ANALYSIS OF SEVERAL VARIABLES

Jaydeb Sarkar, notes by Ramdas Singh

Third Semester

List of Symbols

Placeholder

Contents

Chapter 1

$\mathbb{R}^n$: LIMITS, CONTINUITY, AND DIFFERENTIABILITY

1.1 Translation into Higher Dimensions

July 21st.

We begin with a definition.

Definition 1.1. The space $\mathbb{R}^n$ is defined as $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ (n times) $= \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}\}$.

In the context of analysis, we will talk about open sets, closed sets, sequences, compact sets, and connected sets. In contrast, algebra considers $\mathbb{R}^n$ as a vector space with the operators $+$ and $\cdot$. Combining both these aspects results in the study of analysis of several variables. In this course, we will mainly focus on dealing with functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$, and talk about their properties such as continuity, differentiability, and integrability.

1.1.1 Algebraic and Analytic Structure

We note that $\mathbb{R}^n$ is also an inner product space with the following properties:

- $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ for all $x, y \in \mathbb{R}^n$.
- The set $\{e_i\}_{i=1}^n$ consisting of the unit vectors is an orthonormal basis for $\mathbb{R}^n$.
- The simplest maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ are linear maps that send lines to lines.

Example 1.2. Suppose the function f is a linear map from $\mathbb{R}$ to $\mathbb{R}$. This implies that $f(x) = xf(1)$ for all $x \in \mathbb{R}$. Thus, $f(x) = cx$ for all $x \in \mathbb{R}$, where c is a constant. Conversely, if $c \in \mathbb{R}$, then $x \mapsto cx$ is a linear map. Therefore, we conclude that $\{f : \mathbb{R} \to \mathbb{R}, \text{linear}\} \leftrightarrow \mathbb{R}$, with a possible bijection given by $f \mapsto f(1)$.

In the above example, we note that 1 is not special; we could simply fix any $\alpha \in \mathbb{R} \setminus \{0\}$, and notice that $f(x) = \frac{x}{\alpha} f(\alpha)$ for all $x \in \mathbb{R}$. Here, replacing $f(1)$ by $f(\alpha)$ is a kind of 'change of variable'.

Remark 1.3. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then, $Le_j = \sum_{i=1}^m a_{ij} e_i$ for all $j = 1, 2, \ldots, n$. We may write L as $(a_{ij})_{m\times n} \in M_{m\times n}(\mathbb{R})$, the set of all $m \times n$ matrices with real entries.

Coming to the analysis side, there is a need for defining a distance between points in $\mathbb{R}^n$. Previously, we have seen that the *norm* may be given as $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$ for all $x \in \mathbb{R}^n$. We can use this norm to define our required distance function.

Definition 1.4. The *distance* function between two points $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is defined as $d(x, y) = \|x - y\| = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$ for all $x, y \in \mathbb{R}^n$.

For $n = 1$, we note that $d(x, y) = |x - y|$, from the previous analysis courses. $\mathbb{R}^n$ equipped with the function d is called a *metric space*. Coming to the properties of the inner product, we have

- $\|x\| = \langle x, x\rangle^{1/2}$ for all $x \in \mathbb{R}^n$.
- $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in \mathbb{R}^n$.
- The function $\langle , \rangle$ is linear with respect to the first and second arguments.

We also have the important Cauchy-Schwarz inequality.

Theorem 1.5 (*Cauchy-Schwarz inequality*)**.** *For all $x, y \in \mathbb{R}^n$, we have $|\langle x, y\rangle| \leq \|x\| \|y\|$.*

Proof. Note that

$$0 \leq \sum_{i=1}^n \sum_{j=1}^n (x_i y_j - x_j y_i)^2 = 2\left(\sum_{i,j} x_i^2 y_j^2 - \sum_{i,j} x_i x_j y_i y_j\right) = 2\left(\|x\|^2 \|y\|^2 - \langle x, y\rangle^2\right) \tag{1.1}$$

$$\implies |\langle x, y\rangle| \leq \|x\| \|y\| . \tag{1.2}$$

■

We note that equality occurs if and only if the first quantity in the above equation is zero, *i.e.*, if and only if $x_i y_j = x_j y_i$ for all i, j, or $\frac{x_i}{y_i} = \frac{x_j}{y_j}$ for all i, j showing that x and y are linearly dependent.

Corollary 1.6 (*Triangle inequality*)**.** *For all $x, y \in \mathbb{R}^n$, we have $\|x + y\| \leq \|x\| + \|y\|$.*

Proof. We have

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \tag{1.3}$$

where the inequality follows from Cauchy-Schwarz. ■

The following will prove to be an important result.

Theorem 1.7. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then, there exists a $M > 0$ such that $\|Lx\| \leq M\|x\|$ for all $x \in \mathbb{R}^n$.*

Proof. Rewriting x as $x = \sum_{i=1}^n x_i e_i$, we have

$$Lx = \sum_{i=1}^n x_i L e_i$$

$$\implies \|Lx\| = \left\|\sum_{i=1}^n x_i L e_i\right\| \leq \sum_{i=1}^n |x_i| \|Le_i\| \leq \|x\| \sum_{i=1}^n \|Le_i\| = \|x\| M. \tag{1.4}$$

The first inequality follows from the triangle inequality, and the second from Cauchy-Schwarz. In the last step, M is set to be $\sum_{i=1}^n \|Le_i\|$, which is a constant. ■

We also term $(\mathbb{R}^n, d)$ as a Euclidean metric space. There is now a need to define open sets in $\mathbb{R}^n$ to talk more about the analysis of several variables.

Definition 1.8. For $a \in \mathbb{R}^n$ and $r > 0$, the *open ball* centred at a of radius r is $B_r(a) := \{x \in \mathbb{R}^n : d(x, a) < r\}$, the set of all points in $\mathbb{R}^n$ that are at a distance less than r from a.

From the notion of open balls, we can define open sets.

Definition 1.9. A set $S \subseteq \mathbb{R}^n$ is said to be an *open set* if for all $x \in S$, there exists an $r > 0$ such that $B_r(x) \subseteq S$.

We now bring the notion of convergence of sequences.

Definition 1.10. Let $\{x_m\} \subseteq \mathbb{R}^n$ be a sequence and $x \in \mathbb{R}^n$. We say that $\{x_m\}$ *converges* to x if for every $\varepsilon > 0$, there exists a natural N such that $\|x_m - x\| < \varepsilon$ for all $m \geq N$.

July 23rd.

Definition 1.11. Let $S \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$. We say that a is a *limit point* of S if $S \cap (B_r(a) \setminus \{a\})$ is non-empty for all $r > 0$.

We introduce more notation; for all $i = 1, 2, \ldots, n$, the mapping $\Pi_i : \mathbb{R}^n \to \mathbb{R}$ is called the i^{th} *projection* where $\Pi_i(x) = x_i$. Note that $x = (x_1, x_2, \ldots, x_n) = (\Pi_1(x), \Pi_2(x), \ldots, \Pi_n(x))$. This notation allows us to formulate the following useful fact a little more neatly.

Theorem 1.12. *Let $\{x_m\} \subseteq \mathbb{R}^n$ be a sequence and $x \in \mathbb{R}^n$. Then $x_m \to x$ if and only if $\Pi_i(x_m) \to \Pi_i(x)$ for all $i = 1, 2, \ldots, n$.*

Proof. Suppose $x_m \to x$. Then for all $\varepsilon > 0$, there exists a natural N such that $\|x_m - x\| < \varepsilon$ for all $N \geq N$. Restating, we have

$$\sum_{i=1}^{n} (\Pi_i(x_m) - \Pi_i(x))^2 < \varepsilon^2 \text{ for all } n \geq N \tag{1.5}$$

$$\implies \text{For all } i,\ |\Pi_i(x_m) - \Pi_i(x)| < \varepsilon^2 \text{ for all } n \geq N. \tag{1.6}$$

For the converse, we simply work backwards with $\varepsilon/\sqrt{n}$ as our choice of epsilon. ■

For example, the sequence $\{(\frac{1}{n}, \frac{1}{2n+3})\}_{n=1}^{\infty}$ converges to $(0, 0)$. However, the sequence $\{(\frac{1}{n}, n^2)\}_{n=1}^{\infty}$ does not.

Definition 1.13. Let $S \subseteq \mathbb{R}^n$. $a \in S$ is termed an *interior point* of S if for some $r > 0$, $B_r(a) \subseteq S$ holds. Thus, a set S is open if a is an interior point for all $a \in S$. The *interior* of set S is defined as $\operatorname{int} S := \{a \in S \mid a \text{ is an interior point.}\}$.
If $a \in \operatorname{int}(S^c)$, then a is termed an *exterior point* of S. a is termed a *boundary point* if $B_r(a)$ meets both S and S^c for all $r > 0$. The set of *boundary points* of S is denoted as ∂S.

We also term a set $S \subseteq \mathbb{R}^n$ as a *closed set* if $\mathbb{R}^n \setminus S$ is open. The following facts will only be stated and will be left as an exercise to the reader:

- A set $C \subseteq \mathbb{R}^n$ is closed if and only if for all sequences $\{x_m\}_{m=1}^{\infty} \subseteq C$ that converge to x implies $x \in C$.
- The open ball $B_r(a)$ is an open set.
- The intersection of an arbitrary collection of closed sets is closed; likewise, the union of an arbitrary collection of open sets is open.
- The set $S \subseteq \mathbb{R}^n$ is open if and only if $S = \operatorname{int} S$.

Fix $O \subseteq \mathbb{R}^n$.

- O is open if and only if $O \cap \partial O = \emptyset$.
- O is closed if and only if $\partial O \subseteq O$.

For $S \subseteq \mathbb{R}^n$, we define the *closure* of set S as $\overline{S} = \operatorname{int} S \cup \partial S$.

