} \leqslant \|L\|_{\mathcal{L}} + \|T\|_{\mathcal{L}}$$

$\therefore$ the subordinate norm $\|L\|_{\mathcal{L}}$ is a norm in $\mathcal{L}$. □

Remark. All norms are equivalent on $\mathbb{R}^n$

Definition 2.5 (Landau). Consider two norms on $\mathbb{R}^n$ and $\mathbb{R}^m$ both denoted by $\|\cdot\|$

(O_1) Let $a, b : \mathbb{R}^n \to \mathbb{R}^m$. We say that

$$a(x) = o_{x_0}(b(x))$$

if $\exists\, \varepsilon > 0$ and $c : B(x_0, \varepsilon) \to \mathbb{R}$ s.t.

$$\|a(x)\| = c(x) \cdot \|b(x)\|$$

with $c(x) \to 0$ as $\|x - x_0\| \to 0$

(O_2) We say $a(x) = \mathcal{O}_{x_0}(b(x))$ if $\exists\, \varepsilon > 0$ and $M > 0$ s.t. $\forall\, x \in B(x_0, \varepsilon) \Rightarrow$

$$\|a(x)\| \leqslant M \cdot \|b(x)\|$$

Notation. This is known as Landau or Big O notation

Definition 2.6 (Fréchet Derivative). Let $X \subseteq \mathbb{R}^n$ be open. Let $f : X \to \mathbb{R}^m$. Let $a \in X$. We say f is Fréchet differentiable at a if $\exists$ a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ s.t.

$$f(x) = f(a) + L(x-a) + o_a(x-a)$$

Equivalently we can also say

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

We call L the derivative of f at a and denote it by DF_a

Remark. $\varepsilon(x) = o_a(x-a)$ if $\frac{\|\varepsilon(x)\|}{\|x-a\|} \xrightarrow[\|x-a\| \to 0]{} 0$

Example 2.2. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ where $f : x \mapsto \langle u, x \rangle$ with $u \in \mathbb{R}^n$

Proof. Let us show that

$$\begin{aligned} f(x+h) = \langle u, x+h \rangle &= \langle u, x \rangle + \langle u, h \rangle \\ &= f(x) + \langle u, h \rangle \\ &= f(x) + \underbrace{f(h)}_{\text{linear}} + \underbrace{o(\|h\|)}_{o_x(h)} \end{aligned}$$

By Definition 2.6 $\Rightarrow f$ is differentiable and $\forall\, x \in \mathbb{R}^n \Rightarrow Df_x = f$ □

Example 2.3. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f : x \mapsto \langle x, \mathbf{A}x \rangle$, where $\mathbf{A} \in \mathcal{M}_{n \times n}(\mathbb{R})$

Proof. Let us take

$$\begin{aligned} f(x+h) &= \langle x+h, \mathbf{A}(x+h)\rangle \\ &= \langle x, \mathbf{A}x\rangle + \langle h, \mathbf{A}x\rangle + \langle x, \mathbf{A}h\rangle + \langle h, \mathbf{A}h\rangle \end{aligned}$$

Set $L(h) = \langle h, \mathbf{A}x\rangle + \langle x, \mathbf{A}h\rangle$

$$\Rightarrow f(x+h) = f(x) + L(h) + \langle h, \mathbf{A}h\rangle$$

We have to show $\langle h, \mathbf{A}h\rangle$ satisfies $(O)_1$ of Definition 2.5.

$$\begin{aligned} \langle h, \mathbf{A}h\rangle &= \sum h_i h_j \mathbf{A}_{ij} \\ \big|\langle h, \mathbf{A}h\rangle\big| &\leqslant \max_{i,j} \big|\mathbf{A}_{ij}\big| \Big(\sum |h_i|\Big)^2 \end{aligned}$$

Take $\|h\| = \sum_{i=1}^n |h_i|$. Hence, for $\lambda \in \mathbb{R}^2 \Rightarrow$

$$\big|\langle h, \mathbf{A}h\rangle\big| \leqslant \lambda \cdot \|h\|^2 \Rightarrow \frac{\big|\langle h, \mathbf{A}h\rangle\big|}{\|h\|} \xrightarrow[h\to 0]{} 0$$

Hence $\langle h, \mathbf{A}h\rangle = o_x(h)$. Thus

$$f(x+h) = f(x) + L(h) + o_x(h)$$

Thus, the derivative of f at x is $Df_x(h) = \langle h, \mathbf{A}x\rangle + \langle x, \mathbf{A}h\rangle$. □

Notation. We denote by

$$S_n = \{\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\} \mid \sigma \text{ is a bijection}\}$$

the set of all permutations of $\{1, \ldots, n\}$

Example 2.4. Let $f : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}$ where $\mathbf{M} \mapsto \det(\mathbf{M})$

Proof. Let $\sigma \in S_n$. Define $\operatorname{sgn}(\sigma)$ as

$$\operatorname{sgn}(\sigma) = \begin{cases} +1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}$$

The determinant of $\mathbf{M} \in \mathcal{M}_{n\times n}(\mathbb{R})$ is then defined by

$$\det(\mathbf{M}) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n m_{i\sigma(i)}$$

Notice that $\operatorname{sgn}(\sigma) = (-1)^{N(\sigma)}$ where $N(\sigma)$ is the number of inversions

$$N(\sigma) = \#\{x < y \mid \sigma(x) > \sigma(y)\}$$

Now, by multilinearity and antisymmetry of det we have

$$\begin{aligned} \det(\mathrm{id} + \mathbf{H}) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n \underbrace{(\mathrm{id} + \mathbf{H})_{i\sigma(i)}}_{\underbrace{(\mathrm{id})_{i\sigma(i)}}_{\mathbb{1}_{\sigma(i)=i}} + \mathbf{H}_{i\sigma(i)}} \\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n (\mathbb{1}_{\sigma(i)=i} + \mathbf{H}_{i\sigma(i)}) \end{aligned}$$

If $\sigma = \text{id}$

$$\prod_{i=1}^{n}(\mathbb{1}_{\sigma(i)=i} + \mathbf{H}_{i\sigma(i)}) = \prod_{i=1}^{n}(\mathbb{1} + \mathbf{H}_{ii}) = \sum_{E \subseteq \{1,\ldots,n\}} \prod_{i \in E} \mathbf{H}_{ii}$$

$$\implies \underbrace{\mathbb{1}}_{E=\varnothing} + \underbrace{\sum_{i=1}^{n} \mathbf{H}_{ii}}_{|E|=1} + \underbrace{o(\|\mathbf{H}\|^2)}_{|E|\geqslant 2}$$

Hence when $\sigma \neq \text{id} \Rightarrow \exists\, i \neq j$ s.t. $\sigma(i) \neq i$ and $\sigma(j) \neq j$. Hence

$$\prod_{k=1}^{n}(\underbrace{\mathbb{1}_{\sigma(k)=k}}_{0 \text{ for } k=j \text{ and } i} + \mathbf{H}_{k\sigma(k)}) = o(\|\mathbf{H}\|^2)$$

Hence

$$\begin{aligned}\det(\mathbf{M}) &= \underbrace{\text{sgn}(\text{id})}_{=\mathbb{1}}(\mathbb{1} + \text{trace}(\mathbf{H}) + o(\|\mathbf{H}\|^2)) \\ &= \det(\text{id}) + \text{trace}(\mathbf{H}) + o(\|\mathbf{H}\|^2) \\ &= \det(\text{id}) + \text{trace}(\mathbf{H}) + o(\|\mathbf{H}\|)\end{aligned}$$

Which is a coarser asymptotic. □

Intuition. This is the derivative. A linear approximation and a remainder.

Notation. When $F : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ we have $F = (f_1, \ldots, f_m)$

Theorem 2.4. Let $F : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with X open.

(F_1) $\forall\, a \in X \Rightarrow F$ is differentiable at $a \implies F$ is $\mathcal{C}^0$ at a

(F_2) F constant $\implies F$ is differentiable with derivative zero.

(F_3) F linear $\implies F$ is differentiable and $\forall\, a \in X \Rightarrow DF_a = F$

(F_4) F, G differentiable at $a \implies F + G$ differentiable at a

(F_5) F differentiable $\Leftrightarrow f_i$ differentiable at $a\, \forall\, i \in \{1, \ldots, m\}$

Proof. Let us prove Theorem 2.4

(F_1) Let us remember that $F(a+h) = F(a) + DF_a(h) + \varepsilon(h)$ with

$$\frac{\varepsilon(h)}{\|h\|} \xrightarrow[h\to 0]{} 0 \qquad \text{hence} \qquad \varepsilon(h) \xrightarrow[h\to 0]{} 0$$

Now $L(h)$ is linear, and because of finite dimensions, it is continuous

$$L(h) \xrightarrow[h\to 0]{} L(0) = 0 \qquad \text{thus} \qquad F(a+h) \xrightarrow[h\to 0]{} F(a)$$

which makes F continuous at a

(F_2) Notice the following

$$\begin{aligned}F(x+h) &= F(x) = c \\ &= F(x) + L(h) + 0 = c + L(h)\end{aligned}$$

Which means $\forall\, h \Rightarrow L(h) = 0$. Hence F is differentiable and $\forall\, x \Rightarrow DF_x = 0$

(F_3) The proof is the same as (F_2)

(F_4) The proof is allegedly very simple

(F_5) For the norm of $\mathbb{R}^n$ choose

$$\|x\|_\infty = \max_{1 \leqslant i \leqslant n} |x_i|$$

$\Longrightarrow$ Suppose F is differentiable at a and let $A = DF_a$ with $A = (\ell_1 \cdots \ell_m)$.

Now notice the following

$$\big\|F(x+h) - F(x) - A(h)\big\| = \max_{i \leqslant 1 \leqslant n} \big|f_i(x+h) - f_i(x) - \ell_i(h)\big|$$

By assumption of F

$$\frac{\big\|F(x+h) - F(x) - A(h)\big\|}{\|h\|} \to 0$$

Hence $\forall\, i \in \{1, ..., m\}$ we have that

$$\frac{\big|f_i(x+h) - f_i(x) - \ell_i(h)\big|}{\|h\|} \to 0$$

Hence f_i is differentiable with derivative ℓ_i

$\Longleftarrow$ is the same proof as the necessity.

$\therefore$ Theorem 2.4 is true. $\square$

Notation. For a square matrix $\mathbf{M} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and an integer $p \geqslant 1$, we define

$$\mathbf{M}^p = \underbrace{\mathbf{M}\,\mathbf{M}\cdots\mathbf{M}}_{p\text{ times}}.$$

Definition 2.7. A norm $\|\cdot\|$ on $\mathcal{M}_{n\times n}(\mathbb{R})$ is sub-multiplicative if

$$\forall\, \mathbf{A}, \mathbf{B} \in \mathcal{M}_{n\times n}(\mathbb{R}) \Rightarrow \|\mathbf{AB}\| \leqslant \|\mathbf{A}\|\,\|\mathbf{B}\|$$

Example 2.5. Let $p \geqslant 1$ and $n \geqslant 1$. Let $f : \mathcal{M}_{n\times n}(\mathbb{R}) \longrightarrow \mathcal{M}_{n\times n}(\mathbb{R})$ with $\mathbf{M} \longmapsto \mathbf{M}^p$

Proof. Let us show that

$$(\mathbf{M}+\mathbf{H})^p = \mathbf{M}^p + L_{\mathbf{M}}(\mathbf{H}) + o(\|\mathbf{H}\|)$$

We start by expanding

$$\begin{aligned}(\mathbf{M}+\mathbf{H})^p &= \mathbf{M}^p + \sum_{k=0}^{p-1} \mathbf{M}^k\mathbf{H}(\mathbf{M}+\mathbf{H})^{p-1-k} \\ &= \mathbf{M}^p + \underbrace{\sum_{k=0}^{p-1} \mathbf{M}^k\mathbf{H}\mathbf{M}^{p-1-k}}_{L_{\mathbf{M}}(\mathbf{H})} + \underbrace{R(\mathbf{H})}_{\text{Remainder}}\end{aligned}$$

Which means we need to show the following is true

$$R(\mathbf{H}) = (\mathbf{M}+\mathbf{H})^p - \mathbf{M}^p - L_{\mathbf{M}}(\mathbf{H}) = \sum_{k=0}^{p-1} \mathbf{M}^k\,\mathbf{H}\Big((\mathbf{M}+\mathbf{H})^{\,p-1-k} - \mathbf{M}^{\,p-1-k}\Big) = o(\|\mathbf{H}\|)$$

Fixing $q \geqslant 1$ we expand this take norm from Definition 2.4 that satisfies Definition 2.7

$$\left\|(\mathbf{M}+\mathbf{H})^q - \mathbf{M}^q\right\| \leqslant C_q \sum_{j=1}^{q} \|\mathbf{H}\|^j \Rightarrow \left\|R(\mathbf{H})\right\| \leqslant C\|\mathbf{H}\|^2$$

Notice that

$$\frac{\|(\mathbf{M}+\mathbf{H})^p - \mathbf{M}^p - L_{\mathbf{M}}(\mathbf{H})\|}{\|\mathbf{H}\|} \leqslant C\|\mathbf{H}\| \xrightarrow[\|\mathbf{H}\|\to 0]{} 0$$

Thus f is differentiable. □

Definition 2.8. The inverse image of $B \subseteq Y$ under the function $f : X \to Y$ is the set

$$f^{\leftarrow}[B] := \{x \in X \mid f(x) \in B\}$$

Theorem 2.5. Let (X, d), (Y, ρ) and $f : X \to X$. The following are equivalent:

(o_1) f is continuous

(o_2) $\forall\, W$ open in $(Y, \rho) \Rightarrow f^{\leftarrow}[W]$ is open in (X, d)

(o_3) $\forall\, \mathcal{F}$ closed in $(Y, \rho) \Rightarrow f^{\leftarrow}[\mathcal{F}]$ is closed in (X, d)

Proof. $(o_1) \implies (o_2)$ Let W be any open subset of Y. Let $x \in f^{\leftarrow}[W]$. Since W is open, $\exists\, \varepsilon > 0$ s.t. $B_\rho(f(x), \varepsilon) \subseteq W$. Since we assumed f is continuous at x,

$$\exists\, \delta > 0 \text{ s.t } \forall\, z \in X,\ d(x, z) < \delta \Rightarrow \rho(f(x), f(z)) < \varepsilon$$

Observe that $B_d(x, \delta) \subseteq f^{\leftarrow}[W]$. Take $z \in B(x, \delta)$, which implies

$$f(z) \in B_\rho(f(x), \varepsilon) \subseteq W$$

$(o_2) \implies (o_3)$. Suppose $F \subseteq Y$ is any closed set. Then $Y \setminus F$ is open in Y. By the previous implication, $f^{\leftarrow}[Y \setminus F]$ is open in X. Moreover, note that

$$f^{\leftarrow}[Y \setminus F] = f^{\leftarrow}[Y] \setminus f^{\leftarrow}[F] = X \setminus f^{\leftarrow}[F]$$

But this set is open, so its complement $\therefore f^{\leftarrow}[F]$ is closed in X.

$(o_3) \implies (o_1)$ Suppose $x \in X$ is any element and $\varepsilon > 0$ is arbitrary. Consider $B(f(x), \varepsilon)$, which is open in Y, since all balls are open. Then $Y \setminus B(f(x), \varepsilon)$ is closed in Y. By (o_2), $f^{\leftarrow}[Y \setminus B(f(x), \varepsilon)]$ is closed in X, and moreover,

$$f^{\leftarrow}[Y \setminus B(f(x), \varepsilon)] = f^{\leftarrow}[Y] \setminus f^{\leftarrow}[B(f(x), \varepsilon)] = X \setminus f^{\leftarrow}[B(f(x), \varepsilon)]$$

$\Rightarrow f^{\leftarrow}[B(f(x), \varepsilon)]$ is open in $X \Rightarrow x \in f^{\leftarrow}[B(f(x), \varepsilon)]$, and since it is open

$$\exists\, \delta > 0 \ \text{ s.t. } \ B(x, \delta) \subseteq f^{\leftarrow}[B(f(x), \varepsilon)]$$

It follows that

$$f[B(x, \delta)] \subseteq f^{\leftarrow}[B(f(x), \varepsilon)]$$

Indeed, suppose $z \in f[B(x, \delta)]$ is arbitrary.

$$\Rightarrow\ \exists\, y \in B(x, \delta) \ \text{ s.t. } \ z = f(y) \Rightarrow y \in f^{\leftarrow}[B(f(x), \varepsilon)]$$

Then $f(y) \in B(f(x), \varepsilon)$, but $f(y) = z$, which proves the inclusion

$$f[B(x, \delta)] \subseteq f^{\leftarrow}[B(f(x), \varepsilon)]$$

$\therefore f$ is continuous at x, and it follows that f is continuous on all of X. □

Definition 2.9. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. We define the norm of $x \in V$ as the number $\in \mathbb{R}$

$$\|x\| := \sqrt{\langle x, x\rangle}$$

Note that $\|\cdot\| : V \to \mathbb{R}$ is a function.

Lemma 2.1. The set of invertible matrices $\mathcal{G}_{n\times n}(\mathbb{R}) \subseteq \mathcal{M}_{n\times n}(\mathbb{R})$ and is open.

Proof. Remember the set of invertible matrices is

$$\mathcal{G}_{n\times n}(\mathbb{R}) = \{\mathbf{M} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \det(\mathbf{M}) \neq 0\}$$

We know det is continuous and $\mathbb{R} \setminus \{0\}$ to be open on $\mathbb{R}$.

$$\mathcal{G}_{n\times n}(\mathbb{R}) = \det^{\leftarrow}(\mathbb{R} \setminus \{0\})$$

By (o_2) of Theorem 2.5 $\Rightarrow \mathcal{G}_{n\times n}(\mathbb{R})$ is open since it is the inverse image of an open set. □

Example 2.6. Let $g : \mathcal{G}_{n\times n}(\mathbb{R}) \to \mathcal{G}_{n\times n}(\mathbb{R})$ with $\mathbf{M} \longmapsto \mathbf{M}^{-1}$

Proof. We know from Lemma 2.1 that $\mathcal{G}_{n\times n}(\mathbb{R})$ is open in $\mathcal{M}_{n\times n}(\mathbb{R})$. Now

$$\mathbf{M} \mapsto \mathbf{M}^{-1}$$

is a rational function, so it is differentiable. We have

$$(\mathbf{X} + \mathbf{H})^{-1} = \mathbf{X}^{-1}(\mathbf{I} + \mathbf{H}\mathbf{X}^{-1})^{-1}$$

Then we have $u = \mathbf{H}\mathbf{X}^{-1}$ with $\|\mathbf{u}\| < 1$ and

$$(\mathbf{I} + \mathbf{u})^{-1} = \sum_{n=0}^{\infty} (-1)^n \mathbf{u}^n$$

$$\implies (\mathbf{X} + \mathbf{H})^{-1} = \mathbf{X}^{-1} \underbrace{-\mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}}_{L_{\mathbf{X}(\mathbf{H})}} + o(\|\mathbf{H}\|^2)$$

Which is the linear map we need from Definition 2.6

$\therefore Dg_{(\mathbf{X})}(\mathbf{H}) = -\mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}$ is the derivative of this function. □

Definition 2.10. For $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, we define

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

the diagonal matrix in $\mathcal{M}_n(\mathbb{R})$ whose diagonal entries are $\lambda_1, \ldots, \lambda_n$

Definition 2.11. A matrix $\mathbf{M} \in \mathcal{M}_{n\times n}(\mathbb{R})$ is said to be diagonalizable if there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and a diagonal matrix $\mathbf{D} \in \mathcal{M}_{n\times n}(\mathbb{R})$ such that

$$\mathbf{M} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

where $\mathbf{D} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$

Definition 2.12 (Partial Derivative)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open. We call partial derivative of f at $a \in X$ w.r.t. x_i the limit

$$\lim_{h\to 0} \frac{f(a+h_{e_i}) - f(a)}{h}$$

whenever it exists, and is denoted by

$$\frac{\partial f}{\partial x_i}(a)$$

Remark.

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \xleftarrow{i\text{-th entry}}$$

Definition 2.13 (Directional Derivative)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open. Let $v \in \mathbb{R}^n$. We call directional derivative of f at $a \in X$ along v the limit

$$\lim_{h\to 0} \frac{f(a+hv) - f(a)}{h}$$

whenever it exists.

Lemma 2.2. Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open. Suppose f is differentiable at $a \implies$ the directional derivative exists and $\forall\, v \in \mathbb{R}^{\ltimes}$

$$\lim_{h\to 0} \frac{f(a+hv) - f(a)}{h} = Df_a(v)$$

In particular $\forall\, \{1, ..., n\} \Rightarrow$

$$\frac{\partial f}{\partial x_i}(a) = Df_a(e_i)$$

Proof. We have $f(a + \underbrace{tv}_{h}) = f(a) + Df_a(h) + \varepsilon(h)$ where

$$\frac{\varepsilon(h)}{\|h\|} \to 0 \qquad \text{and} \qquad \|h\| = |t|\|v\|$$

$$\frac{\varepsilon(h)}{|t|} \xrightarrow[t\to 0]{} 0 \implies \varepsilon(h) = o(t) \Rightarrow f(a+tv) = f(a) + tDf_a(v) + o(t)$$

Hence

$$\frac{f(a+tv) - f(a)}{t} = Df_a(v) + \frac{o(t)}{t} \xrightarrow[t\to 0]{} Df_a(v)$$

So the directional derivative exists and equals $Df_a(v)$ □

Example 2.7. Let $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

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

f has directional derivatives in every direction at $(0,0)$, yet is not continuous at this point.

Proof. Let $v = (v_1, v_2) \in \mathbb{R}^2$. Consider from Definition 2.13 the directional derivative

$$\lim_{h\to 0} \frac{f(0 + hv_1, 0 + hv_2) - f(0,0)}{h}$$

The first case is $v_1 \neq 0$

$$\frac{f(hv_1, hv_2) - f(0,0)}{h} = \frac{\frac{(hv_2)^2}{hv_1} - 0}{h} = \frac{\frac{h^2 v_2^2}{hv_1}}{h} = \frac{hv_2^2}{hv_1} = \frac{v_2^2}{v_1}$$

Second case is $v_1 = 0$

$$\frac{f(0, hv_2) - f(0,0)}{h} = \frac{hv_2 - 0}{h} = v_2$$

So the directional derivatives are

$$\begin{cases} \frac{v_2^2}{v_1}, & v_1 \neq 0 \\ v_2, & v_1 = 0 \end{cases}$$

Now suppose f were continuous at 0. Then for $\varepsilon = \frac{1}{2}\ \exists\, \delta$ s.t.

$$\sqrt{x^2 + y^2} < \delta \implies |f(x,y) - f(0,0)| < \frac{1}{2}$$

Consider $y = x^{1/2}$ which means $f(x, x^{1/2}) = 1$ so

$$\sqrt{x^2 + y^2} = \sqrt{x^2 + x} < \delta \Rightarrow |f(x, x^{1/2}) - f(0,0)| = |1 - 0| = 1 \not< \frac{1}{2}$$

Which means f is not continuous at $(0,0)$ while having directional derivatives. $\square$

Corollary. Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with X open. Let $a \in X$. Suppose f is differentiable at a

$$\mathcal{J}_a(f) := \mathcal{M}_{B_c}(Df_a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

where $\mathcal{J}_a(f)$ is called the Jacobian matrix

Proof. Let $L : \mathbb{R}^n \to \mathbb{R}^m$. Now

$$\mathcal{M}_{B_c}(L) = \begin{pmatrix} L_1(e_1) & \cdots & L_1(e_n) \\ \vdots & \ddots & \vdots \\ L_m(e_1) & \cdots & L_m(e_n) \end{pmatrix}$$

Therefore $(\mathcal{M}_{B_c}(L))_{ij} = \langle L(e_j), e_i \rangle$

$$L = Df_a(e_j) = \begin{pmatrix} D(f_1)_a(e_j) \\ \vdots \\ D(f_m)_a(e_j) \end{pmatrix}$$

And by Lemma 2.2 we have that

$$\langle Df_a(e_j), e_i \rangle = D(f_i)_a(e_j) = \frac{\partial f_i}{\partial x_j}(a)$$

Thus $\mathcal{M}_{B_c}(Df_a)_{ij} = \mathcal{J}_a(f)_{ij} = \frac{\partial f_i}{\partial x_j}(a)$ $\square$

Notation. $\mathcal{M}_{B_c}(L)$ denotes the matrix of $L : \mathbb{R}^n \to \mathbb{R}^m$ in the canonical bases

Theorem 2.6. Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Let $a \in X$. Suppose the partials exist on $B(a, \varepsilon)$ for some $\varepsilon \in \mathbb{R}^+$ and suppose that $\forall\, i \in \{1, ..., m\}$ and $\frac{\partial f}{\partial x_i}$ is $\mathcal{C}^0$ at $a \Rightarrow f$ is differentiable at a

Proof. Let us consider the case $m = 2$ so $a = (a_1, a_2)$

$$\begin{aligned}f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) =& f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2)\\ &+ f(a_1, a_2 + h_2) + f(a_1, a_2)\end{aligned}$$

For $\|(h_1, h_2)\|$ small enough we have $x \mapsto f(x, a_2 + h_2)$ differentiable on $[a_1, a_1 + h_1]$.

By Theorem 2.3 $\exists\, c_1 \in (a_1, a_1 + h_1)$ s.t.