- $S \subseteq \mathbb{R}^n$ is closed if and only if $\overline{S} = S$.
- Let $C_i \subseteq R$ be closed sets and $O_i \subseteq \mathbb{R}$ be open sets, for $i = 1, 2, \ldots, n$. Then $C_1 \times C_2 \times \cdots \times C_n \subseteq \mathbb{R}^n$ is a closed set, and $O_1 \times O_2 \times \cdots \times O_n \subseteq \mathbb{R}^n$ is an open set.
- The n *dimensional unit sphere* $S^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ is closed in $\mathbb{R}^n$.

Definition 1.14. For $S \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$, a is termed an *isolated point* if a is *not* a limit points; that there exists an $r > 0$ such that $S \cap (B_r(a) \setminus \{a\}) = \emptyset$.

With the pesky definitions and translation of one dimensional concept into being defined over several variables, we come to limits and continuity.

1.2 Limits and Continuity

Recall that given $f : (a, b) \setminus \{c\} \to \mathbb{R}$, we say that $\lim_{x\to c} f(x) = b$ if for every $\varepsilon 0$, there exists a $\delta > 0$ such that $|f(x) - b| < \varepsilon$ for all x satisfying $0 < |x - c| < \delta$. Note that in this definition of the limit, we have $f(x) \in B_\varepsilon(b)$ and $x \in B_\delta(c) \setminus \{c\}$; this can easily be rewritten as $f(B_\delta(c) \setminus \{c\}) \subseteq B_\varepsilon(b)$. However, for our definition we would not require f to be defined on an open set. We defind it over any arbitrary set.

Definition 1.15. Let $a \in S \subseteq \mathbb{R}^n$ be a limit point of S and let $f : S \setminus \{a\} \to \mathbb{R}^m$ be a function and $b \in \mathbb{R}^m$. We say $\lim_{x\to a} f(x) = b$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $f((B_\delta(a)\setminus\{a\})\cap S) \subseteq B_\varepsilon(b)$. Again, this is equivalent to saying that $\|f(x) - b\| < \varepsilon$ for all $x \in S\setminus\{a\}$ satisfying $\|x - a\| < \delta$.

It is important to get accustomed to the definition that works with open balls.

Remark 1.16. In the above definition, if we instead write $x - a = h$, then $\lim_{x\to a} f(x) = b$ is equivalent to saying that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|f(a + h) - b\| < \varepsilon$ for all $\|h\| < \delta$. We can further rewrite to get the usual notation of

$$\lim_{\|h\|\to 0} \|f(a + h) - b\| = 0. \tag{1.7}$$

Note that the above limit is in the real numbers, making it eaiser to deal with.

A notion of continuity also comes in handy.

Definition 1.17. For $S \subseteq \mathbb{R}^n$, let $f : S \to R^m$ with $a \in S$. We say f is *continuous* at a if $\lim_{x\to a} f(x) = f(a)$. In other words, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\|f(x) - f(a)\| < \varepsilon \text{ for all } x \in S \text{ satisfying } \|x - a\| < \delta \tag{1.8}$$

or

$$f(B_\delta(a) \cap S) \subseteq B_\varepsilon(f(a)). \tag{1.9}$$

Note that if a is an isolated point of S, then any $f : S \to \mathbb{R}^m$ is continuous at a since $f(\{a\}) \subseteq B_\varepsilon(f(a))$ holds true, trivially.

Remark 1.18. Similar to the previous remark, f is continuous at a if and only if

$$\lim_{\|h\|\to 0} \|f(a + h) - f(a)\| = 0. \tag{1.10}$$

Functions defined on $S \subseteq \mathbb{R}^n$ can be broken down into components; given $f : S \to R^m$, define $f_j := \Pi_j \circ f$ for all $j = 1, 2, \ldots, m$. Thus, f can be rewritten as $(f_1, f_2, \ldots, f_m)$. We can conclude that f is continuous at $a \in S$ if and only if $f_j : S \to \mathbb{R}$ is continuous at a for all $j = 1, 2, \ldots, m$. The proof of this observation is left as an exercise to the reader.

Theorem 1.19. *Let $a \in \mathbb{R}^n$ be a limit point of a set $S \subseteq \mathbb{R}^n$, with $b \in \mathbb{R}^m$ and $f : S \to R^m$ a function. Then, the following are equivalent—*

1. $\lim_{x\to a} f(x) = b$.
2. *If* $\{x_p\} \subseteq S \setminus \{a\}$ *and* $x_p \to a$, *then* $f(x_p) \to b$.
3. $\lim_{x\to a} \|f(x) - b\| = 0$.

The proof of this theorem is left as an exercise to the reader.

Definition 1.20. For a set $S \subseteq \mathbb{R}^n$, a function $f : S \to \mathbb{R}^m$ is termed a continuous function if f is continuous at a for all $a \in S$.

Theorem 1.21. *Let $f : S \to \mathbb{R}^m$ be a function, where $S \subseteq \mathbb{R}^n$. The following are, then, equivalent—*

1. f *is continuous.*
2. *For all* $a \in S$ *and* $\{x_n\} \subseteq S$ *with* $x_n \to a$, *we have* $f(x_n) \to f(a)$.
3. *For all open sets* $O \subseteq \mathbb{R}^m$, *the set* $f^{-1}(O) \subseteq S$ *is also open.*
4. *For all closed sets* $C \subseteq \mathbb{R}^m$, *the set* $f^{-1}(C) \subseteq S$ *is also closed.*

Proof. For 1. implies 3. , let $O \subseteq \mathbb{R}^m$ be open. Pick some $a \in f^{-1}(O)$. Then, since $f(a) \in O$, there exists $r > 0$ such that $B_r(f(a)) \subseteq O$. Also, f is continuous at a; for $\frac{r}{2} > 0$, there exists $\delta > 0$ such that

$$f(B_\delta(a)) \subseteq B_{\frac{r}{2}}(f(a)) \subseteq B_r(f(a)) \implies a \in B_\delta(a) \subseteq f^{-1}(B_r(f(a))) \subseteq f^{-1}(O). \tag{1.11}$$

Thus, $f^{-1}(O)$ is open. For 3. implies 1. , let $a \in S$. Fix $\varepsilon > 0$. Then the set $f^{-1}(B_\varepsilon(f(a)))$ is open; there exists a $\delta > 0$ such that $B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a)))$. ∎

July 28th.

We look at a few examples.

Example 1.22. Let $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ be defind as $f(x,y) = \frac{2xy}{x^2+y^2}$. We find the limit $\lim_{(x,y)\to(0,0)} f(x,y)$. Let us approach from different directions, starting with the line L_1 defined as $y = 0$ with $x > 0$. Then $\lim_{(x,y)\to(0,0);(x,y)\in L_1} f(x,y) = \lim_{(x,y)\to(0,0)} 0 = 0$. However, along the line L_2 defned as $\{(x,y) \mid x = y\ x,y > 0\}$, we have $\lim_{(x,y)\to(0,0);(x,y)\in L_2} f(x,y) = 1$. We conclude that this limit cannot exist. Going along the line $y = mx$ gives several possible values for the limit.

The above method is good only for showing that the limit does not exist; if the limit does exist, we need to use theory.

Example 1.23. We compute the limit $\lim_{(x,y)\to(0,0)} \frac{x^3}{x^2+y^2}$. Here, we can prove that the limit exists as follows:

$$\left|\frac{x^3}{x^2+y^2}\right| \leq \left|\frac{x^3}{x^2}\right| = |x| \to 0 \implies \lim_{(x,y)\to(0,0)} \frac{x^3}{x^2+y^2} = 0. \tag{1.12}$$

Example 1.24. We solve the limit $\lim_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{x^2+y^2}$. Simply rewriting $z = x^2 + y^2$ gives us $z \to 0$ as $(x,y) \to (0,0)$, so $\lim_{z\to 0} \frac{\sin z}{z} = 1$.

Before the next example, we write down a few properties. Let $S \subseteq \mathbb{R}^n$, let $f, g : S \to \mathbb{R}$ be functions, and let $a \in \mathbb{R}^n$ be a limit point of S. Suppose $\lim_{x\to a} f(x) = \alpha$ and $\lim_{x\to b} g(x) = \beta$. Then,

1. $\lim_{x\to a}(cf(x) + g(x)) = c\alpha + \beta$ for all $c \in \mathbb{R}$,

2. $\lim_{x\to a} f(x)g(x) = \alpha\beta$,

3. $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\alpha}{\beta}$, provided that $\beta \neq 0$,

4. if $f(x) \leq h(x) \leq g(x)$ for all $x \in S$ and if $\alpha = \beta$, then $\lim_{x\to a} h(x)$ exists and equals α.

A similar set of corresponding statements also hold true for continuous functions. Note that the function $\Pi_i : \mathbb{R}^n \to \mathbb{R}$ is also continuous.

Example 1.25. The function $f(x,y) = \frac{\sin(x^2+y^2)}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $f(x,y) = 1$ otherwise is a continuous function since it has been assigned its limit at $(x,y) = (0,0)$. However, the function $f(x,y) = \frac{2xy}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $f(x,y) = \alpha$ otherwise is continuous only at $\mathbb{R}^2 \setminus \{(0,0)\}$ for all $\alpha \in \mathbb{R}$.

Example 1.26. We look at the continuity of the function

$$f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{1.13}$$

Then, we have

$$\left| \frac{xy}{\sqrt{x^2+y^2}} \right| \leq \frac{1}{2} \cdot \frac{x^2+y^2}{\sqrt{x^2+y^2}} = \frac{1}{2} \|(x,y)\| \tag{1.14}$$

which shows that $\lim_{(x,y)\to(0,0)} f(x,y) = 0 = f(0,0)$. Thus, f is continuous on $\mathbb{R}^2$.

Example 1.27. Set $\mathcal{D} = \{(x,y) \in \mathbb{R}^2 \mid y \neq 0\}$. Since $\mathcal{D} = (\Pi_2^{-1}(\{0\}))^c$, $\mathcal{D}$ is an open set. Define $f : \mathcal{D} \to \mathbb{R}$ by $f(x,y) = x \sin \frac{1}{y}$. Here, we simply work as

$$|f(x,y)| = \left| x \sin \frac{1}{y} \right| \leq |x| \text{ on } \mathcal{D}. \tag{1.15}$$

Thus, the limit becomes $f(x,y)$.

Hereforth, O_n denotes an open subset of $\mathbb{R}^n$.

Remark 1.28. Let $(a,b) \in \mathbb{R}^2$ be a limit point of O_2. Suppose $\lim_{(x,y)\to(a,b)} f(x,y)$ exists and equals $\alpha \in \mathbb{R}$. It is natural to ask whether

$$\alpha = \lim_{y\to b} \lim_{x\to a} f(x,y) = \lim_{x\to a} \lim_{y\to b} f(x,y). \tag{1.16}$$

For someone looking at multivariable limits for the first time, it is tempting to believe this holds true always. We leave this question unanswered for now, and come back to it later.

A notion of uniform continuity may also be explored.

Definition 1.29. A function $f : S \to \mathbb{R}$, for $S \subseteq \mathbb{R}^n$, is said to be a *uniformly continuous* function if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|f(x) - f(y)\| < \varepsilon \text{ for all } \|x - y\| < \delta \text{ in } S. \tag{1.17}$$

We urge the reader to compute examples for uniformly continuous functions. The exercise of uniform contiuity implying continuity but not the other way around is left as an exercise to the reader.

1.3 Differentiability

As a little convention, for any $f : O_n \to \mathbb{R}^m$, we prefer to rewrite it as $f = (f_1, \dots, f_n)$ where $f_j = \Pi_j f$. We now ask the question of derivatives; what does it mean for the derivative of a function $f : O_n \to \mathbb{R}^m$? What about $f'(a)$ for some $a \in O_n$?

For the case of $n = m = 1$, we recall that f is termed differentiable at a if and only if $\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$ exists. If the limit is λ, then this limit exists if and only if $\lim_{h\to 0} \frac{f(a+h)-f(a)-h\lambda}{h}$, which is really a function of h. Thus, the function $h \mapsto \lambda h$ matters the most, that is, $L : \mathbb{R} \to \mathbb{R}$ where $Lh = \lambda h$. So, we can twist our words a little and say that f is differentiable at a if and only if there exists a linear map $L : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Lh}{h} = 0. \tag{1.18}$$

In this case, $f'(a) = L1 = \lambda$. We translate this exact idea into higher dimensions.

Definition 1.30. Let $f : O_n \to \mathbb{R}^m$. We say that f is *differentiable* at $a \in O_n$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, that depends on a, such that

$$\lim_{h\to 0} \frac{1}{\|h\|} (f(a+h) - f(a) - Lh) = 0. \tag{1.19}$$

In this case, we write $Df(a) = L$ and call it the total derivative of f at a. We say f is differentiable on O_n if f is differentiable at a for all $a \in O_n$.

Observe that the above limit is equivalent to saying that $\lim_{h\to 0} \frac{1}{\|h\|} \|f(a+h) - f(a) - Lh\| = 0$. Note that $Df(a) = L$ is unique. To show this, suppose there exists another linear map $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that $\lim_{h\to 0} \frac{1}{\|h\|} \left\|f(a+h) - f(a) - \tilde{L}h\right\| = 0$. Let there exist $h_0 \in \mathbb{R}^n$ such that $Lh_0 \neq \tilde{L}h_0$ and $\|h_0\| = 1$. Define $h : \mathbb{R} \to \mathbb{R}^n$ by $ht = th_0$. Then as $t \to 0$, $ht \to 0$. Therefore,

$$\left\|L(h(t)) - \tilde{L}(h(t))\right\| \leq \|f(a+h) - f(a) - Lh(t)\| + \left\|f(a+h) - f(a) - \tilde{L}h(t)\right\| \tag{1.20}$$

$$\implies \lim_{t\to 0} \frac{\left\|Lh(t) - \tilde{L}h(t)\right\|}{\|h(t)\|} = 0 \implies \lim_{t\to 0} \frac{1}{|t|} |t| \cdot \left\|Lh_0 - \tilde{L}h_0\right\| = 0 \implies Lh_0 = \tilde{L}h_0 \tag{1.21}$$

which is a contradiction.

Example 1.31. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. In this case, $\frac{f(a+h)-f(a)-f(h)}{\|h\|} \to 0$ as $\|h\| \to 0$. Thus, f is differentiable at a and $Df(a) = f$ for all $a \in \mathbb{R}^n$.

Example 1.32. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ be defined as $f(x) = c$ for all $x \in \mathbb{R}^n$. Then, we simply have $Df(a) = 0$, the null linear mapping.

We now truly ask how to compute $Df(a)$. Observe that $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Then, one can represent it as a matrix $Df(a) \in M_{m\times n}(R)$.

July 30th.

Theorem 1.33. *Let $f : O_n \to \mathbb{R}^m$ be a function. Then f is differentiable at $a \in O_n$ if and only if $f_i : O_n \to \mathbb{R}$ is differentiable at a for all $i = 1, 2, \dots, n$. Moreover, in this case,*

$$[Df(a)]_{m\times n} = \begin{bmatrix} [Df_1(a)]_{1\times n} \\ \vdots \\ [Df_m(a)]_{1\times n} \end{bmatrix}_{m\times n}. \tag{1.22}$$

From the above theorem it is clear that $D\Pi_i f = \Pi_i Df$. We now provide a proof.

Proof. For the forward implication, let f be differentiable at $a \in O_n$. Set $L := Df(a)$ and $L_i := \Pi_i Df(a)$. Note that $L_i : \mathbb{R}^n \to \mathbb{R}$ since $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ and $\Pi_i : \mathbb{R}^m \to \mathbb{R}$. Observe

$$f(a+h) - f(a) - Df(a)h = (\tilde{f}_1(h), \tilde{f}_2(h), \ldots, \tilde{f}_m(h)) \tag{1.23}$$

where $\tilde{f}_i(h) = f_i(a+h) - f_i(a) - L_i h$ for $i = 1, 2, \ldots, m$. Thus, for all $i = 1, 2, \ldots, m$,

$$\left|\tilde{f}_i(h)\right| \leq \left(\sum_{j=1}^{m} \left|\tilde{f}_j(h)\right|^2\right)^{1/2} = \|f(a+h) - f(a) - Lh\|. \tag{1.24}$$

Dividing by $\|h\|$ and taking $h \to 0$, we have

$$\lim_{h \to 0} \frac{\left|\tilde{f}_i(h)\right|}{\|h\|} = 0 \tag{1.25}$$

which shows that f_i is differentiable with $Df_i(a) = L_i$ $(= \Pi_i Df(a))$.

For the converse, let f_i be differentiable at a for all $i = 1, 2, \ldots, m$ and set $L_i = Df_i(a) : \mathbb{R}^n \to \mathbb{R}$. Set $L = \begin{bmatrix} L_1 \\ \vdots \\ L_m \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^m$, a linear map. Therefore,

$$\frac{1}{\|h\|} \|f(a+h) - f(a) - Lh\| = \frac{1}{\|h\|} \left(\sum_{j=1}^{m} \left|\tilde{f}_i(h)\right|^2\right)^{1/2} \to 0. \tag{1.26}$$

where $\tilde{f}_i(h) = f_i(a+h) - f_i(a) - L_i h$ for all i. ■

Corollary 1.34. *Let $f : O_1 \to \mathbb{R}^m$ be a function. f is, then, differentiable at $ain O_1$ if and only if f_i is differentiable at a for all $1 \leq i \leq m$. Moreover, in this case,*

$$Df(a) = \begin{bmatrix} f_1'(a) \\ \cdots \\ f_m'(a) \end{bmatrix}. \tag{1.27}$$

This is just a special case when $n = 1$.

Remark 1.35. Let $f : O_n \to \mathbb{R}^m$ be differentiable at a. Then f is continuous at a.

Proof. We have

$$\begin{aligned} \|f(x) - f(a)\| &\leq \|f(x) - f(a) - (Df(a))(x-a)\| + \|(Df(a))(x-a)\| \\ &\leq \frac{1}{\|x-a\|} \|f(x) - f(a) - (Df(a))(x-a)\| \cdot \|x-a\| + M\|x-a\| \to 0 \end{aligned} \tag{1.28}$$

as $x \to a$. Note that such an $M > 0$ exists because $Df(a)$ is a linear map. ■

1.3.1 Chain Rule

To simplify our study of derivatives in highder dimensions, we look at the so called chain rule. This will prove to be a very important tool to study the differentiability of any function of several variables.

Theorem 1.36 (The *chain rule*)**.** *Let $f : O_n \to O_m$ be differentiable at $a \in O_n$, and $g : O_m \to \mathbb{R}^p$ be differentiable at $b = f(a) \in O_m$. Then $g \circ f : O_n \to \mathbb{R}^p$ is differentiable at $a \in O_n$, and*

$$(Dg \circ f)(a) = Dg(f(a)) \circ Df(a). \tag{1.29}$$

For the proof, we will denote $A := Df(a)$ and $B := Dg(f(a))$, and $b = f(a)$. Moreover, we will write $r_f(x) := f(x) - f(a) + Df(a)(x-a)$.