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2) = h_1 \frac{\partial f}{\partial x}(c_1, a_2 + h_2)$$

Similarly $\exists\, c_2 \in (a_2, a_2 + h_2)$ s.t.

$$f(a_1, a_2 + h_2) - f(a_1, a_2) = h_2 \frac{\partial f}{\partial y}(a_1, c_2)$$

Thus

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = h_1 \frac{\partial f}{\partial x}(c_1, a_2 + h_2) + h_2 \frac{\partial f}{\partial y}(a_1, c_2)$$

Since $(x, y) \mapsto \frac{\partial f}{\partial x}(x, y)$ is $\mathcal{C}^0$ at (a_1, a_2)

$$\frac{\partial f}{\partial x}(c_1, a_2 + h_2) = \frac{\partial f}{\partial x}(a_1, a_2) + \underbrace{o(1)}_{h \to 0}$$

Since $(x, y) \mapsto \frac{\partial f}{\partial y}(x, y)$ is $\mathcal{C}^0$ at (a_1, a_2)

$$\frac{\partial f}{\partial y}(a_1, c_2) = \frac{\partial f}{\partial y}(a_1, a_2) + \underbrace{o(1)}_{h \to 0}$$

Thus

$$\begin{aligned}f(a_1 + h_1, a_2 + h_2) - &f(a_1, a_2) =\\ &h_1 \frac{\partial f}{\partial x}(a_1, a_2) + h_2 \frac{\partial f}{\partial y}(a_1, a_2) + \underbrace{h_1 o(1) + h_2 o(1)}_{\varepsilon(h)}\end{aligned}$$

$$\frac{\varepsilon(h)}{\|h\|} \xrightarrow[h \to 0]{} 0 \qquad \text{and} \qquad \|h\| \geqslant \frac{1}{c} \max(|h_1|, |h_2|)$$

Thus

$$\begin{aligned}f(a_1 + h_1, a_2 + h_2) - &f(a_1, a_2) =\\ &h_1 \frac{\partial f}{\partial x}(a_1, a_2) + h_2 \frac{\partial f}{\partial y}(a_1, a_2) + o(\|h\|)\end{aligned}$$

$\therefore f$ is differentiable. $\varepsilon(h) = o(1)$ and $\frac{\varepsilon(h)}{1} \to 0$ $\square$

Remark. We deduce from Theorem 2.6 that polynomials are differentiable.

Example 2.8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ with $f : (x, y) \mapsto xy + x^2$

Proof. Let us look at the tangent plane to the graph of f at $(1,1)$

$$\frac{\partial f}{\partial x} = y + 2x \qquad \text{and} \qquad \frac{\partial f}{\partial y} = x$$

Which by Definition 2.12 are the relevant partial derivatives. At $(1,1)$

$$\frac{\partial f}{\partial x} = 1 + 2 \cdot 1 = 3 \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 1$$

We know the equation for a tangent plane at $(1,1)$ is

$$z - f(1,1) = \frac{\partial f}{\partial x}(x-1) + \frac{\partial f}{\partial y}(y-1)$$

$$z - 2 = 3(x-1) + (y-1) \Rightarrow z = 3x + y - 2$$

This is the tangent plane. □

Theorem 2.7. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at 0 and satisfy

$$\forall\, x \in \mathbb{R}^n \Rightarrow x \neq 0 \qquad \text{and} \qquad \forall\, t \in \mathbb{R}_+^* \Rightarrow f(tx) = tf(x)$$

$\Longrightarrow f$ is linear.

Proof. Since f is differentiable at 0 by Definition 2.6 $\exists\, L : \mathbb{R}^n \to \mathbb{R}$ s.t.

$$f(x) = f(0) + L(x) + o(\|x\|)$$

By our second assumption about this function we have that

$$\forall\, t > 0 \Rightarrow f(0) = f(t \cdot 0) = tf(0) = 0 \Longrightarrow f(x) = L(x) + o(\|x\|)$$

Using $f(tx) = tf(x)$

$$\frac{f(tx)}{t} = L(x) + \frac{o(t\|x\|)}{t} \xrightarrow[t \to 0]{} L(x) + o(\|x\|)$$

$\forall\, x \in \mathbb{R}^n \Rightarrow f(x) = L(x)$, so f is linear. □

Definition 2.14 (Gradient)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open. Suppose the partial derivatives exist at $a \in X$. The gradient of f at a is

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a) \right)$$

Corollary. If f is diff at $a \Rightarrow \forall\, h \in \mathbb{R}^n$

$$Df_a(h) = \langle \nabla f(a), h \rangle = \mathcal{J}_a(f)$$

Indeed $\mathcal{J}_a(f) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n} \right) \in \mathcal{M}_{1 \times n}(\mathbb{R})$

2.3 The Chain Rule

Intuition. We have $X \subseteq \mathbb{R}^n \xrightarrow{F} F(X) \subseteq Y \subseteq \mathbb{R}^m \xrightarrow{G} \mathbb{R}^k$. Which means $G \circ F : X \to \mathbb{R}^k$

Definition 2.15 (Lipschitz)**.** Let (X, d) and (Y, ρ) be any metric spaces. We say that a function $f : X \to Y$ is Lipschitz continuous if there exists a Lipschitz constant $c < 0$ such that

$$\forall\, x, z \in X \Rightarrow \rho(f(x).f(z)) \leqslant c \cdot d(x, z)$$

Corollary. Let (X, d) and (Y, ρ) be metric spaces. If $f : X \to Y$ is Lipschitz $\Rightarrow f$ is $\mathcal{C}^0$

Proof. Let $x_0 \in X$ be arbitrary. Let $\varepsilon \in \mathbb{R}^+$ be any. Define $\delta = \frac{\varepsilon}{c}$, where $c > 0$ is such that

$$\forall\, x, y \in X \Rightarrow \rho(f(x).f(y)) \leqslant c \cdot d(x, y)$$

It follows that

$$\forall\, x \in X \Rightarrow d(x, x_0) < \delta \Rightarrow \rho(f(x).f(x_0)) < \varepsilon$$

Indeed, let $x \in X$ s.t. $d(x, x_0) < \delta \Rightarrow$

$$\rho(f(x).f(x_0)) \leqslant c \cdot d(x, x_0) < c \cdot \delta = \varepsilon$$

$\therefore f$ is continuous at x_0. Since x_0 is arbitrary, it is continuous on the whole space. □

Remark. Linear maps are Lipschitz because all linear maps are continuous.

Theorem 2.8 (Chain Rule)**.** Let $X \subseteq \mathbb{R}^n$ be open, $Y \subseteq \mathbb{R}^m$ open, $F : X \to \mathbb{R}^m$ s.t. $F(X) \subseteq Y$ and $G : Y \to \mathbb{R}^k$. Let $a \in X$. Suppose F is differentiable at a and G is differentiable at $F(a) \Rightarrow G \circ F$ is differentiable at a and

$$D(G \circ F)_a = DG_{F(a)} \circ DF_a$$

Proof. We have $G(F(a + h)) = \cdots$. Since F is differentiable at a

$$F(a + h) + \underbrace{DF_a(h) + \varepsilon(h)}_{:=h'} \qquad \text{where } \frac{\left\|\varepsilon(h)\right\|}{\|h\|} \to 0$$

Since G is differentiable at $F(a)$

$$G(F(a + h)) = G(F(a) + h') = G(F(a) + DG_{F(a)}(h') + \tilde{\varepsilon}(h') \quad \text{where } \frac{\left\|\tilde{\varepsilon}(h')\right\|}{\|h'\|} \xrightarrow[h' \to 0]{} 0$$

$$DG_{F(a)}(h') = DG_{F(a)}(DF_a(h)) + DG_{F(a)}(\varepsilon(h))$$

Since $DG_{F(a)}$ is linear and finite dimensional it satisfies Definition 2.15 $\exists\, c > 0$ s.t.

$$\forall\, y \in \mathbb{R}^n \Rightarrow \left\|DG_{F(a)}(y)\right\| \leqslant c \cdot \|y\|$$

$$\Rightarrow \left\|DG_{F(a)}(\varepsilon(h))\right\| \leqslant c \cdot \left\|\varepsilon(h)\right\| \Rightarrow \frac{\left\|DG_{F(a)}(\varepsilon(h))\right\|}{\|h\|} \to 0$$

Thus $DG_{F(a)}(\varepsilon(h)) = o(\|h\|)$. Let us also show $\tilde{\varepsilon}(h') = o(\|h\|)$

$$h' = DF_a(h) + \varepsilon(h)$$

Using the Lipschitz property of DF_a we know there $\exists\, c > 0$ s.t. $\left\|h'\right\| \leqslant c \cdot \|h\|$

$$\frac{\left\|\tilde{\varepsilon}(h')\right\|}{\|h\|} = \frac{\left\|\tilde{\varepsilon}(h')\right\|}{\|h\|} \cdot \frac{\left\|h'\right\|}{\|h'\|} \leqslant c \cdot \frac{\left\|\tilde{\varepsilon}(h')\right\|}{\|h\|} \xrightarrow[h \to 0]{} 0$$

Thus $\tilde{\varepsilon}(h') = o(\|h\|)$ and hence

$$G \circ F(x+h) = G(F(x)) + DG_{F(x)} \circ DF_x(h) + o(\|h\|)$$

which satisfies Definition 2.6 □

Remark. Let $L : \mathbb{R}^n \to \mathbb{R}^k$ and $\|x\|_\infty = \max|x_i|$ so $x = x_1e_1 + \cdots x_ne_n$

$$L(x) = x_1L(e_1) + \cdots + x_nL(e_n)$$

$$\|L(x)\|_\infty \leqslant \|x\|_\infty \underbrace{\sum_{i=1}^n \|L(e_i)\|_\infty}_{=c}$$

Theorem 2.9. Let $\langle\cdot\rangle$ be a scalar product in $\mathbb{R}^n$ and $\|\cdot\|$ the associated norm by Definition 2.9

(d_1) $\|\cdot\|$ is differentiable on $\mathbb{R}^n \setminus \{0\}$

(d_2) $\|\cdot\|$ is not differentiable on 0

Proof. Let us proceed with the proof.

(d_1) We can write the norm from Definition 2.9 as the composition $\|\cdot\| = G \circ F$

$$F : \mathbb{R}^n \to \mathbb{R} \text{ with } F : x \mapsto \langle x, x\rangle \quad \text{and} \quad G : (0,\infty) \to \mathbb{R} \text{ with } G : x \mapsto \sqrt{x}$$

Both of which are differentiable

$$\forall\, a \in \mathbb{R}^n \Rightarrow DF_a(h) = 2\langle a, h\rangle \quad \text{and} \quad \forall\, a \in \mathbb{R}^n \Rightarrow DG_{F(a)}(s) = \frac{1}{2\sqrt{t}}s$$

By Theorem 2.8 we have $d(\|\cdot\|)_a = DG_{F(a)} \circ DF_a$ hence

$$d(\|\cdot\|)_a(h) = DG_{\|a\|^2}\big(DF_a(h)\big) = \frac{1}{2\|a\|} \cdot 2\langle a, h\rangle = \frac{\langle a, h\rangle}{\|a\|}$$

$d(\|\cdot\|)_a(h) = \frac{\langle a,h\rangle}{\|a\|}$ for $a \neq 0$

(d_2) Suppose by contradiction, $\|\cdot\|$ satisfies Definition 2.6 and take $a = 0$

$$\|x\| = \|0\| + L(x-0) + o_0(x) = L(x) + o_0(x)$$

Since L is linear $v \in \mathbb{R}^n$ s.t. $L(x) = \langle v, x\rangle$. Suppose $\|u\| = 1$ and $x = tu$ with $t \to 0$

$$\frac{|\|tu\| - L(tu)|}{\|tu\|} = \frac{||t| - t\langle v, u\rangle|}{|t|} = |1 - \operatorname{sgn}(t)\langle v, u\rangle|$$

$$\underbrace{|1 - \langle v, u\rangle| = 0 \Rightarrow \langle v, u\rangle = 1}_{t>0} \quad \text{and} \quad \underbrace{|1 + \langle v, u\rangle| = 0 \Rightarrow \langle v, u\rangle = -1}_{t<0}$$

$\Longrightarrow\!\Longleftarrow$ since $\langle v, u\rangle$ can't have two different values.

$\therefore$ Theorem 2.9 is true. □

Intuition. We now will prove that Definition 2.14 is orthogonal to the level sets.

Theorem 2.10. Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open, f differentiable, and $x \in X$. Suppose $\nabla f(x) \neq 0 \Rightarrow \nabla f(x)$ points in direction of sharpest increase of f.