Proof. There exists r_f in the neighbourhood of a and r_g in the neighbourhood of b such that $r_f(x) = f(x) - f(a) - A(x-a)$ and $r_g(y) = g(y) - g(b) - B(y-b)$. Now set

$$r(x) = g(f(x)) - g(b) - BA(x-a). \tag{1.30}$$

We claim that $\lim_{x\to a} \frac{r(x)}{\|x-a\|} = 0$. We know that $\lim_{x\to a} \frac{\|r_f(x)\|}{\|x-a\|} = 0 = \lim_{y\to b} \frac{\|r_g(y)\|}{\|y-b\|}$. Now,

$$\begin{aligned} r(x) &= g(f(x)) - g(b) - B(A(x-a)) = g(f(x)) - g(b) + B(r_f(x) - f(x) + f(a)) \\ &= (g(f(x)) - g(f(a)) - B(f(x) - f(a))) + Br_f(x) = r_g(f(x)) + Br_f(x). \end{aligned}$$

We show that both terms on the right hand side, when divided by $\|x-a\|$, tend to zero. Now,

$$\frac{\|Br_f(x)\|}{\|x-a\|} \le M \frac{\|r_f(x)\|}{\|x-a\|} \to 0. \tag{1.31}$$

For the remaining term, we work as follows: for $\varepsilon > 0$ fixed, there exists a $\delta > 0$ such that $\|r_g(y)\| < \varepsilon \|y-b\|$ for all $0 < \|y-b\| < \delta$. By continuity of f at a, there exists a $\tilde{\delta} > 0$ such that $\|f(x) - f(a)\| < \delta$ for all $\|x-a\| < \tilde{\delta}$. Thus, for all $0 < \|x-a\| < \tilde{\delta}$, we have

$$\|r_g(f(x))\| < \varepsilon \|f(x) - f(a)\| < \varepsilon \|r_f(x) + A(x-a)\| \tag{1.32}$$

$$\implies \|r_g(f(x))\| < \varepsilon (\|r_f(x)\| + \|A(x-a)\|) < \varepsilon \left(\|r_f(x)\| + \tilde{M}\|x-a\|\right)$$

$$\implies \frac{\|r_g(f(x))\|}{\|x-a\|} \to 0 \text{ as } x \to a. \tag{1.33}$$

■

August 4th.

The derivative also satisfies nice properties in higher dimensions.

Proposition 1.37. *Let $f, g : O_n \to \mathbb{R}^m$ be differentiable at $a \in O_n$. Then*

1. *$D(\alpha f + g)(a) = \alpha Df(a) + Dg(a)$ for all $\alpha \in \mathbb{R}$.*
2. *If $m = 1$, then $(f \times g)'(a) = f(a)g'(a) + f'(a)g(a)$.*
3. *If $m = 1$, then $\left(\frac{f}{g}\right)'(a) = \frac{1}{(g(a))^2}(f'(a)g(a) - f(a)g'(a))$.*

Proof. The proof is left as an exercise to the reader. ■

1.4 Partial Derivatives

Given $a \in O_n$ and $f : O_n \to \mathbb{R}$, define $\eta_i : (-\varepsilon, \varepsilon) \to \mathbb{R}$ for $i = 1, 2, \ldots, n$ by

$$\eta_i(t) = f(a + te_i). \tag{1.34}$$

We define $\frac{\partial f}{\partial x_i}(a) = \frac{d\eta_i}{dt}(0)$, if the latter term is defined. This is called the *partial derivative* of f in the direction, or with respect to, x_i at a. Therefore,

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t\to 0} \frac{f(a + te_i) - f(a)}{t}. \tag{1.35}$$

Remark 1.38.

1. Note that $\frac{\partial f}{\partial x_i}$ is easy to compute since we are essentially holding all other x_j's with $j \neq i$ as constant and just differentiating f with respect to x_i.
2. $\frac{\partial f}{\partial x_i}$ is the total derivative of f with the limit taken in the x_i-direction.

We want $Df(a)$ for a function $f : O_n \to \mathbb{R}$. We show further that the idea of partial derivatives solves this issues of computing $Df(a)$.

Definition 1.39. $f : O_n \to \mathbb{R}$ is termed a function in $C^1(O_n)$ if $\frac{\partial f}{\partial x_i}$ for all $i = 1, 2, \ldots, n$ exists and $x \mapsto \frac{\partial f}{\partial x_i}(x)$ is continuous on O_n.

We discuss some examples.

Example 1.40. Let $f(x, y) = x^3 + y^4 + \sin(xy)$ on $\mathbb{R}^2$. Taking y as a constant, $x \mapsto f(x, y)$ is differentiable. Thus,

$$\frac{\partial f}{\partial x} = 3x^2 + y\cos(xy). \tag{1.36}$$

Similarly,

$$\frac{\partial f}{\partial y} = 4y^3 + x\cos(xy). \tag{1.37}$$

The above two partial derivatives are continuous functions. Thus, $f \in C^1(\mathbb{R}^2)$.

Example 1.41. Recall the function

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \tag{1.38}$$

which was discontinuous at the origin. At the origin, we have

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{t \to 0} \frac{f(t, 0) - f(0, 0)}{t} = 0. \tag{1.39}$$

Similarly,

$$\frac{\partial f}{\partial y}(0, 0) = 0. \tag{1.40}$$

Thus we conclude that the partial derivatives exist at $(0, 0)$. But f is not continuous at $(0, 0)$. Thus, we conclude that the existence of partial derivatives does *not* imply the existence of the total derivative.

1.4.1 Higher Order Partials

Hereforth, we will denote $f_{x_i} := \frac{\partial f}{\partial x_i}$. Let us assume that the partials f_{x_i} exist for all $i = 1, 2, \ldots, n$. Therefore, $f_{x_i} : O_n \to \mathbb{R}$ are functions. Define

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial}{\partial x_j}(f_{x_i}) \text{ for all } j = 1, 2, \ldots, n. \tag{1.41}$$

This is known as the second order partial derivative. We may write this as $f_{x_i x_j} = (f_{x_i})_{x_j}$. Similarly, $\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}$ may be defined as $\frac{\partial}{\partial x_i}(f_{x_j x_k})$. This idea can be extended even further into higher dimensions.

Example 1.42. Define $f(x, y) = \sin x + e^y + xy$. Then $f_x = \cos x + y$ and $f_y = e^y + x$. From here, we further have $f_{xy} = 1$ and $f_{yx} = 1$. Coincidentally, we have $f_{xy} = f_{yx}$. Thus, we question whether the order of the variables even matters.

The following example shows that the order of the variables does matter.

Example 1.43. Define

$$f(x, y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases} \tag{1.42}$$

In this case, we get $f_{xy}(0,0) = 1 \neq f_{yx}(0,0) = -1$. Partial differentiation is not commutative.

The following result shows that with a little more constraints, the commutativity does hold.

Theorem 1.44 (*Clairaut's theorem*)**.** *Let $(a,b) \in O_2$, $f : O_2 \to \mathbb{R}$, and assume that f_{xy} and f_{yx} exist on O_2. Also suppose that f_{xy} is continuous at (a,b). Then $f_{xy}(a,b) = f_{yx}(a,b)$.*

The result does hold in higher dimensions too, but we only show for two dimensions. The rest of the theorem and proof are left as exercises for the reader.

Proof. Without the loss of generality, let $(a,b) = (0,0)$ and $O_2 = B_1(0,0)$. Choose $h,k > 0$ such that $[0,h] \times [0,k] \subseteq B_1(0,0)$. Then,

$$\begin{aligned}
\frac{\partial^2 f}{\partial y \partial x} &= \lim_{k\to 0} \frac{1}{k}\left(\frac{\partial f}{\partial x}(x, y+k) - \frac{\partial f}{\partial x}(x,y)\right) && (1.43)\\
&= \lim_{k\to 0} \frac{1}{k} \lim_{h\to 0} \frac{1}{h}\left(f(x+h,y+k) - f(x,y+k) - f(x+h,y) + f(x,y)\right)\\
&= \lim_{k\to 0} \lim_{h\to 0} \frac{1}{hk}\left(f(x+h,y+k) - f(x,y+k) - f(x+h,y) + f(x,y)\right) && (1.44)\\
\implies \frac{\partial^2 f}{\partial y \partial x}(0,0) &= \lim_{k\to 0} \lim_{h\to 0} \frac{1}{hk}\left(f(h,k) - f(0,k) - f(h,0) + f(0,0)\right) := \lim_{k\to 0} \lim_{h\to 0} \frac{1}{hk} F(h,k). && (1.45)
\end{aligned}$$

Similarly,

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h\to 0} \lim_{k\to 0} \frac{1}{hk} F(h,k). \tag{1.46}$$

Fix k and h for a moment. Set $f_1(x) = f(x,k) - f(x,0)$ for all $x \in [0,h]$. Then f_1 is continuous on $[0,h]$, since f_x exists, and is differentiable on $(0,h)$. By the mean value theorem, there exists $c_1 \in (0,h)$ such that $f_1(h) - f_1(0) = f_1'(c_1) \times h_1$

$$\implies \frac{1}{h} F(h,k) = \left(\frac{\partial f}{\partial x}(c_1,k) - \frac{\partial f}{\partial x}(c_1,0)\right). \tag{1.47}$$

Next, consider $f_2(y) = \frac{\partial f}{\partial x}(c_1,y)$ for all $y \in [0,k]$ which is continuous on $[0,k]$ and differentiable on $(0,k)$. Again, by the mean value theorem, there exists $c_2 \in (0,k)$ such that $f_2(k) - f_2(0) = f_2'(c_2) \times k$

$$\implies \frac{1}{hk} F(h,k) = \frac{\partial^2 f}{\partial y \partial x}(c_1,c_2) \tag{1.48}$$

with $0 < c_1 < h$ and $0 < c_2 < k$. Similarly, if we had redefined f_1 and f_2, we would have received

$$\frac{1}{hk} F(h,k) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{c}_1,\tilde{c}_2) \tag{1.49}$$

with $0 < \tilde{c}_1 < h$ and $0 < \tilde{c}_2 < k$. Thus,

$$\frac{\partial^2 f}{\partial y \partial x}(c_1,c_2) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{c}_1,\tilde{c}_2). \tag{1.50}$$

As $(h,k) \to (0,0)$, $f_{xy}(0,0) = f_{yx}(0,0)$. ∎

August 6th.

Theorem 1.45 (*Schwarz theorem*)**.** *Let $f : O_2 \to \mathbb{R}$ be a function such that $(0,0) \in O_2$. Also suppose that f_x, f_y, f_{xy} exist on O_2 and f_{xy} is continuous on O_2. Then $f_{yx}(0,0)$ exists and $f_{yx}(0,0) = f_{xy}(0,0)$.*

Proof. As f_{xy} is continuous at $(0,0)$, for $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f_{xy}(s,t) - f_{xy}(0,0)| < \varepsilon \text{ for all } \sqrt{s^2+t^2} < \delta. \tag{1.51}$$

We already know $F(h,k) = f_{xy}(c_1,c_2)$ with $0 < c_1 < h$ and $0 < c_2 < k$. Choose h,k small enough such that $\sqrt{h^2+k^2} < \delta$. Therefore, for $\sqrt{c_1^2+c_2^2} < \delta$,

$$|f_{xy}(c_1,c_2) - f_{xy}(0,0)| < \varepsilon \implies |F(h,k) - f_{xy}(0,0)| < \varepsilon \text{ for } \sqrt{h^2+k^2} < \varepsilon. \tag{1.52}$$

From the above, we infer $-\varepsilon + f_{xy}(0,0) < F(h,k) < \varepsilon + f_{xy}(0,0)$. Rewriting the middle term,

$$F(h,k) = \frac{1}{h}\left(\frac{f(h,k)-f(h,0)}{k} - \frac{f(0,k)-f(0,0)}{k}\right) \xrightarrow{k\to 0} \frac{1}{h}(f_y(h,0) - f_y(0,0)) \tag{1.53}$$

Rebounding gives us

$$\left|\frac{1}{h}(f_y(h,0) - f_y(0,)) - f_{xy}(0,0)\right| \le \varepsilon \implies \lim_{h\to 0}\frac{1}{h}(f_y(h,0) - f_y(0,0)) = f_{xy}(0,0). \tag{1.54}$$

■

Note that in the above theorem, $(0,0)$ was chosen for the sake of simplifying the proof. Any $(\alpha,\beta) \in O_2$ would have worked. We move our focus back to the total derivative.

Theorem 1.46. *Let $f : O_n \to \mathbb{R}^m$ be differentiable at $a \in O_n$. Then $\frac{\partial f_i}{\partial x_j}$ exists at a for all $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. Moreover,*

$$\big[Df(a)\big]_{m\times n} = \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{m\times n}. \tag{1.55}$$

Proof. Let $m = 1$, and fix $j \in \{1,2,\ldots,n\}$. Suppose $a = (a_1,\ldots,a_j,\ldots,a_n)$. Consider the mapping $\eta_j : (a_j-\varepsilon, a_j+\varepsilon) \to \mathbb{R}^n$ defined as $x \mapsto (a_1,\ldots,a_{j-1},x,a_{j+1},\ldots,a_n)$. Since this image is in the neighbourhood of a, we can apply f on it to get an image in $\mathbb{R}$; the mapping maps x to $f(a_1,\ldots,a_{j-1},x,a_{j+1},\ldots,a_n)$.

As $x \to a_1$, $x \to a_2$, …, $x \to x$, $x \to a_{j+1}$, … differentiable on $(a_j-\varepsilon, a_j+\varepsilon)$, η_j is differentiable on a_j and $\eta_j'(a_j) = e_j$. Thus, $f \circ n_j$ is differentiable at a_j by the chain rule.

$$(Df \circ n_j)(a_j) = \frac{d}{dx}(f\circ n_j)(a_j) = \lim_{h\to 0}\frac{f(\eta_j(a_j+h)) - f(a)}{h} = \lim_{h\to 0}\frac{f(a_1,\ldots,a_j,\ldots,a_n) - f(a)}{h} = \frac{\partial f}{\partial x_j}(a). \tag{1.56}$$

The chain rule implies that $(Df \circ n_j)(a_j) = Df(\eta_j(a_j))D\eta_j(a_j) \implies \frac{\partial f}{\partial x_j}(a) = Df(a)e_j$. Thus, we must have

$$\big(Df(a)\big) = \begin{pmatrix} f_{x_1(a)} & \cdots & f_{x_n}(a)\end{pmatrix}. \tag{1.57}$$

We now work the case for a general m; write f as $(f_1,\ldots,f_m)$. f is differentiable at a implies that

$$\big[Df(a)\big] = \begin{bmatrix} Df_1(a) \\ \vdots \\ Df_m(a)\end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)\end{bmatrix}. \tag{1.58}$$

■

Definition 1.47. Let $f : O_n \to \mathbb{R}^m$ be differentiable at a. The matrix representation $\left(\frac{\partial f_i}{\partial x_j}(a)\right)_{m\times n}$ of the total derivative $Df(a)$ is termed the *Jacobian* of f at a. We prefer to write it as $J_f(a)$.

Since we have a matrix to deal with now, it is only natural to ask questions regarding its nature. For instance, what does the rank of the Jacobian tell us? What about its determinant?

Theorem 1.48. *Let $f : O_n \to \mathbb{R}^m$ be a function with $a \in O_n$. Suppose f is a C^1 function in the neighbourhood of a. Then f is differentiable at a.*

There is a *gap* between this theorem and the previous one; we have the extra requirement of continuity of the partial derivatives here.

August 11th.

Example 1.49. Consider the function

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\frac{1}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{1.59}$$

Then f is continuous at $(0,0)$. Moreover, one can show that f is also differentiable at $(0,0)$ with $Df(0,0) = \begin{bmatrix} 0 & 0 \end{bmatrix}$. However, both f_x and f_y are not continuous at $(0,0)$.

We now provide the proof of the above theorem, setting $a = 0$ without the loss of generality.

Proof. Let $m = 1$; the general case will be handled later. We claim that

$$\lim_{h\to 0} \frac{1}{\|h\|}\left| f(h) - f(0) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(0)h_i \right| = 0. \tag{1.60}$$

For $h \in \mathbb{R}^n$, with $\|h\|$ sufficiently small, we write $\hat{h}_i = (h_1, h_2, \ldots, h_i, 0, \ldots, 0)$ with $(n-i)$ zeroes at the end, for all $i = 1, 2, \ldots, n$, and $\hat{h}_0 = (0, \ldots, 0)$. Therefore,

$$\begin{aligned} f(h) - f(0) = f(\hat{h}_n) - f(\hat{h}_0) &= (f(\hat{h}_1) - f(\hat{h}_0)) + (f(\hat{h}_2) - f(\hat{h}_1)) + \cdots + (f(\hat{h}_n) - f(\hat{h}_{n-1})) \\ &= \sum_{i=1}^{n} (f(\hat{h}_i) - f(\hat{h}_{i-1})). \end{aligned} \tag{1.61}$$

Define $\eta_i(t) = f(h_1, \ldots, h_{i-1}, t, 0, \ldots, 0)$ for $t \in [0, h_i]$. Thus, each η_i is a single variable function and the chain rule tells us $\eta_i : [0, h_i] \to \mathbb{R}$ is a C^1-function. The mean value theorem then tell us that there exists $c_i \in (0, h_i)$ such that

$$\eta_i(h_i) - \eta_i(0) = h_i \eta_i'(c_i) \tag{1.62}$$

$$\implies f(\hat{h}_i) - f(\hat{h}_{i-1}) = h_i \frac{\partial f}{\partial x_i}(h_1, \ldots, h_{i-1}, c_i, 0, \ldots, 0). \tag{1.63}$$

Therefore,

$$\begin{aligned} \frac{1}{\|h\|}\left| f(h) - f(0) - \sum_{i=1}^{n} f_{x_i}(0)h_i \right| &= \frac{1}{\|h\|}\left| \sum_{i=1}^{n} h_i \left(\frac{\partial f}{\partial x_i}(h_1, \ldots, h_{i-1}, c_i, 0, \ldots, 0) - f_{x_i}(0) \right) \right| \\ &\leq \sum_{i=1}^{n} \frac{|h_i|}{\|h\|} \left| f_{x_i}(h_1, \ldots, h_i, c_i, 0, \ldots, 0) - f_{x_i}(0) \right| \xrightarrow{h\to 0} 0. \end{aligned} \tag{1.64}$$

Hence, the function f is differentiable at 0. ∎

Example 1.50. We compute the total derivative of $f(x,y,z) = (x+2y+3z, xyz, \cos x, \sin x)$. Clearly, f is a C^1 function. Thus,

$$J_f(x,y,z) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \\ \frac{\partial f_4}{\partial x} & \frac{\partial f_4}{\partial y} & \frac{\partial f_4}{\partial z} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ yz & xz & xy \\ -\sin x & 0 & 0 \\ \cos x & 0 & 0 \end{bmatrix}. \tag{1.65}$$

1.5 Gradient and The Chain Rule

We first extend the notion of the derivative along the coordinate directions to the derivative along *any* direction.

Definition 1.51. Let $u \in \mathbb{R}^n$ be a unit vector, that is, $\|u\| = 1$. Also let $f : O_n \to \mathbb{R}$ be a function. The *directional derivative* of f at $x \in O_n$ in the direction of u is defined as

$$D_u f(x) = \lim_{t\to 0} \frac{f(x+tu) - f(x)}{t}, \text{ if exists.} \tag{1.66}$$

The reader may verify that $D_{e_i} f(x) = f_{x_i}(x)$. If we denote $t \mapsto f(x+tu)$ as $\eta(t)$, then we directional derivative is simply $D_u f(x) = \eta'(0)$. If we additionally assume that f is differentiable at x, then the chain rule gives us

$$D_u f(x) = \eta'(0) = Df(x) \circ u. \tag{1.67}$$

Notice that u can be thought of as a linear map from $\mathbb{R}$ to $\mathbb{R}^n$ and $Df(x)$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}$.