Proof. Let $v \in \mathbb{R}^+$ s.t. $\|v\| = 1$

$$\Rightarrow \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} = Df_x(v) = \langle \nabla f(x), v \rangle$$

We can then observe that for

$$v = \frac{\nabla f(x)}{\|\nabla f(x)\|} \Rightarrow \sup_{\|v\|=1} \langle \nabla f(x), v \rangle \text{ is attained}$$

which means v is in direction of the gradient. $\square$

Definition 2.16 (Level Set). Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\alpha \in \mathbb{R}$. The level set of f at α is

$$S_\alpha := \{x \in \mathbb{R}^n \mid f(x) = \alpha\}.$$

Theorem 2.11. Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable and $\alpha \in \mathbb{R}$. Set S_α to be the level set. Also suppose $\nabla f(x) \neq 0 \implies \nabla f(x) \perp S_\alpha$ at $x \in S_\alpha$. Meaning $\forall\, \gamma : (-\varepsilon, \varepsilon) \to S_\alpha$ that is differentiable s.t. $\gamma(0)$ we have

$$\langle \gamma'(0), \nabla f(x) \rangle = 0$$

Proof. Since $\forall\, t \in (-\varepsilon, \varepsilon) \Rightarrow \gamma(t) = S_\alpha \Rightarrow f(\gamma(t)) = \alpha$. We know $p(t) := f(\gamma(t)) = \alpha$ is differentiable. By Theorem 2.8

$$p'(t) = \langle \nabla f(\gamma(t)), \gamma'(t) \rangle = 0$$

At $t = 0 \implies \langle \nabla f(x), \gamma'(0) \rangle = 0$ $\square$

Notation. Remember $\|\cdot\|_2$ is the Euclidean norm.

Theorem 2.12. Let $r > 0$ and set $S_r := \{x \in \mathbb{R}^n \mid \|x\|_2 = r\}$. Then $\forall\, x \in S_r \implies x \perp S_r$

Proof. Set $f := \|x\|_2$ and notice it is differentiable by Theorem 2.9

$$\Rightarrow \nabla f(x) = \frac{x}{\|x\|_2}$$

Take $x \in S_r \Rightarrow \nabla f(x) = \frac{x}{r}$. But we know from Theorem 2.11 that is we take $\gamma : (-\varepsilon, \varepsilon) \to S_r$

$$\langle \gamma'(0), \nabla f(x) \rangle = 0 = \langle \gamma'(0), \frac{x}{r} \rangle$$

$\therefore x \perp S_r$ since multiplying by r preserves orthogonality. $\square$

Theorem 2.13. Let $u \neq 0 \in \mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ differentiable s.t. $\forall\, x \in \mathbb{R}^n\ \exists\, \lambda_x \in \mathbb{R}$ s.t. $\nabla f(x) = \lambda_x u$. Show that $\exists\, \varphi : \mathbb{R} \to \mathbb{R}$ s.t. $\forall\, x \in \mathbb{R}^n \Rightarrow f(x) = \varphi(\langle x, u \rangle)$

Proof. Let $\alpha \in \mathbb{R}$ and $H := \{x \in \mathbb{R}^n \mid \langle x, u \rangle = \alpha\}$. Let $x, y \in H$, which is the affine hyperplane, and we can link through a path $\gamma(t) = (1-t)x + ty$

$$\Rightarrow\ \forall\, t \in (0,1) \Rightarrow \gamma'(t) = y - x$$

Set $p : [0,1] \to \mathbb{R}$ with $p = t \mapsto f(\gamma(t))$, notice p is differentiable

$$\Rightarrow\ \forall\, t \in (0,1) \Rightarrow p'(t) = \langle \nabla f(\gamma(t)), \gamma'(t) \rangle = \langle \nabla f(\gamma(t)), y - x \rangle$$

By assumption $\nabla f(\gamma(t))$ is proportional to $\gamma(t)$ which is orthogonal to H

$$\Rightarrow p'(t) = \langle \lambda_x u, y - x\rangle = \lambda_x \langle u, y - x\rangle = 0$$

$\therefore f(x) = f(y)$ which means f is constant $\Rightarrow$ affine hyperplane is a level set. $\square$

Theorem 2.14 (MVT Reloaded)**.** Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open. If f is $\mathcal{C}^1$ on the segment $[a, b]$ and $\exists\, M \in \mathbb{R}^+$ s.t $\forall\, c \in (a, b) \Rightarrow \|Df_c\| \leqslant M \Longrightarrow$

$$\|f(b) - f(a)\| \leq M\|b - a\|$$

Proof. Let $\varepsilon > 0$. Consider the following set.

$$S = \{t \in [a, b] \mid \big\|f(t) - f(a)\big\| \leqslant M(t - a) + \varepsilon(t - a) + \varepsilon\}$$

Since f is $\mathcal{C}^0 \Rightarrow$ s.t.

$$\exists\, \delta > 0 \text{ s.t. } \;\forall\, s \in [a, a + \delta] \Longrightarrow \big\|f(s) - f(a)\big\| \leqslant \varepsilon$$

Hence $a + \delta \in S$. Let $c := \sup S$. Note $c \in S \Rightarrow a + \delta \leqslant c \leqslant b$ Suppose, by contradiction, that $c < b \Rightarrow f$ is differentiable in c and

$$\exists\, \delta_0 \in (0, \min\{c - a, b - c\}) \text{ s.t if} |s - c| < \delta_0 \Longrightarrow \big\|f(s) - f(c) - Df_c(s - c)\big\| < \varepsilon|s - c|$$

By assumption, if $s \in (c, c + \delta_0)$

$$\begin{aligned}
\big\|f(s) - f(a)\big\| &\leqslant \big\|f(s) - f(c)\big\| + \big\|f(c) - f(a)\big\| \\
&\leqslant \big\|f(s) - f(c) - Df_c(s - c)\big\| + \big\|Df_c(s - c)\big\| + \big\|f(c) - f(a)\big\| \\
&< \varepsilon(s - c) + (s - c)\big\|Df_c(c)\big\| + M(c - a) + \varepsilon(c - a) + \varepsilon \\
&\leqslant M(s - a) + \varepsilon(s - a) + \varepsilon
\end{aligned}$$

But this shows $c \neq \sup S \Rightarrow c = b$ which implies

$$\big\|f(b) - f(a)\big\| \leqslant M(b - a) + \varepsilon(b - a) + \varepsilon \Longrightarrow \big\|f(b) - f(a)\big\| \leqslant M(b - a)$$

Hence Theorem 2.14 is true. $\square$

2.4 Clairaut Theorem

Definition 2.17 (Second Derivative)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$. We call

$$\forall\, 1 \leqslant i, j \leqslant n \Rightarrow \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

the second derivative of f whenever f has differentiable first partial derivatives.

Definition 2.18 ($\mathcal{C}^1$ and $\mathcal{C}^2$)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$

($\mathcal{C}_1$) We say f is $\mathcal{C}^1$ if f is differentiable with continuous partials i.e. if

$$\forall\, i \in \{1, \cdots, n\} \Rightarrow \frac{\partial f}{\partial x_i} \text{ is } \mathcal{C}^0$$

($\mathcal{C}_2$) We say f is $\mathcal{C}^2$ if

$$\forall\, i \in \{1, \cdots, n\} \Rightarrow \frac{\partial^2 f}{\partial x_j x_i} \text{ is } \mathcal{C}^0$$

Theorem 2.15 (Clairaut)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open be $\mathcal{C}^2 \implies$

$$\forall\, i, j \in \{1, \cdots, n\} \Rightarrow \frac{\partial^2 f}{\partial x_i x_j} = \frac{\partial^2 f}{\partial x_j x_i}$$

Proof. We take $n = 2$ to simplify notation so $\mathbb{R}^n = \mathbb{R}^2$. Let $(a_1, a_2) \in X$. Notice that

$$\frac{\partial f}{\partial x_1}(a_1, a_2) = \lim_{h_1 \to 0} \frac{f(a_1 + h_1, a_2) - f(a_1, a_2)}{h_1}$$

Now set $S = f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2) - f(a_1 + h_1, a_2) + f(a_1, a_2)$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(a_1, a_2) = \lim_{h_2 \to 0} \lim_{h_1 \to 0} \frac{S(h_1, h_2)}{h_1 h_2}$$

Similarly

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(a_1, a_2) = \lim_{h_2 \to 0} \lim_{h_1 \to 0} \frac{S(h_1, h_2)}{h_1 h_2}$$

Set $g : x \mapsto f(a_1 + h_1, x) - f(a_1, x)$. Notice that $S(h_1, h_2) = g(a_2 + h_2) - g(a_2)$. We can apply Theorem 2.14 to g that is $\mathcal{C}^1$ so $\exists\, c_2 \in (a_2, a_2 + h_2)$ s.t.

$$\begin{aligned} S(h_1, h_2) &= h_2 g'(c_2) \\ &= h_2 \left(\frac{\partial f}{\partial x_2}(a_1 + h_1, c_2) - \frac{\partial f}{\partial x_2}(a_1, c_2) \right) \end{aligned}$$

Set $h : x \mapsto \frac{\partial f}{\partial x_2}(x, c_2)$ and apply Theorem 2.14 to h so $\exists\, c_1 \in (a_1, a_1 + h_1)$ s.t.

$$S(h_1, h_2) = h_1 h_2 \frac{\partial f}{\partial x_1 \partial x_2}(c_1, c_2)$$

Since $c_1 \in (a_1, a_1 + h_1)$ and $c_2 \in (a_2, a_2 + h_2)$ and f is $\mathcal{C}^2 \Rightarrow$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(a_1, a_2) \implies \lim_{(h_1, h_2) \to (0,0)} \frac{S(h_1, h_2)}{h_1 h_2} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(a_1, a_2)$$

Proceeding similarly setting $k : x \mapsto f(x, a_2 + h_2) - f(x, a_2)$

$$\lim_{(h_1, h_2) \to (0,0)} \frac{S(h_1, h_2)}{h_1 h_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}(a_1, a_2)$$

Thus

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(a_1, a_2) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(a_1, a_2)$$

$\therefore$ Theorem 2.15 is true. $\square$

Definition 2.19 (Hessian Matrix)**.** Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open be twice differentiable at $x \in X$. The Hessian of f at x is defined as

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} \in \mathcal{M}_{n \times n}(\mathbb{R})$$

Corollary. If f is $\mathcal{C}^2 \Rightarrow$ by Theorem 2.15 we have that $\forall\, x \in X \implies \nabla^2 f(x)$ is symmetric.

Definition 2.20. We say $x \in X$ is a critical point of f if $\nabla f(x) = 0$

Lemma 2.3. Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open. Suppose f is differentiable. Suppose $x \in X$ is a local minmax of $f \implies \nabla f(x) = 0$

Proof. Let $v \in \mathbb{R}^n$ ang $g : t \mapsto f(x + tv)$. By assumption g has a local minmax at 0 since g is differentiable $\implies g'(0) = 0$. Note that

$$g'(0) = \langle \nabla f(x), v\rangle$$

Hence $\forall\, v \in \mathbb{R}^n \implies \langle \nabla f(x), v\rangle = 0$. Thus $\nabla f(x) = 0$ □

Notation. Let $I \subseteq \mathbb{R}$ open $\Rightarrow \mathcal{C}^k(I, \mathbb{R}) = \{f : I \to \mathbb{R} \mid f^{(k)}$ exists and is $\mathcal{C}^0\}$

Theorem 2.16 (Taylor). Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be $\mathcal{C}^k$ on an open interval I containing $x_0 \implies \forall\, x \in I\, \exists\, \xi$ between x_0 and x s.t.

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}$$

where the last term is the Lagrange remainder.

Intuition. We have $\nabla f(x) = 0$. When can we asses $\nabla f(x) = 0$ is minmax?