Theorem 1.52. *Let $f : O_n \to \mathbb{R}$ be differentiable at $x \in O_n$ and let u be a unit vector in $\mathbb{R}^n$. Then the directional derivative $D_u f(x)$ exists and is given by*

$$(D_u f)(x) = Df(x)u. \tag{1.68}$$

The idea of the gradient is introduced.

Definition 1.53. Let $f : O_n \to \mathbb{R}$ be a function with $a \in O_n$. Also suppose $f_{x_i}(a)$ exists for all $i = 1, 2, \ldots, n$. Then

$$(\nabla f)(a) = (f_{x_1}(a), \ldots, f_{x_n}(a)) \tag{1.69}$$

is called the *gradient* of f at a. Therefore,

$$\nabla : \{f : \mathbb{R}^n \to \mathbb{R} \mid f_{x_i} \text{ exists}\} \to \mathbb{R}^n. \tag{1.70}$$

Corollary 1.54. *Let $f : O_n \to \mathbb{R}$ be differentiable at $x \in O_n$ and let u be a unit vector. Then, $(D_u f)(x) = (\nabla f)(x) \cdot u$.*

Remark 1.55. Suppose $f : O_n \to \mathbb{R}$ is a function whose partial derivatives exist at $x \in O_n$. Then

$$(\nabla f)(x) \cdot u = \|(\nabla f)(x)\| \cdot \|u\| \cos\theta = \|(\nabla f)(x)\| \cos\theta. \tag{1.71}$$

As $|\cos\theta| \le 1$, the right hand side is maximum if $\theta = 0$.

August 13th.

What follows is an important result.

Theorem 1.56. *Let $f : O_n \to \mathbb{R}$ be differentiable at $x \in O_n$ and suppose $(\nabla f)(x) \neq 0$. Then the vector $(\nabla f)(x)$ points in the direction of the steepest ascent of the function f at x and $\|(\nabla f)(x)\|$ is the greatest possible rate of change.*

In other words, the maximum possible directional derivative of f at a occurs at $\nabla f(a)$.

Example 1.57. Let $f(x, y, z) = x^2yz$. We find the directional derivatives of f at $(1, 1, 0)$ in the direction of $\langle 1, 1, -1\rangle$. Note that the unit vector in this case is $u = \frac{1}{\sqrt{3}}\langle 1, 1, -1\rangle$, and that f is differentiable since f is a polynomial. Thus, $f_x = 2xyz$, $f_y = x^2z$, and $f_z = x^2y$. Therefore,

$$\nabla f = \langle 2xyz, x^2z, x^2y\rangle \implies \nabla f(1,1,0) = \langle 0,0,1\rangle. \tag{1.72}$$

Finally, we get

$$D_u f(1,1,0) = \langle 0,0,1\rangle \cdot \frac{1}{\sqrt{3}}\langle 1,1,-1\rangle = -\frac{1}{\sqrt{3}}. \tag{1.73}$$

Also, the maximum possible derivative of f at $(1,1,0)$ occurs at $\langle 0,0,1\rangle$ and it is $\langle 0,0,1\rangle \cdot \langle 0,0,1\rangle = 1$.

Example 1.58. We look at

$$f(x,y) = \begin{cases} \frac{x^2y}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{1.74}$$

Clearly, $|f(x,y) - f(0,0)| = \frac{x^2|y|}{x^2+y^2} \leq |y|$, so f is continuous at $(0,0)$. For a given direction $u = \langle u_1, u_2\rangle$, we have

$$D_u f(0,0) = \lim_{t\to 0} \frac{f(tu_1, tu_2) - f(0,0)}{t} = \lim_{t\to 0} \frac{t^3u_1^2u_2}{t^2(u_1^2+u_2^2)} \cdot \frac{1}{t} = u_1^2u_2 < \infty. \tag{1.75}$$

Therefore, $D_u f(0,0) = u_1^2u_2$. Also, $f_x(0,0) = 0 = f_y(0,0)$ implying that $\nabla f(0,0) = \langle 0,0\rangle$. This gives us

$$\nabla f(0,0) \cdot u = 0 \neq u_1^2u_2 = D_u f(0,0) \tag{1.76}$$

making f *not* differentiable at $(0,0)$.

Hereforth, given $a, b \in \mathbb{R}^n$, we denote $L_{a,b}$ to be the line segment joining a to b. Essentially, $L_{a,b} = \{(1-t)a + tb \mid 0 \leq t \leq 1\}$.

Theorem 1.59 (The *mean value theorem*)**.** *Let $f : O_n \to \mathbb{R}$ be differentiable and $L_{a,b} \subseteq O_n$. Then there exists $c \in L_{a,b}$ such that*

$$f(b) - f(a) = \nabla f(c) \cdot (b-a). \tag{1.77}$$

Proof. Define $\eta : [0,1] \to L_{a,b}$ as $\eta(t) = (1-t)a + tb$. Then $f \circ \eta : [0,1] \to \mathbb{R}$ is differentiable and $\eta'(t) = b - a$, a column vector. We apply the one-dimensional mean value theorem on $f \circ \eta$ to get

$$(f\circ\eta)(1) - (f\circ\eta)(0) = (f\circ\eta)'(t_0) \tag{1.78}$$

for some $t_0 \in (0,1)$. This implies that

$$f(b) - f(a) = f'(\eta(t_0)) \cdot \eta'(t_0) = \nabla f(\eta(t_0)) \cdot (b-a) = \nabla f(c) \cdot (b-a) \tag{1.79}$$

where $c = \eta(t_0) \in L_{a,b}$. ■

1.5.1 More Partials

Suppose we have $f : O_n \to O_m$ and $g : O_m \to \mathbb{R}^p$ where $a \in O_n$ and $b = f(a) \in O_m$. Assume that f and g are differentiable at a and b, respectively. Then

$$J_{g\circ f}(a) = J_g(b) J_f(a). \tag{1.80}$$

Comparing the $(i,j)^{\text{th}}$ entry of both sides (for $1 \leq i \leq p$ and $1 \leq j \leq n$), we get

$$\frac{\partial (g\circ f)_i}{\partial x_j}(a) = \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(b) \frac{\partial f_k}{\partial x_j}(a). \tag{1.81}$$

In a more natural way, set $y_k = f_k(x_1, \ldots, x_n)$ for $k = 1, \ldots, m$ and set $z_i = g_i(y_1, \ldots, y_m)$ for $i = 1, \ldots, p$. Then we can write

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial z_i}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial z_i}{\partial y_m} \cdot \frac{\partial y_m}{\partial x_j} = \sum_{t=1}^{m} \frac{\partial z_i}{\partial y_t} \cdot \frac{\partial y_t}{\partial x_j}. \tag{1.82}$$

This is termed the *chain rule for partials*.

Example 1.60. Suppose on $O_1 \to O_m \to \mathbb{R}$, we have the mapping(s) $t \mapsto (x_1(t), \ldots, x_m(t)) \mapsto f(x_1(t), \ldots, x_m(t))$. If we call the first mapping $\eta(t)$, and second mapping $(f(\eta(t)))$, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_m}\frac{dx_m}{dt} = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}. \tag{1.83}$$

Example 1.61. Suppose $f(x, y, z) = xy^2z$ with $x = t$, $y = e^t$, and $z = 1 + t$. Then

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = y^2z + 2xyze^t + xy = e^{2t}((1 + 2t) + 2t(1 + t)). \tag{1.84}$$

Of course, in this example, it would have been preferable to substitute back in the variables in terms of t and then derivating. This may not always be the case, especially in abstract computations.

Example 1.62. Suppose $g(z, w) = f(x(z, w), y(z, w))$. Making all the necessary assumptions, we have

$$\frac{\partial g}{\partial z} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z}. \tag{1.85}$$

Example 1.63. We introduce the idea of *polar coordinates*. $g(y, x)$ can be written as $f(r, \theta)$ where $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(\frac{y}{x})$. The converse may also be done. Here,

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial g}{\partial \theta}\frac{\partial \theta}{\partial x}. \tag{1.86}$$

Example 1.64. Suppose $z = z(u, v)$ with $u = x^2y$ and $v = 3x + 2y$. Then

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} = \frac{\partial z}{\partial u}x^2 + \frac{\partial z}{\partial v} \cdot 2. \tag{1.87}$$

One can reuse the chain rule and think of double partial differentiation too as

$$\frac{\partial^2 y}{\partial y^2} = x^2 \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial u}\right) + 2\frac{\partial}{\partial y}\left(\frac{\partial z}{\partial v}\right). \tag{1.88}$$

August 18th.

The Laplacian is introduced.

Definition 1.65. For $u \in C^2(O_n)$, the *Laplacian* of u is defined as

$$\Delta u := \nabla \circ \nabla u = \langle \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \rangle \cdot \langle u_{x_1}, \ldots, u_{x_n} \rangle = \sum_{j=1}^{n} \frac{\partial^2 u}{\partial x_j^2}. \tag{1.89}$$

Example 1.66. Let $x = r\cos\theta$ and $y = r\sin\theta$ and let $u := u(x, y)$. The Laplacian of u is $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$.

Now,

$$\frac{\partial u}{\partial r} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial u}{\partial y}\sin\theta. \tag{1.90}$$

In a case where u is considered in its polar coordinates, the Laplacian of u is also written in its polar coordinate form.

1.6 Extremum and Critical Points

Of course, one can extend the idea of the maximum (minimum) value of a function into higher dimensions.

Definition 1.67. Let $S_n \subseteq \mathbb{R}^n$ and let $a \in S_n$ be an interior point. We say $f : S_n \to \mathbb{R}$ attains a *local maximum* at a if there exists $r > 0$ such that $f(x) \leq f(a)$ for all $x \in B_r(a) \subseteq S_n$. Similarly, it attains a *local minimum* at a if there exists $r > 0$ such that $f(x) \geq f(a)$ for all $x \in B_r(a) \subseteq S_n$.
a is, instead, termed a *saddle point* of f if for all $r > 0$ satisfying $B_r(a) \subseteq S_n$, there exist $h_1, h_2 \in B_r(a)$ such that $f(h_1) > f(a)$ and $f(h_2) < f(a)$.

Similarly, one has critical points.

Definition 1.68. $a \in S_n$ is called a *critical point* of $f : S_n \to \mathbb{R}$ if $\nabla f(a) = 0$. In other words, $f_{x_i}(a) = 0$ for all $i = 1, 2, \ldots, n$.

Theorem 1.69. *Suppose $f : O_n \to \mathbb{R}$ is differentiable at $a \in O_n$. If a is a local extremum, then $\nabla f(a) = 0$.*

Proof. We claim that $f_{x_i}(a) = 0$ for each i. Fix i and define $\varphi_i(t) = f(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n)$ for $t \in (a_i - \varepsilon, a_i + \varepsilon)$. This implies that a_i is a local extremum of φ_i giving $\frac{d\varphi_i}{dt}(a_i) = 0$. Thus, $f_{x_i}(a) = 0$. ∎

To find these extrema, one would find the critical points and apply the second derivative test in the one variable case. In higher dimensions, we do the same; however, one needs to formulate the idea of a 'second derivative test' here.

Definition 1.70. Given $f \in C^2(O_n)$ with $a \in O_n$. The *Hessian matrix*, or Hessian, of f at a is defined as

$$H_f(a) = \left[\frac{\partial^2 f(a)}{\partial x_i \partial x_j}\right]_{n\times n} \tag{1.91}$$

Note that H_f is a symmetric matrix.

Example 1.71. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x, y) = \sin^2 x + x^2 y + y^2$. Then $f_x = \sin 2x + 2xy$, $f_y = x^2 + 2y$, and $f_{xy} = 2x$. The Jacobian in this case is

$$J_f = \begin{bmatrix} \sin 2x + 2xy & x^2 + 2y \end{bmatrix} \tag{1.92}$$

and the Hessian is

$$H_f = \begin{bmatrix} 2\cos 2x + 2y & 2x \\ 2x & 2 \end{bmatrix}. \tag{1.93}$$

Definition 1.72. Given $A = (a_{ij})_{n\times n} \in M_n(\mathbb{R})$, we define $Q_A(x) = x^t A x = \langle Ax, x\rangle = \sum_{i,j=1}^n a_{ij} x_i x_j$ for all $x \in \mathbb{R}^n$. A *quadratic form* is Q_A when A is symmetric.

Thus, $x^t H_f(a) x$ is a quadratic form which is also a homogenous polynomial of degree 2. One calls a symmetric matrix $A \in M_n(\mathbb{R})$ a *positive definite matrix* if $h^t A h > 0$ for all $h \in \mathbb{R}^n \setminus \{0\}$ and, likewise, a *negative definite matrix*. It is termed a *positive semidefinite matrix* if $h^t A h \geq 0$ for all $h \in \mathbb{R}^n$ and, likewise, a *negative semidefinite matrix*. It is called a *indefinite matrix* if it satisfies none of the above conditions.

Theorem 1.73. *Consider a symmetric matrix* $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \in M_2(\mathbb{R})$. *Then*

1. *A is positive definite if and only if* $a > 0$ *and* $ac - b^2 > 0$,
2. *A is negative definite if and only if* $a < 0$ *and* $ac - b^2 > 0$,
3. *A is indefinite if and only if* $ac - b^2 < 0$.

Lemma 1.74. *Let* $a \in O_n$, *and suppose* $A(x) = \begin{bmatrix} a_{11}(x) & a_{12}(x) \\ a_{21}(x) & a_{22}(x) \end{bmatrix} \in M_2(\mathbb{R})$ *is a symmetric matrix for all* $x \in O_n$. *Suppose A is continuous at a, that is, the functions* a_{ij} *are continuous at a for all pairs* i, j. *If* $A(a)$ *is positive definite, then A is positive definite in a neighbourhood of a.*

Proof. $A(a)$ is positive definite implies that $a_{11}(a) > 0$ and $a_{11}(a)a_{22}(a) - a_{12}(a)^2 > 0$. As $x \mapsto a_{ij}(x)$ is continuous at a, there exists a neighbourhood of a such that $a_{11}(x) > 0$ and $a_{11}(x)a_{22}(x) - a_{12}(x)^2 > 0$ for all x in that neighbourhood. Therefore, $A(x)$ is positive definite in a neighbourhood of a. ∎

Recall Taylor's polynomial and approximation; given a function $f \in C^k(I)$ with $a \in I \subseteq \mathbb{R}$, we would have

$$p_{a,k}(a+h) = \sum_{m=0}^{k} \frac{f^{(m)}(a)}{m!} h^m \tag{1.94}$$

where $a + h \in I$. In terms of x, it was

$$p_{a,k}(x) = \sum_{m=0}^{k} \frac{f^{(m)}(a)}{m!} (x-a)^m \tag{1.95}$$

One would also have Taylor's theorem, where if the above f were a $C^{k+1}(I)$-function, then for all $x \in I$ there exists a $\zeta \equiv \zeta(x, a)$ such that

$$f(x) = p_{a,k}(x) + \frac{f^{(k+1)}(\zeta)}{(k+1)!} (x-a)^{k+1}. \tag{1.96}$$

This theorem extends to higher dimensions.

Theorem 1.75 (*Taylor's theorem*)**.** *Let* $a \in O_n$ *where* O_n *is a convex set. Also suppose* $f : O_n \to \mathbb{R}$ *is a* C^{k+1}*-function. If* $h \in O_n$ *and* $a + h \in O_n$, *then*

$$f(a+h) = \sum_{|\alpha| \leq k} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha + \gamma_{a,k}(h) \tag{1.97}$$

where

$$\gamma_{a,k}(h) = \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(a+ch)}{\alpha!} h^\alpha \tag{1.98}$$

for some $c \in (0, 1)$.

We clear up some notation. Here, $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ is termed a multi-index, with $|\alpha| = \sum \alpha_i$. The notation $\partial^\alpha f(a)$ is shorthand for $\frac{\partial^{|\alpha|} f(a)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$. Also, $\alpha! = \alpha_1! \cdots \alpha_n!$, and finally, $h^\alpha = h_1^{\alpha_1} \cdots h_n^{\alpha_n}$.

Proof. Fix $a, a+h \in O_n$. Consider the mappings $t \mapsto a+th \mapsto f(a+th)$, with $\eta : [0,1] \to \mathbb{R}$ defined as $\eta(t) = f(a+th)$. Note that η is also a C^{k+1}-function, and $\eta'(t) = \nabla f(a+th) \circ h = \sum_{i=1}^n h_i f_{x_i}(a+th)$. The double derivatives is

$$\eta''(t) = \sum_{i=1}^n h_i f_{x_i}(a+th) = \nabla f_{x_i}(a+th) \circ h = \sum_{i,j=1}^n h_i h_j f_{x_i x_j}(a+th). \tag{1.99}$$

Call $\nabla \circ h = \sum_{i=1}^n h_i \frac{\partial}{\partial x_i}$. Therefore, $\eta'(t) = (\nabla \circ h) f(a+th)$ and $\eta''(t) = (\nabla \circ h)^2 f(a+th)$; we have

$$\eta^{(m)}(t) = (\nabla \circ h)^m f(a+th) \tag{1.100}$$

for all $0 \leq m \leq k+1$. Note that

$$(\nabla \circ h)^m = \sum_{|\alpha|=m} \frac{\partial^\alpha}{\alpha!} h^\alpha. \tag{1.101}$$

We apply the one-dimensional Taylor's theorem to get

$$\eta(1) = \eta(0) + \eta'(0) + \frac{1}{2!}\eta''(0) + \cdots + \frac{1}{k!}\eta^{(k)}(0) + \frac{1}{(k+1)!}\eta^{(k+1)}(c) \tag{1.102}$$

for some $c \in (0,1)$. Expanding in terms of f gives the desired result. ■

August 22nd.

Theorem 1.76 (The *min-max theorem*)**.** *Let $f \in C^2(O_2)$ with $a \in O_2$, and suppose $Df(a) = 0$. Consider the Hessian matrix $H_f(a) = \begin{bmatrix} f_{xx}(a) & f_{xy}(a) \\ f_{xy}(a) & f_{yy}(a) \end{bmatrix}$. Then*

1. *$f(a)$ is a local maximum if $f_{xx}(a) < 0$ and $\det H_f(a) > 0$,*
2. *$f(a)$ is a local minimum if $f_{xx}(a) > 0$ and $\det H_f(a) > 0$,*
3. *$f(a)$ is a saddle point if $\det H_f(a) < 0$.*

Proof. We show only for the first part; the rest follow a similar logic. Let $f_{xx}(a) > 0$ and $\det H_f(a) > 0$. Then $f_{xx}(x,y) > 0$ for all (x,y) in a neighbourhood of a. Now, $\det H_f(x,y) = f_{xx}(x,y) f_{yy}(x,y) - f_{xy}(x,y)^2$ implies that $(x,y) \mapsto \det H_f(x,y)$ is a continuous function. Thus $H_f(a) > 0$ will force $H_f(x,y) > 0$ for all (x,y) in a neighbourhood of a. Thus, for some $r > 0$, $f_{xx} > 0$ and $\det H_f > 0$ in $B_r(a)$. We conclude that H_f is positive definite on $B_r(a)$. Thus, for all $a+h \in B_r(a)$, we use Taylor's polynomial to get

$$f(a+h) - f(a) = \frac{1}{2} h^t H_f(a+ch) h > 0 \implies f(a+h) > f(a) \text{ for all } a+h \in B_r(a). \tag{1.103}$$

Thus, $f(a)$ is a local minimum. ■

Example 1.77. We find the critical points and discuss the nature of the function $f(x,y) = x^3 - 6x^2 - 8y^2$. We have

$$f_x = 3x^2 - 12x,\ f_y = -16y,\ f_{xx} = 6x - 12,\ f_{yy} = -16,\ f_{xy} = 0. \tag{1.104}$$

Setting $f_x = 0 = f_y$ gives $x = 0, 4$ and $y = 0$; the critical points are $(0,0)$ and $(4,0)$. Moreover, $H_f(x,y) = \begin{bmatrix} 6x-12 & 0 \\ 0 & -16 \end{bmatrix}$. At $(0,0)$, $\det H_f(0,0) > 0$ and $f_{xx}(0,0) < 0$, so $f(0,0)$ is a local maximum. At $(4,0)$, $\det H_f(4,0) < 0$, so $f(4,0)$ is a saddle point.