Theorem 2.17 (Taylor Expansion). Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open. Let $v \in \mathbb{R}^n$

$$\implies f(x + tv) = f(x) + t\langle \nabla f(x), v\rangle + \frac{t^2}{2}\langle v, \nabla^2 f(x)v\rangle + o(t^2)$$

Proof. Set $g(t) = f(x + tv)$. Note f and g are $\mathcal{C}^2$ and by Theorem 2.16 we have

$$g(t) = g(0) + t\underbrace{g'(0)}_{=\langle \nabla f(x), v\rangle} + \frac{t^2}{2}\underbrace{g''(0)}_{=\langle v, \nabla^2 f(x)v\rangle} + o(t^2)$$

By Theorem 2.8 $g'(t) = \langle \nabla f(x + tv), v\rangle$ so

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x + tv)v_i = \langle \nabla f(x + tv), v\rangle$$

$$\frac{d}{dt}\frac{\partial f}{\partial x_i}(x + tv) = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x + tv)v_j$$

Thus

$$g''(t) = \sum_{i,j} \frac{\partial^2 f}{\partial x_j \partial x_i}(x + tv)v_i v_j = \langle v, \nabla^2 f(x + tv)v\rangle = \underbrace{\langle v, \nabla^2 f(x)v\rangle}_{\text{at } t=0}$$

Which proves Theorem 2.17. □

Lemma 2.4 (Hessian Test). Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X open be s.t. $\nabla f(x) = 0$

(∇_1) $\forall\, v \in \mathbb{R}^n$, if x is a local min $\implies \langle v, \nabla^2 f(x)v\rangle \geqslant 0$

(∇_2) $\forall\, v \neq 0 \implies \langle v, \nabla^2 f(x)\rangle > 0$

Proof. Recall $f(x+tv) = f(x) + \frac{t^2}{2}\langle v, \nabla^2 f(x)v\rangle + o(t^2)$

(∇_1) Suppose by contradiction $\exists\, v \in \mathbb{R}^n$ s.t. $\langle v, \nabla^2 f(x)v\rangle \leqslant 0$

$$\implies f(x+tv) - f(x) = \frac{t^2}{2}\langle v, \nabla^2 f(x)v\rangle + o(t^2)$$

Hence for small enough $t \implies \langle v, \nabla^2 f(x)v\rangle + o(t^2) < 0$. Thus for t small enough $f(x+tv) - f(x) < 0 \implies$ not a local minimum. Hence $\langle v, \nabla^2 f(x)v\rangle \geqslant 0$

(∇_2) Consider $\inf_{\|v\|=1}\langle v, \nabla^2 f(x)v\rangle$. Notice that $v \mapsto \langle v, \nabla^2 f(x)v\rangle$ is $\mathcal{C}^0$ and $\{x \in \mathbb{R}^n \mid \|x\| = 1\}$ which is a unit sphere. This set is compact, hence $\exists\, v_0$ s.t. $\|v_0\| = 1 \implies$

$$\inf_{\|v\|=1}\langle v, \nabla^2 f(x)v\rangle = \langle v_0, \nabla^2 f(x)v_0\rangle$$

This v_0 is the minimum. Now define $c_0 := \inf_{\|v\|=1}\langle v, \nabla^2 f(x)v\rangle$. We deduce that

$$\forall\, v \text{ s.t. } \|v\| = 1 \implies \langle v, \nabla^2 f(x)v\rangle \geqslant c_0$$

Hence

$$\forall\, v \neq 0 \in \mathbb{R}^n \implies \left\langle \frac{v}{\|v\|}, \nabla^2 f(x)\frac{v}{\|v\|}\right\rangle \geqslant c_0$$

Thus $\forall\, v \neq 0 \implies \langle v, \nabla^2 f(x)v\rangle \geqslant c_0\|v\|^2$ and hence

$$f(x+tv) - f(x) \geqslant \frac{t^2}{2}c_0\|v\|^2 + o(t^2)$$

Since $\frac{o(t^2)}{t^2} \to 0$ as $t \to 0 \Rightarrow \exists\, \varepsilon_0 > 0$ depending only on c_0 s.t. $\forall\, v \in \mathbb{R}^n \Rightarrow \|v\| = 1$ and $\forall\, |t| \leqslant \varepsilon_0 \implies o(t^2) \geqslant -\frac{t^2}{4}c_0 \Rightarrow$

$$f(x+tv) - f(x) \geqslant \frac{t^2}{2}c_0 - \frac{t^2}{4}c_0 = \frac{t^2}{4}c_0 > 0$$

Indeed $\forall\, y \in B(x, \varepsilon_0) \implies f(y) - f(x) \geqslant \|x-y\|^2$

$$f(y) - f(x) \geqslant \frac{\|x-y\|^2}{4}c_0$$

Then x is a local minimum. Note $y = x + \frac{y-x}{\|y-x\|}\|y-x\|$

Thus Lemma 2.4 is true. □

Definition 2.21. Let $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$. A number $\lambda \in \mathbb{R}$ is called an eigenvalue of $\mathbf{A}$ if $\exists\, v \neq 0 \in \mathbb{R}^n$ s.t.

$$\mathbf{A}v = \lambda v$$

Then v is called an eigenvector associated to λ.

Remark. Let $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \Rightarrow \langle x, \mathbf{A}y\rangle = x^\top(\mathbf{A}y) = (\mathbf{A}^\top x)^\top y = \langle \mathbf{A}^\top x, y\rangle$

Lemma 2.5. Let $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $\mathbf{A} = \mathbf{A}^\top$

$$\implies \inf_{\|x\|=1}\langle x, \mathbf{A}x\rangle = \lambda_{\min}$$

where $\lambda_{\min}$ is the minimal eigenvalue of $\mathbf{A}$

Proof. By Definition 2.11 $\mathbf{A} = \mathbf{PDP}^{-1}$ and since $\mathbf{A}$ is symmetric $\mathbf{A} = \mathbf{PDP}^\top$

$$\langle x, \mathbf{A}x \rangle = \langle x, \mathbf{PD}(\mathbf{P}^\top x) \rangle = \langle \mathbf{P}^\top x, \mathbf{D}(\mathbf{P}^\top x) \rangle$$

Notice $\mathbf{P} = (v_1, v_2, \dots, v_n)$ are the eigenvectors. Now $y = \mathbf{P}^\top x$ and $\|y\| = \|x\| = 1$ since $\mathbf{P}^\top$ is orthogonal. Thus

$$\inf_{\|x\|=1} \langle x, \mathbf{A}x \rangle = \inf_{\|y\|=1} \langle y, \mathbf{D}y \rangle = \inf_{\|y\|=1} \left(\sum_{i=1}^{n} \lambda_i y_i^2 \right)$$

Clearly the infimum is for $y_1 = \pm 1, y_2 = 0, \cdots, y_n = 0$. Thus

$$\inf_{\|x\|=1} \langle x, \mathbf{A}x \rangle = \lambda_1$$

Where λ_1 is of course the minimal eigenvalue. □

Example 2.9. Let $f : (x, y) \mapsto e^x + xy$

Proof. Notice $\nabla f(x, y) = (e^x + y, x)$ and $\nabla f(x, y) = (0, 0) \Leftrightarrow x = 0$ and $y = -1$

$$\nabla^2 f(x, y) = \begin{pmatrix} e^x & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow \nabla^2 f(0, -1) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Notice now that

$$\mathbf{X}(\lambda) = \det \begin{pmatrix} \lambda - 1 & -1 \\ -1 & \lambda \end{pmatrix} = (\lambda - 1)\lambda - 1 = \lambda^2 - \lambda - 1$$

Notice $\lambda_1 < 0 < \lambda_2$. Hence $(0, -1)$ is neither a local max or min: it is a saddle point. □

2.5 Inverse Function Theorem

Definition 2.22. Let (X, d) be a metric space. We say that $f : X \to X$ is a contraction if

$$\forall\, x, y \in X \implies d(f(x), f(y)) \leqslant d(x, y)$$

We say f is a strict contraction if

$$\exists\, c \in (0, 1) \text{ s.t. } \forall\, x, y \in X \implies d(f(x), f(y)) \leqslant c \cdot d(x, y)$$

Definition 2.23. Let $f : X \to X$. We say that $x \in X$ is a fixed point of f if $f(x) = x$

Theorem 2.18 (Picard Fixed Point)**.** Let (X, d) be a complete metric space. Let f be a strict contraction $\implies f$ has a unique fixed point.

Proof. Let us begin with uniqueness. Suppose by contradiction $f(x) = x$ and $f(y) = y$. By Definition 2.23 we have that

$$\underbrace{d(f(x), f(y))}_{=d(x,y)} \leqslant \underbrace{c}_{\in(0,1)} \cdot d(x, y) \implies d(x, y) = 0$$

And by (d_2) of Definition 1.1 $\Rightarrow d(x, y) = 0 \Leftrightarrow x = y$.

Now we have to show existance. Let $u_0 \in X \Rightarrow \forall\, n \geqslant 0$ we set $u_{n+1} = f(u_n)$

$$\implies d(u_{n+1}, u_n) = d(f(u_n), f(u_{n+1})) \leqslant c \cdot d(u_n, u_{n+1})$$

Iterating gives $d(u_{n+1}, u_n) \leqslant c^n d(u_1, u_0)$ by induction. Let $p, q \geqslant n_0$ with $q \geqslant p$

$$\Longrightarrow d(u_p, u_q) \leqslant \sum_{k=p}^{q-1} d(u_{k+1}, u_k) \leqslant \left(\sum_{k=p}^{q-1} c^k\right) d(u_1, u_0)$$
$$\leqslant \left(\sum_{k=n_0}^{+\infty} c^k\right) d(u_1, u_0) = \frac{c^{n_0}}{1-c} d(u_1, u_0)$$

Fix $\varepsilon > 0 \Rightarrow \exists\, n_o \in \mathbb{N}$ s.t. $\frac{c^{n_0}}{1-c} d(u_1, u_0) < \varepsilon$

$$\Rightarrow \forall\, p, q \geqslant n_0 \Longrightarrow d(u_p, u_q) \leqslant \varepsilon$$

This means that (u_n) is a Cauchy sequence, so by Definition 1.12 (u_n) converges

$$\Longrightarrow \exists\, \ell \in X \text{ s.t. } u_n \to \ell$$

Notice f is $\mathcal{C}^0$ since it is a contraction. If $d(x_n, x) \to 0$

$$\Longrightarrow d(f(x_n), f(x)) \leqslant d(x_n, x) \to 0$$

Remember all Lipschitz functions are continuous. Now $u_{n+1} = f(u_n)$

$$u_{n+1} \xrightarrow[n \to +\infty]{} \ell \Rightarrow u_n \xrightarrow[n \to +\infty]{} \ell \underbrace{\Longrightarrow}_{f \text{ is } \mathcal{C}^0} f(u_n) \xrightarrow[n \to +\infty]{} f(\ell)$$

But $u_{n+1} \to \ell$, so $\ell = f(\ell)$. This proves Theorem 2.18 □

Theorem 2.19 (Brouwer Fixed Point)**.** Let $D \subseteq \mathbb{R}^n$ be a nonempty, compact, and convex set. If $f : D \to D$ is $\mathcal{C}^0 \Longrightarrow \exists$ at least one point $x \in D$ such that $f(x) = x$.

Intuition. While no one has been able to axiomatize reality, I was able to notice Earth is a 2-sphere embedded in $\mathbb{R}^3$. I am currently in New York. Print a map of the city, which is of course a shrunken version of the city, and also a continuous function $f : NY \to NY$ that sends each point of New York to a point in the map. Then by Theorem 2.19, there exists at least one point $x \in NY$ such that $f(x) = x$, a point of New York that coincides with its representation on the map. For more information consult Borges.

Definition 2.24. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$. We say that f is strictly monotonic $\Leftrightarrow$ f is either strictly increasing or strictly decreasing, that is

$$\forall\, x < y \in A \Longrightarrow \begin{cases} f(x) < f(y) & \text{if } f \text{ is strictly increasing} \\ f(x) > f(y) & \text{if } f \text{ is strictly decreasing} \end{cases}$$

Definition 2.25 (Injective)**.** Let $f : A \to B$. We say that f is an injection (or one-to-one function) $\Leftrightarrow$

$$\forall\, x_1, x_2 \in A \text{ s.t. } f(x_1) = f(x_2) \Longrightarrow x_1 = x_2$$

Equivalently, if $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$.

Definition 2.26 (Surjective)**.** Let $f : A \to B$. We say that f is a surjection (or onto function) $\Leftrightarrow$

$$\forall\, y \in B\, \exists\, x \in A \text{ s.t. } f(x) = y.$$

That is, every element of B is the image of at least one element of A under f.

Definition 2.27 (Bijection)**.** Let $f : A \to B$. We say that f is a bijection $\Leftrightarrow$ f is both Definition 2.25 and Definition 2.26, so

$$\forall\, y \in B\, \exists!\, x \in A \text{ s.t } f(x) = y.$$

Intuition. The motivation to have this here is the Inverse Function Theorem Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is $\mathcal{C}^1$. Suppose Df_{x_0} is invertible i.e. $f'(x) \neq 0$, which is a bijection. Now

$$f(x) \simeq \underbrace{f(x_0) + Df_{x_0}(x - x_0)}_{\text{is a bijection}} + o(\|x - x_0\|)$$

We expect then locally that around x_0 f is a bijection, since it is strictly monotonic.

Lemma 2.6. Let $g : B(0, r) \subseteq \mathbb{R}^n \to \mathbb{R}^n$ s.t. $g(0) = 0$ and for which $\forall\, x, y \in B(0, r) \Rightarrow$

$$\big\|g(x) - g(y)\big\| \leqslant \frac{1}{2}\|x - y\|$$

$\implies f : B(0, r) \to \mathbb{R}^n$ with $f : x \mapsto x + g(x)$ is an injective function and

$$B\left(0, \frac{r}{2}\right) \subseteq f(B(0, r))$$

Proof. Let $x, y \in B(0, r)$ s.t. $f(x) = f(y)$.

$$x + g(x) = y + g(y) \implies \|x - y\| = \big\|g(x) - g(y)\big\| \leqslant \|x - y\|$$

$\Rightarrow \|x - y\| = 0 \Leftrightarrow x = y$. This satisfies Definition 2.25. Now let $y \in \left(0, \frac{r}{2}\right)$. We want to show $\exists\, x \in B(0, r)$ s.t.

$$\underbrace{f(x)}_{x+g(x)} = y \Leftrightarrow x = y - g(x)$$

We want a fixed point of $F : B(0, r) \to \mathbb{R}^n$ with $F : x \mapsto y - g(x)$

$$\big\|F(x)\big\| = \big\|y - g(x)\big\| \leqslant \|y\| + \big\|g(x) - g(0)\big\| < \frac{r}{2} + \frac{\|x\|}{2} < r$$

This means $F(B(0, r)) \subseteq B(0, r)$. Now, remember $\|y\| < \frac{r}{2}$ hence $\exists \varepsilon > 0$ s.t. $\|y\| \leqslant \frac{r}{2}(1-\varepsilon)$. Let $x \in B[0, r(1 - \varepsilon)]$ which as per Definition 1.2 is a closed ball $\Rightarrow$

$$\begin{aligned}\big\|F(x)\big\| &\leqslant \|y\| + \frac{1}{2}\|x\| \\ &\leqslant \frac{r}{2}(1 - \varepsilon) + \frac{r}{2}(1 - \varepsilon) = r(1 - \varepsilon)\end{aligned}$$

Hence $F(B[0, r(1 - \varepsilon)]) \subseteq B[(0, r(1 - \varepsilon)]$. Now

$$F(x) - F(x') = g(x') - g(x) \Rightarrow$$

$$\big\|F(x) - F(x')\big\| \leqslant \frac{1}{2}\big\|x - x'\big\|$$

Which means F is a strict contraction from X to itself where $X = B[(0, r(1 - \varepsilon)]$ is closed. This means that it satisfies Definition 1.12 and is complete. By Theorem 2.18 $\exists\, x \in B[0, r(1 - \varepsilon)]$ s.t.

$$F(x) \Leftrightarrow f(x) = y \implies y \in f(B(0, r))$$

This shows $B\left(0, \frac{r}{2}\right) \subseteq f(B(0, r))$ $\square$

Remark. Remember $f^{\leftarrow} \circ f = \text{id}$ so

$$D(f^{\leftarrow} \circ f)_x = \text{id} \Rightarrow D(f^{\leftarrow})_{f(x)} \circ Df_x \Rightarrow D(f')_{f(x)} = (Df_x)^{\leftarrow}$$

Definition 2.28 ($\mathcal{C}^1$-diffeomorphism)**.** A map $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is a local $\mathcal{C}^1$-diffeomorphism if $\forall\, x \in X\ \exists\, U \subseteq X$ open s.t. $x \in U$ and $\exists\, V \subseteq \mathbb{R}^n$ open s.t. $f(x) \in V$ s.t. $f\mid_U : U \to V$ is a bijection with a $\mathcal{C}^1$ inverse.

Theorem 2.20 (Inverse Function)**.** Let $f : E \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be $\mathcal{C}^1$ with E open. Let $x_0 \in E$. Suppose Df_{x_0} is invertible $\Longrightarrow \exists U \subseteq E$ open s.t. $x_0 \in U$ and $V \subseteq \mathbb{R}^n$ open with $f(x_0) \in V$ s.t. $f\mid_U$ is a bijection i.e. $f(U) = V$. Hence $\exists$ an inverse map $f^{\leftarrow} : V \to U$ that is $\mathcal{C}^1$ on V

$$\Rightarrow \forall\, x \in V \Longrightarrow D(f^{\leftarrow})_{f(x)} = (Df_x)^{\leftarrow}$$

Proof. Since Df_{x_0} is invertible, consider the map $\tilde{f}(x) := (Df_{x_0})^{\leftarrow}\big(f(x + x_0) - f(x_0)\big)$ For this new map $\tilde{f}$, we have $\tilde{f}(0) = 0$ and $D\tilde{f}_0 = \text{id}$. Hence, without loss of generality, we can assume from now on that

$$x_0 = 0 \Rightarrow f(0) = f(x_0) = 0 \Longrightarrow Df_0 = \text{id}$$

Let $g(x) = f(x) - x$ so that $f(x) = x + g(x)$

$$\Rightarrow \underbrace{Df_0}_{=\text{id}} = \text{id} + Dg_0 \Rightarrow Dg_0 = 0 \Rightarrow \|Dg_0\|_{\mathcal{L}} = 0$$

Where this norm satisfies Definition 2.4, and by its continuity

$$\exists\, r > 0 \text{ s.t. } \forall\, x \in B(0, r) \Longrightarrow \|Dg_x\| < \frac{1}{2}$$

Let $x, y \in B(0, r)$. Define $\gamma(t) = (1 - t)x + ty$

$$\Rightarrow \underbrace{g(\gamma(1))}_{f(y)} - \underbrace{g(\gamma(0))}_{f(x)} = \int_0^1 \frac{d}{dt} \underbrace{g(\gamma(t))}_{\text{is } \mathcal{C}^1} dt$$

By Theorem 2.8 then $\frac{d}{dt} g(\gamma(t)) = Dg_{\gamma(t)}(\gamma'(t))$

$$\Rightarrow g(y) - g(x) = \int_0^1 Df_{\gamma(t)}(g - x)dt$$

$$\Rightarrow \|g(y) - g(x)\| \leqslant \int_0^1 \underbrace{\left\|Df_{\gamma(t)}(y - x)\right\|}_{\leqslant \frac{1}{2}\|y-x\|} dt \leqslant \frac{1}{2}\|x - y\|$$

We apply Lemma 2.6 $\Rightarrow f\mid_{B(0,r)}$ is injective and $B\left(0, \frac{r}{2}\right) \subseteq f(B(0, r))$. We guess and set

$$V = B\left(0, \frac{r}{2}\right) \qquad \text{and} \qquad U = B(0, r) \cap f^{\leftarrow}\left(B\left(0, \frac{r}{2}\right)\right)$$

Since $U \subseteq B(0, r) \Rightarrow f\mid_U : U \to V$ is one-to-one. Moreover

$$\forall\, v \in V\ \exists\, x \in B(0, r) \text{ s.t. } f(x) = v \Longrightarrow x \in B(0, r) \cap f^{\leftarrow}(V) = U$$

$f\mid_{U:U\to V}$ is also surjective $\Longrightarrow$ it is a bijection. Since the ball $B\left(0, \frac{r}{2}\right)$ is open and f is $\mathcal{C}^0 \Longrightarrow$ by Theorem 2.5 $f^{\leftarrow}(B\left(0, \frac{r}{2}\right))$ is open, thus U is open and so is V. Now, let

$f^{\leftarrow} : V \to U$. We have to show $f^{\leftarrow}$ is differentiable at 0. By definition $f^{\leftarrow}(x) = 0$l Let $x_n \in V \to 0$. Let $y_n = f^{\leftarrow}(x_n) \in U$.

$$\Rightarrow \frac{\left\| f^{\leftarrow}(x_n) - x_n \right\|}{x_n} = \frac{\left\| y_n - f(y_n) \right\|}{\|x_n\|}$$

And since $x_n = f(y_n) = y_n + g(y_n)$

$$\|x_n\| \leqslant \|y_n\| + \left\| g(y_n) - g(0) \right\| \leqslant \|y_n\| + \frac{1}{2}\|y_n\| \leqslant \frac{3}{2}\|y_n\|$$

And notice that $g(0) = 0$ and then we have

$$\begin{aligned} \|x_n\| &\geqslant \|y_n\| - \left\| g(y_n) \right\| \\ &\geqslant \|y_n\| - \frac{1}{2}\|y_n\| = \frac{1}{2}\|y_n\| \end{aligned}$$

By both of these inequalities

$$\frac{1}{2}\|y_n\| \leqslant \|x_n\| \leqslant \frac{3}{2}\|y_n\| \Longrightarrow \frac{\left\| y_n - f(y_n) \right\|}{\|x_n\|} \leqslant 2\frac{\left\| y_n - f(y_n) \right\|}{\|x_n\|}$$

Since f is differentiable at 0 with $f(0) = 0$ and $Df_0 = \mathrm{id}$

$$\Longrightarrow \lim_{y \to 0} \frac{\left\| f(y) - f(0) - (y - 0) \right\|}{\|y - 0\|} = 0$$

Which is why we can conclude

$$\frac{\left\| y_n - f(y_n) \right\|}{\|y_n\|} \to 0 \Longrightarrow \frac{\left\| f^{\leftarrow}(x_n) - x_n \right\|}{\|x_n\|} \to 0$$

$\therefore f^{\leftarrow}$ is differentiable at 0 and $D(f^{\leftarrow})_0 = \mathrm{id}$ ☐

Intuition. Theorem 2.20 says that if f is $\mathcal{C}^1$ and $\forall\, x \in X$, the differential Df_x is invertible, then f satisfies Definition 2.28. In other words, f is a local $\mathcal{C}^1$-diffeomorphism.

Example 2.10. $F(x, y) = (e^x \cos, e^x \sin y)$. Let us show F is a local $\mathcal{C}^1$-diffeomorphism.

Proof. F is $\mathcal{C}^1$ since partials exist and are $\mathcal{C}^0$. We compute

$$\mathcal{J}(F)_{(x,y)} = \begin{pmatrix} \frac{\partial F_1}{\partial x} = e^x \cos y & \frac{\partial F_1}{\partial y} = -e^x \sin y \\ \frac{\partial F_2}{\partial x} = e^x \sin y & \frac{\partial F_2}{\partial y} = e^x \cos y \end{pmatrix}$$

Notice $\det\Big(\mathcal{J}(F)_{(x,y)}\Big) = e^{2x} \neq 0 \Longrightarrow \forall\, (x, y) \in \mathbb{R}^2 \Longrightarrow DF_{(x,y)}$ is invertible.

By Theorem 2.5 $\Longrightarrow F$ is a local $\mathcal{C}^1$-diffeomorphism. ☐

Theorem 2.21. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be $\mathcal{C}^1$ s.t. $\forall\, x \in \mathbb{R}^n \Rightarrow Df_x$ is invertible $\forall\, V \subseteq \mathbb{R}^n$ open $\Longrightarrow f(V)$ is open.

Proof. Let $V \subseteq \mathbb{R}^n$ be open and set $y \in f(V) \Rightarrow \exists\, x_o \in V$ s.t. $f(x_0) = y$. Since f is $\mathcal{C}^1$ and Df_{x_0} is invertible, we can apply Theorem 2.20. That is $\exists\, U \subseteq \mathbb{R}^n$ open s.t. $x_0 \in U$ and $\exists\, W \subseteq \mathbb{R}^n$ open with $f(x_0) \in W$ s.t. $f\,|_U \colon U \to W$ is a bijection. Now since $x_0 \in V$ and U we can pick U small enough such that

$$U \subseteq V \Longrightarrow W = f(U) \subseteq f(V)$$

But W is open and $f(x_0) = y \in W$. Thus $f(V)$ is open. ☐

2.6 Implicit Function Theorem

Theorem 2.22 (Implicit Function). Let $f : E \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^1$. Let $y \in E$ s.t. $f(y) = 0$. Suppose additionally that $\frac{\partial f}{\partial x_n}(y) \neq 0 \implies \exists\, V \subseteq \mathbb{R}^n$ open s.t. $y \in V$ and $\exists\, U \subseteq \mathbb{R}^{n-1}$ open s.t. $(y_1, \cdots, y_n) \in U$ and $\exists\, g : U \to \mathbb{R}$ that is $\mathcal{C}^1$ s.t.

$$\{x \in V \mid f(x) = 0\} = \{(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1}) \mid (x_1, \cdots, x_{n-1}) \in U\}$$

Moreover $\forall\, j \in \{1, \cdots, n-1\} \Rightarrow$

$$\frac{\partial g}{\partial x_j}(y_1, \cdots, y_{n-1}) = -\frac{\frac{\partial f}{\partial x_j}(y)}{\frac{\partial f}{\partial x_n}(y)}$$

Proof. Let $F : E \to \mathbb{R}^n$ with $F : (x_1, \cdots, x_n) \mapsto (x_1, \cdots, x_{n-1}, f(x_1, \cdots, x_n))$. By assumption f is $\mathcal{C}^1$ so F is $\mathcal{C}^1$ as well. Computing the Jacobian

$$\mathcal{J}(F)_{(y)} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ \frac{\partial f}{\partial x_1}(y) & \frac{\partial f}{\partial x_2}(y) & \cdots & \frac{\partial f}{\partial x_{n-1}}(y) & \frac{\partial f}{\partial x_n}(y) \end{pmatrix}(y)$$

This matrix is invertible. Notice that $\det(\mathcal{J}(F)_{(y)}) = 1 \times \cdots \times \frac{\partial f}{\partial x_n}(y) \neq 0$. This is because by assumption the derivative is nonzero $\Rightarrow DF_y : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. By Theorem 2.20 $\exists\, V \subseteq E$ and $W \subseteq \mathbb{R}^n$ open sets s.t. $y \in V$ and $F(x) \in W$ s.t. $F\,|_V : V \to W$ is a bijection and $F^{\leftarrow} : W \to V$ is $\mathcal{C}^1$. Let $h_1, \cdots, h_n : W \to \mathbb{R}$ s.t. $F^{\leftarrow}(x) = (h_1, \cdots, h_n)(x)$ with $x \in V$

$$\Rightarrow \underbrace{F(F^{\leftarrow}(x))}_{(h_1(x), \cdots, h_{n-1}(x), f((h_1(x), \cdots, h_n(x))))} = x = (x_1, \cdots, x_n)$$

$\Rightarrow h_1(x) = x_1, \cdots, h_{n-1}(x) = x_{n-1}$ and $\Rightarrow f(x_1, \cdots, x_{n-1}, h_n(x)) = x_n$. Set $U = \{(x_1, \cdots, x_{n-1}) \mid (x_1, \cdots, x_{n_1}, 0) \in W\}$. We want to prove the equality that was stated.

$\subseteq$. Let $x \in V$ s.t. $f(x) = 0$.

$$\Rightarrow F(x) \in V = \underbrace{(x_1, \cdots, x_{n-1})}_{\in W}$$

Since $F\,|_V : V \to W$. By definition of $U \Rightarrow (x_1, \cdots, x_{n-1}) \in W$. Notice then

$$x = F^{\leftarrow}(x_1, \cdots, x_{n-1}, 0) = (x_1, \cdots, x_{n-1}, h_n(x_1, \cdots, x_{n-1}, 0)$$

And since $g : U \to \mathbb{R}$ with $g(x_1, \cdots, x_{n-1}) = h_n(x_1, \cdots, x_{n-1}, 0)$. This shows $x \in \{(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1}) \mid (x_1, \cdots, x_{n-1}) \in U\}$.

$\supseteq$. Let $x \in \{(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1}) \mid (x_1, \cdots, x_{n-1}) \in U\}$. Recall that $\forall\, x' \in W \Rightarrow$

$$f(x'_1, \cdots, x'_{n-1}, h_n(x'_1, \cdots, x'_n) = x'_n$$

Suppose $x'_n \neq 0$. If $(x'_1, \cdots, x'_{n-1}, 0) \in W \Rightarrow$

$$f(x'_1, \cdots, x'_{n-1}, g(x'_1, \cdots, x'_{n-1})) = 0$$

Since $(x_1, \cdots, x_{n-1}) \in U \Rightarrow (x_1, \cdots, x_{n-1}, 0) \in W \Rightarrow$

$$f(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1})) = 0$$

So then we have $F^{\leftarrow}(x_1, \cdots, x_{n-1}, 0) \in V$

$$= (x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1})) = x \in V$$

Thus $x \in \{x' \in V \mid f(x') = 0\}$. Finally $\forall\, (x_1, \cdots, x_{n-1}) \in U \Rightarrow$

$$f(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1})) = 0$$

By Theorem 2.8, and since g is $\mathcal{C}^1$ since $F^{\leftarrow}$ is $\mathcal{C}^1$ we have that $\forall\, j \in \{1, \cdots, n-1\} \Rightarrow$

$$\frac{\partial f}{\partial x_j}(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1})) + \frac{\partial f}{\partial x_n}(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1}))\frac{\partial f}{\partial x_j}(x) = 0$$

We now argue that $g(y_1, \cdots, y_{n-1}) = y_n$ since $(y_1, \cdots, y_n) \in \{x \in V \mid f(x) = 0\} \in \{(x_1, \cdots, x_{n-1}, g(x_1, \cdots, x_{n-1}) \mid (x_1, \cdots, x_{n-1}) \in U\}$ and $(y_1, \cdots, y_{n-1}) \in U$. Hence we have $g(y_1, \cdots, y_{n-1}) = y_n$. This proves Theorem 2.22. $\square$

Example 2.11. Let $f : (x, y) \in \mathbb{R}^2 \mapsto \sin y + xy^4 + x^2$.

Proof. Let us look at $f(0,0) = \sin 0 + 0 \cdot 0^4 + 0^2 = 0$ and compute the partial derivative

$$\frac{\partial f}{\partial y}(x, y) = \cos y + 4xy^3 \Rightarrow \frac{\partial f}{\partial y}(0, 0) = \cos 0 + 4 \cdot 0 \cdot 0^3 = 1 \neq 0$$

Since f is $\mathcal{C}^1$ and $\frac{\partial f}{\partial y}(0,0) \neq 0$ we use Theorem 2.22 to get $U, V \subseteq \mathbb{R}$ open s.t. $0 \in U$ and V, as well as $\varphi : U \to V$ that is $\mathcal{C}^1$ s.t.

$$\forall\, x \in U \implies f(x, \varphi(x)) = 0$$

And $\varphi(0) = 0$. Now let's use Theorem 2.16 around 0 that is $\varphi(x) = \varphi(0) + \varphi'(0)x + o(x)$

$$\Rightarrow \frac{\partial f}{\partial x}(x, \varphi(x)) + \frac{\partial f}{\partial y}(x, \varphi(x)) \cdot \varphi'(x) = 0$$

At $x = 0$ we can get that $\frac{\partial f}{\partial x}(x, y) = y^4 + 2x$ and $\frac{\partial f}{\partial y}(x, y) = \cos y + 4xy^3$

$$\Rightarrow (0^4 + 2 \cdot 0) + (\cos 0 + 4 \cdot 0 \cdot 0^3)\varphi'(0) = 0 + 1 \cdot \varphi'(0) = 0 \Rightarrow \varphi'(0) = 0$$

Thus the Taylor Expansion is $\varphi(x) = 0 + 0 \cdot x + o(x) = o(x)$ $\square$

Example 2.12. Does the relation $x + y + z + \sin(xyz) = 0$ define z as a function of x and y in a neighborhood of the point $(0, 0, 0)$?

Proof. First let us define $f(x, y, z) = x + y + z + \sin(xyz)$ which satisfies $f(0, 0, 0) = 0$ as required. Now let us calculate the partial derivative with respect to z

$$\frac{\partial f}{\partial z} = 1 + \cos(xyz) \cdot \frac{\partial}{\partial z}(xyz) = 1 + \cos(xyz) \cdot (xy) \Rightarrow \frac{\partial f}{\partial z}(0, 0, 0) = 1 + \cos(0) \cdot 0 = 1 \neq 0$$

Thus $\frac{\partial f}{\partial z}(0, 0, 0) \neq 0$ and we can apply Theorem 2.22 to get $(0, 0) \in \mathbb{R}^2$ and $z = h(x, y)$ s.t. $f(x, y, h(x, y)) = 0$. Now

$$\frac{\partial z}{\partial x} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial z}} \qquad \text{and} \qquad \frac{\partial z}{\partial y} = -\frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial z}}$$

We have to get these.

$$\frac{\partial f}{\partial y} = 1 + \cos(xyz) \cdot xz = 1 + \cos(0) \cdot 0 = 1 \qquad \frac{\partial f}{\partial x} = 1 + \cos(xyz) \cdot yz = 1 + \cos(0) \cdot 0 = 1$$

And $\frac{\partial f}{\partial z}(0, 0, 0) = 1$. Therefore $\frac{\partial z}{\partial x}(0, 0) = -1$ and $\frac{\partial z}{\partial y}(0, 0) = -1$ $\square$

2.7 Lagrange Multiplier

Intuition. Let us refresh our memory and apply Lemma 2.4 to an example.

Example 2.13. Let $f(x, y) = \sin x + y^2 - 2y + 1$

Proof. We want to find the critical points. Recall from Definition 2.20 that these are the point that satisfy $\nabla f(x) = 0$. As such, let us compute the gradient.

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

Where (x, y) are the critical points. This gives us the following expression

$$\nabla f(a, b) = (\cos x, 2y - 2) = 0$$

Thus $2y - 2 = 0 \Rightarrow y = 1$ and $\cos x = 0 \Rightarrow \forall\, n \in \mathbb{N} \Rightarrow x = \frac{\pi}{2} + n\pi$. These are the critical points $(a, b) = (\frac{\pi}{2} + n\pi, 1)$. Now let us apply Lemma 2.4.

$$\nabla^2 f(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial x \partial y}(x, y) \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{pmatrix}$$

This matrix can be solved as follows

$$\nabla^2 f(x, y) = \begin{pmatrix} -\sin x & 0 \\ 0 & 2 \end{pmatrix} \Rightarrow \nabla^2 f\left(\frac{\pi}{2} + n\pi, 1\right) = \begin{pmatrix} -\sin\frac{\pi}{2} + n = -(-1)^n & 0 \\ 0 & 2 \end{pmatrix}$$

Thus this diagonal has two options. For n even we have a saddle point, as the diagonal does not share the same sign. If it is odd, both numbers are positive and thus the point is a local minimum. $\square$

Definition 2.29 (Restriction)**.** Let $f : X \to Y$ be a function and let $S \subseteq X$. The restriction of f to S, denoted $f\mid_S$, is the function $f\mid_S : S \to Y$ defined as $\forall\, x \in S \Rightarrow f\mid_S(x) = f(x)$.

Intuition. The setup is the following. We want to maximize/minimize a function $f : \mathbb{R}^n \to \mathbb{R}$ under the constraint $g = 0$ where $g : \mathbb{R}^n \to \mathbb{R}$ where both f, g are $\mathcal{C}^1$. The level set is $S = \{g = 0\}$ where $\nabla g(x) \neq 0$. Now $\forall$ tangent vector v of S at $x \Rightarrow \langle v, \nabla g(x)\rangle = 0$. Suppose x is a local maximum of $f\mid_S$. Theorem 2.10 says $\nabla f(x)$ is the direction of the sharpest increase.

$\exists$ a path $\gamma : (-\varepsilon, \varepsilon) \to S$ that is $\mathcal{C}^1$ s.t. $\gamma(0) = x$ and $\gamma'(0) = v$. Now we look at

$$\frac{d}{dt} f(\gamma(t))\mid_{t=0} = \langle \nabla(\gamma(t)), \gamma'(t)\rangle\mid_{t=0} = \langle \nabla f(x), v\rangle = 0$$

since $\|\nabla f(x)\|\|v\| \cos\theta$ with $\theta \in \left(0, \frac{\pi}{2}\right) \implies\impliedby$. Contradiction, so that implies x is not a local max. Impossible! Hence $\nabla f(x) \perp$ to S at x. Then the tangent plane is a hyperplane of $\dim n-1$ and its orthogonal complement is of $\dim 1$. $\nabla g(x) \neq 0$. Hence $\exists\, \lambda \in \mathbb{R}$ s.t.

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

Take $g(x) = 0$ with $y \approx x \Rightarrow$

$$g(y) \approx \underbrace{g(x)}_{=0} + \langle \nabla g(x), y - x\rangle$$

Hence $y \approx$ in level set if $\langle \nabla g(x), y - x\rangle = 0$. So $\dim n - 1$. Since $y \in \mathbb{R}^n$ with one non-trivial constraint. So, if x is an extremum of f subject to $g = 0$ then f cannot increase in any tangent direction on the constraint surface, so its gradient must be parallel to that of the constraint.

Remark. The dimension of a vector space V is the number of vectors in a basis of V, that is, the maximal number of linearly independent vectors in V. It is denoted by $\dim V$.

Lemma 2.7. Suppose that x is a local minmax of $f\mid_S$. Let T_xS be the set of vectors of S at x. That is $v \in T_xS$ if $\exists\, \varepsilon > 0$ and $\gamma : (-\varepsilon, \varepsilon) \to S$ that is $\mathcal{C}^1$ s.t. $\gamma(0) = x$ and $\gamma'(0) = v \implies \forall\, v \in T_xS \Rightarrow \langle \nabla f(x), v\rangle = 0$.

Proof. Let $v \in T_xS$ and by definition $\exists\, \gamma : (-\varepsilon, \varepsilon) \to S$ that is $\mathcal{C}^1$ s.t. $\gamma(0) = x$ and $\gamma'(0) = v \Rightarrow p : t \mapsto f(\gamma(t))$ has a local minmax at $0 \Rightarrow$

$$p'(0) = 0$$

where $p'(0) = \langle \nabla f(\gamma(0)), \gamma'(0)\rangle = \langle \nabla f(x), v\rangle$ □

Definition 2.30 (Rank)**.** Let $\mathbf{A} \in \mathcal{M}_{m\times n}(\mathbb{R})$ The rank of $\mathbf{A}$, denoted $\operatorname{rank}(\mathbf{A})$, is the dimension of $\operatorname{Im}(\mathbf{A})$, that is, the number of linearly independent columns (or rows) of $\mathbf{A}$.

Lemma 2.8. Suppose $\nabla g(x) \neq 0 \implies T_xS$ is a hyperplane.

Proof. Notice this is true for linear function, hence it is true for $\mathcal{C}^1$ functions. Suppose that $\frac{\partial g}{\partial x_n}(x) \neq 0$. By Theorem 2.22, and since $\nabla g(x) \neq 0 \Rightarrow\ \exists$ open set $V \subseteq \mathbb{R}^n$ s.t. $y \in V$ and $U \subseteq \mathbb{R}^n$ open set s.t. $(x_1, \cdots, x_{n-1}) \in U$ and $h : U \to \mathbb{R}$ s.t.

$$\{g = 0\} \cap V = (y_1, \cdots, y_{n-1}, h(y_1, \cdots, y_{n-1})) = \operatorname{graph}(h)$$

Let's write $\Phi : U \to \mathbb{R}^n$ with $\Phi : (y_1, \cdots, y_{n-1}) \mapsto (y_1, \cdots, y_{n-1}, h(y_1, \cdots, y_{n-1}))$. We claim $T_xS = \operatorname{Im}(D\Phi_{(x_1,\cdots,x_{n-1})}) = \{D\Phi_{(x_1,\cdots,x_{n-1})}(v) \mid v \in \mathbb{R}^{n-1}\}$

$\subseteq$. Let $v \in T_xS \Rightarrow\ \exists\, \gamma : (-\varepsilon, \varepsilon) \to S$ that is $\mathcal{C}^1$ s.t. $\gamma(0) = x$ and $\gamma'(0) = v$ and $\forall\, t \in (-\varepsilon, \varepsilon) \Rightarrow \gamma(t) \in V$. Since $\gamma(t) \in S \cap V = \operatorname{graph}(h) = \Phi(U) \Rightarrow\ \exists\, \tilde{\gamma}(t) \in U$ s.t.

$$\gamma(t) = (\tilde{\gamma}(t), h(\tilde{\gamma}(t)))$$

Now since γ is $\mathcal{C}^1$ so is $\tilde{\gamma} \Rightarrow \gamma(t) = \Phi(\tilde{\gamma}(t))$ with $\tilde{\gamma}$ and Φ both $\mathcal{C}^1$. By Theorem 2.8 $\Rightarrow$

$$\gamma'(t) = D\Phi_{\tilde{\gamma}(t)}(\tilde{\gamma}(t))$$

with $t = 0 \Rightarrow v = D\Phi_{(x_1,\cdots,x_{n-1})}(\tilde{\gamma}'(0)) \in \operatorname{Im}(D\Phi_{(x_1,\cdots,x_{n-1})})$

$\supseteq$. Let $w \in \operatorname{Im}(D\Phi_{(x_1,\cdots,x_{n-1})}) \Rightarrow \exists w \in \mathbb{R}^{n-1}$ s.t. $v = D\Phi_{(x_1,\cdots,x_{n-1})}(w)$. Let $\tilde{\gamma}(t) = \tilde{x} + tw$. Set

$$\gamma(t) = (\tilde{\gamma}(t), h(\tilde{\gamma}(t)))$$

The for $\varepsilon > 0$ small enough $\forall\, t \in (-\varepsilon, \varepsilon) \Rightarrow \tilde{\gamma}(t) \in U$. Hence $\gamma(t) \in S \cap V$ and $\forall\, t \in (-\varepsilon, \varepsilon) \Rightarrow \gamma(t) = \Phi(\tilde{\gamma}(t))$. Then by Theorem 2.8 $\Rightarrow \gamma'(0) = D\Phi_{(x_1,\cdots,x_{n-1})}(\tilde{\gamma}'(0) = w) = v$. Hence $v \in T_xS$ by definition. Thus $T_xS = \operatorname{Im}(D\Phi_{(x_1,\cdots,x_{n-1})})$ is a vector space

$$\mathcal{J}\Phi_{(x_1,\cdots,x_{n-1})} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \frac{\partial h}{\partial y_1} & \frac{\partial h}{\partial y_2} & \cdots & \frac{\partial h}{\partial y_{n-1}} \end{pmatrix}(x_1, \cdots, x_{n-1})$$

$\Rightarrow \Phi(y_1, \cdots, y_{n-1}) = (y_1, \cdots, y_{n-1}, h(y_1, \cdots, y_{n-1}))$. Now since $D\Phi_{(x_1,\cdots,x_{n-1})} : \mathbb{R}^{n-1} \to \mathbb{R}^n$ the rank is $n-1$, since the column vectors are linearly independent. Hence $\dim T_xS = n-1$. This proves Lemma 2.8. □

Theorem 2.23 (Lagrange). Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^1$. Let x be a local minmax of $f\mid_S$ with $S = \{g = 0\}$. Suppose $\nabla g(x) \neq 0 \implies \exists\, \lambda \in \mathbb{R}$ s.t.

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

Proof. Notice that $\dim(T_x S)^\perp = n - \dim T_x S = n - (n-1) = 1$. So

$$A^\perp = \{v \mid \forall\, a \in A \Rightarrow \langle v, a\rangle = 0\}$$

And $\nabla g(x) \in (T_x S)^\perp$ the gradient orthogonal to the level set sine $\nabla g(x) \neq 0$. So $(\nabla g(x))$ is a basis of $(T_x S)^\perp \Rightarrow$ by Lemma 2.8 $\implies \nabla f(x) \in (T_x S)^\perp$. Hence $\nabla f(x) = \lambda \nabla g(x)$ for some $\lambda \in \mathbb{R}$. This proves Theorem 2.23. $\square$

Example 2.14. $f(x, y) = x^2 - y^2$. Optimize on $S = \{(x, y) \mid g(x, y) = x^2 - y^2 - 1 = 0\}$

Proof. Let (x, y) be a local minmax on $f\mid_S$. So f, g are $\mathcal{C}^1 \Rightarrow \nabla g(x, y) = (2x, 2y) \neq 0$ since $x^2 - y^2 = 1$. By Theorem 2.23 $\exists\, \lambda \in \mathbb{R}$ s.t.

$$\nabla f(x, y) = \lambda \nabla g(x, y) \Rightarrow (2x, -2y) = \lambda(2x, 2y)$$

Notice $x = \lambda x \Rightarrow x(\lambda - 1) = 0 \Rightarrow x = 0$ or $\lambda = 1$. If $x = 0 \Rightarrow y \pm 1$ so $(0, 1)$ and $(0, -1)$ are candidates. If $\lambda = 1 \Rightarrow -2y = 2y \Rightarrow y = 0$ so $x = \pm 1$ and $(-1, 0)$ and $(1, 0)$ are also candidates. Since f is $\mathcal{C}^0 \Rightarrow S$ is compact. Hence $f\mid_S$ is bounded and attains both min and max.

$$\underbrace{f(\pm 1, 0) = 1}_{\text{global maxima of } f|_S} \qquad \text{and} \qquad \underbrace{f(0, \pm 1) = 1}_{\text{global minima of } f|_S}$$

This solves the exercise. $\square$

Theorem 2.24. Let $S = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ be the unit sphere, and let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Suppose that the restriction of f to S is constant $\implies \exists\, x_0 \in \mathbb{R}^n$ with $\|x_0\| < 1$ s.t. $\nabla f(x_0) = 0$

Proof. Define $g(x) = \|x\|^2 - 1$. Then $S = \{x \mid g(x) = 0\}$ and $\nabla g(x) = 2x$. Since f is constant on S $\exists\, c$ s.t. $\forall\, x \in\Rightarrow S f(x) = c$. Hence each $x \in S$ is an extremum of $f|_S$. By Theorem 2.23 $\forall\, x \in S\, \exists\, \lambda_x \in \mathbb{R}$ s.t.

$$\nabla f(x) = \lambda_x \nabla g(x) = 2\lambda_x x$$

Take inner product with x and use $\|x\| = 1$

$$\langle \nabla f(x), x\rangle = 2\lambda_x$$

Define $\varphi_x : (-1, 1) \to \mathbb{R}$ by $\varphi_x(t) = f(tx)$. By Theorem 2.8

$$\varphi_x'(t) = \langle \nabla f(tx), x\rangle \Rightarrow \varphi_x'(1) = \langle \nabla f(x), x\rangle = 2\lambda_x$$

But φ_x is constant at $t = 1$ (since f is constant on S) $\Rightarrow \varphi_x'(1) = 0$. Thus $\lambda_x = 0$ and therefore

$$\forall\, x \in S \Rightarrow \nabla f(x) = 0$$

Since $\forall x \in S \Rightarrow \nabla f(x) = 0$, define $F(x) = x - \nabla f(x)$. Then $F(x) = x$ on S, so F maps the closed unit ball $B = \{x \mid \|x\| \leqslant 1\}$ into itself. By Theorem 2.19, $\exists\, x_0 \in B$ s.t. $F(x_0) = x_0$. Hence $\nabla f(x_0) = 0$. If $\|x_0\| = 1 \Rightarrow x_0 \in S$, otherwise $\|x_0\| < 1$. Either way, x_0 exists. $\square$

Remark. $f(a(t), b(t)) = \gamma(t) \Rightarrow \gamma'(t) = a'(t)\frac{\partial f}{\partial x}(a(t), b(t)) + b'(t)\frac{\partial f}{\partial}(a(t), b(t))$ (Theorem 2.8)

Definition 2.31. Let $T : V \to W$ be a linear map between vector spaces. The kernel of T is the set of all vectors in V that are mapped to the zero vector in W, that is

$$\ker(T) = \{v \in V \mid T(v) = 0\}$$

Theorem 2.25 (Lagrange). Let $f : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^1$ and $g_1, \cdots, g_m : \mathbb{R}^n \to \mathbb{R}$ also $\mathcal{C}^1$ with $m \leqslant n$. Set $S = \{g_1 = 0, \cdots, g_m = 0\}$. Let x be a local minmax of $f\mid_S$. Suppose that $\nabla g_1(x), \cdots \nabla g_m(x)$ are linearly independent $\Longrightarrow$

$$\nabla f(x) = \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) + \cdots + \lambda_m \nabla g_m(x)$$

for some $\lambda_1, \lambda_2, \cdots, \lambda_m \in \mathbb{R}$

Chapter 3

Measure Theory

3.1 Preliminaries

Intuition. There are several motivations for this. One of those is we want to integrate functions that are not smooth i.e. functions where we cannot integrate under Riemann integral.

$$\mathbb{1}_{\mathbb{Q}} = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is an example of such a function: it is not Riemann integrable on $[0,1]$. Another motivation is to build a theory closed under limit $\Rightarrow \forall\, n \in \mathbb{N}$ we have f_n integrable and $f_n \to f$. For Riemann integral f is not integrable. For Lebesgue integral it is. This is what we will construct. We are also interested in constructing a notion of volume of $A \subseteq \mathbb{R}^d$. For $\mathbb{R}^2$ we have area, and for $\mathbb{R}$ we have length. Finally, another motivation is taking a point at random between 0 and 1. That is, constructing a probability measure on $[0,1]$ i.e.

$$\mathbb{P}(X \in A) = |A|$$

where X is a random point. There exists no such thing, a probability measure s.t. we can compute $\forall\, A \subseteq [0,1] \Rightarrow \mathbb{P}(X \in A)$. The set of A s.t. we can compute $\mathbb{P}(X \in A)$ is called the set of measurable sets of $[0,1]$. The Lebesgue measure is defined on $\mathcal{A} \subseteq \mathbb{P}([0,1])$ called Borelian.

Definition 3.1 (Rectangle). A closed rectangle in $\mathbb{R}^d$ is a set of the form

$$[a_1, b_1] \times [a_2, b_2] \times \cdots [a_d, b_d]$$

with $a_1 \leqslant b_1 \cdots a_d \leqslant b_d$. An open rectangle in $\mathbb{R}^d$ is a set of the form

$$(a_1, b_1) \times (a_2, b_2) \times \cdots (a_d, b_d)$$

with $a_1 < b_1 \cdots a_d < b_d$.

Corollary (Cube). A cube in $\mathbb{R}^d$ is a rectangle s.t. $b_1 - a_1 = \cdots = b_d - a_d$

Definition 3.2 (Volume). We define the volume of an open or closed rectangle R to be

$$|R| = \prod_{i=1}^{n} (b_i - a_i)$$

Definition 3.3 (Open Set)**.** A set $O \subseteq \mathbb{R}^d$ is open is $\forall\, x \in O\, \exists\, \varepsilon \in \mathbb{R}^+$ s.t.

$$B(x, \varepsilon) \subseteq O$$

Intuition. For $d = 1 \Rightarrow (a, b)$ is open. Any union of open sets is open i.e. $(-\infty, a) \cup (b, \infty)$

Remark. $\mathcal{I}$ is countable if $|\mathcal{I}| < \infty$ or if $\mathcal{I}$ is in bijection with $\mathbb{N}$

Theorem 3.1. Let $O \subseteq \mathbb{R}$ be open $\Rightarrow O$ can be decomposed into a countable union of open (non-empty) disjoint intervals. Moreover, this decomposition is unique.