1.7 Compact Sets

The theory of compact sets proves to be useful in various areas of analysis.

Definition 1.78. A set $K \subseteq \mathbb{R}^n$ is termed a *compact set* if every sequence from K has a convergent subsequence whose limit lies in K.

Remark 1.79.

1. Compact sets are bounded sets; there exists an $M > 0$ such that $\|x\| \leq M$ for all $x \in K$. If it were not bounded, then for every $m \in \mathbb{N}$, one could find $x_m \in K$ such that $\|x_m\| > m$. The sequence $\{x_m\}$ does not have a convergent subsequence.
2. Compact sets are closed sets; let K be compact and let $\{x_m\} \subseteq K$ converge to x. Since K is compact, simply pick the subsequence $x_{m_k} = x_m$, which will force the limit point x to lie in K showing that K is closed.

One then naturally asks whether every closed and bounded subset is compact.

Theorem 1.80. *The box $K := \Pi_{i=1}^n [a_i, b_i]$ is compact.*

Proof. Pick $\{x_m\} \subseteq K$. Look at sequence of the i^{th} projection of the x_m's as $\{\Pi_i(x_m)\} \subseteq [a_i, b_i]$. By the one dimensional Bolzano-Weierstrass theorem, we can find a convergent subsequence such that $\Pi_i(x_{m_l}) \to \alpha_i$. Pick the intersection of all these convergent subsequences, the resulting subsequences will converge to $\alpha = (\alpha_1, \ldots, \alpha_n) \in K$. ■

Theorem 1.81. *Let $K \subseteq \mathbb{R}^n$. Then K is compact if and only if K is closed and bounded.*

Proof. We have already shown the forward direction. For the converse implication, since K is bounded, there exists $M > 0$ such that $K \subseteq [-M, M]^n$. Pick any $\{x_m\} \subseteq K$. As $\{x_m\}$ is also a subset of the box $[-M, M]^n$, a compact set, there exists x_{m_t} such that it converges to $x \in [-M, M]^n$. But K is closed forcing $x \in K$. Thus, K is compact. ■

Theorem 1.82. *The continuous image of a compact set is compact.*

Proof. Let $f : O_n \to \mathbb{R}$ be continuous with $K \subseteq O_n$, a compact set. To show that $f(K)$ is compact, pick $\{y_m\} \subseteq f(K)$ which results in $y_m = f(x_m)$ for some $x_m \in K$. K is compact, so there exists a convergent subsequence $x_{m_t} \to x \in K$. By continuity of f, we have $f(x_{m_t}) \to f(x)$. Thus, $f(K)$ is compact. ■

Theorem 1.83 (The *extreme value theorem*)**.** *Let $f : K \to \mathbb{R}$, where K is a compact set. Then, there exist $a, b \in K$ such that $f(a) \leq f(x) \leq f(b)$ for all $x \in K$.*

Proof. Note that $f(K) \subseteq \mathbb{R}$ is compact. So pick $m := \inf_{x \in K} f(x)$ and $M := \sup_{x \in K} f(x)$. Since $f(K)$ is compact, there exist $a, b \in K$ such that $f(a) = m$ and $f(b) = M$. Thus, for all $x \in K$, we have $m \leq f(x) \leq M$. ■

Lemma 1.84. *Let $f : O_n \to \mathbb{R}^n$ be a C^1-function, with O_n convex. Suppose $\sup_{x \in O_n} \left| \frac{\partial f_i}{\partial x_j}(a) \right| \leq M$ for all $i, j = 1, 2, \ldots, n$. Then*

$$\|f(a) - f(y)\| \leq n^2 M \|x - y\| \text{ for all } x, y \in O_n. \tag{1.105}$$

Proof. Here, using the mean value theorem, we have

$$\|f(x) - f(y)\| \leq \sum_{i=1}^n |f_i(x) - f_i(y)| = \sum_{i=1}^n \left| \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(c_i) \cdot (x_j - y_j) \right| \tag{1.106}$$

for $c_i \in L_{x,y}$. One then bounds as

$$\|f(x) - f(y)\| \leq M \cdot \sum_{i=1}^{n} \left| \sum_{j=1}^{n} (x_j - y_j) \right| \leq M \times \sum_{i=1}^{n} \sum_{j=1}^{n} |x_j - y_j| = n^2 M \|x - y\|. \tag{1.107}$$

■

August 25th.

Theorem 1.85 (The *inverse function theorem*)**.** *Let $f : O_n \to \mathbb{R}^n$ be a C^1-function and $a \in O_n$. Suppose $f'(a)$ is invertible. Then there exist open sets V and W, with $a \in V$ and $f(a) \in W$, such that $f|_V : V \to W$ is a bijection and $(f|_V)^{-1}$ is differentiable on W. Also, $(f^{-1}(y))' = [f'(f^{-1}(y))]^{-1}$ for all $y \in W$.*

Proof. Without the loss of generality, assume $Df(a) = I$ (This can be done easily by setting $L := (Df(a))^{-1}$ implying $L \circ f$ is differentiable and $(L \circ f)'(a) = L'(f(a))f'(a) = LL^{-1} = I$). Note that there exists a closed box U with $a \in U$ such that $f(x) \neq f(a)$ for all $x \in U \setminus \{a\}$. As $x \mapsto \det J_f(x)$ is continuous, and $\det J_f(a) = 1 \neq 0$, we conclude $\det J_f(x) \neq 0$ for all $x \in U$. By continuity of $\frac{\partial f_i}{\partial x_j}$ at a,

$$\left| \frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(a) \right| < \frac{1}{2n^2} \text{ for all } x \in U,\ i, j = 1, 2, \ldots, n. \tag{1.108}$$

We make a bold claim:

$$\|f(x) - f(y)\| \geq \frac{1}{2} \|x - y\| \text{ for all } x, y \in U. \tag{1.109}$$

To show this claim, set $g(x) = f(x) - x$ for all $x \in U$. g is then a C^1-function and $g'(x) = f'(x) - I_n$ shows $\frac{\partial g_i}{\partial x_j}(a) = \frac{\partial f_i}{\partial x_j}(a) - \frac{\partial f_i}{\partial x_j}(a)$, or

$$\left| \frac{\partial g_i}{\partial x_j}(x) \right| < \frac{1}{2n^2} \text{ for all } x \in U,\ i, j = 1, 2, \ldots, n. \tag{1.110}$$

By the previous lemma, $\|g(x) - g(y)\| \leq \frac{1}{2} \|x - y\|$. Therefore,

$$\|x - y\| - \|f(x) - f(y)\| \leq \|(x - y) - (f(x) - f(y))\| \leq \frac{1}{2} \|x - y\| \implies \|f(x) - f(y)\| \geq \frac{1}{2} \|x - y\|. \tag{1.111}$$

Now, ∂U is a compact set and $f(x) - f(a) \neq 0$ for all $x \in \partial U$, so there exists a $d > 0$ such that $\|f(x) - f(a)\| \geq d$ for all $x \in \partial U$. Set $W = B_{d/2}(f(a))$. We make another claim, that $\|y - f(a)\| \leq \|y - f(x)\|$ for all $y \in W$ and $x \in \partial U$. To see this, let

$$\|y - f(a)\| > \|y - f(x)\| \text{ for some } y \in W,\ x \in \partial U. \tag{1.112}$$

Then, by the triangle inequality,

$$d \leq \|f(x) - f(a)\| \leq \|f(x) - y\| + \|y - f(a)\| < 2 \|y - f(a)\| < d, \tag{1.113}$$

which is a contradiction. We show another claim that for each $y \in W$, there exists a unique x_0 in the interior of U such that $f(x_0) = y$. Set $g(x) = \|y - f(x)\|^2$ for $x \in U$. Observe that U is compact and g is continuous; from the previous claim, $g(a) \leq g(x)$ for all $x \in \partial U$ so the minimum of g exists in the interior of U, say, at x_0. Now $g(x) = \sum_{i=1}^n \sum_{i=1}^n (y_i - f_i(x))^2$, and is a C^1-function and $\nabla g(x_0) = \langle \frac{\partial g}{\partial x_1}(x_0), \ldots, \frac{\partial g}{\partial x_n}(x_0) \rangle = 0$. But, at x_0,

$$\frac{\partial g}{\partial x_j}(x_0) = -2 \sum_{i=1}^{n} (y_i - f_i(x_0)) \frac{\partial f_i}{\partial x_j}(x_0) = 0 \text{ for all } j = 1, 2, \ldots, n \tag{1.114}$$

$$\implies \left(\frac{\partial f_i}{\partial x_j}(x_0) \right)^t (y - f(x_0)) = 0. \tag{1.115}$$

Since the matrix is invertible ($x_0 \in U$ and $\det J_f(x) \neq 0$ for all $x \in U$), we conclude that $y - f(x_0) = 0$, or $f(x_0) = y$. Set $V = U \cap f^{-1}(W)$, which is an open set containing a. Therefore, $f|_V : V \to W$ is invertible.

We now show that f^{-1} is differentiable (and continuous) on W. Recall that $\|f(x) - f(y)\| \geq \frac{1}{2}\|x - y\|$ for all $x, y \in U$. Thus, for all $y_1, y_2 \in W$, we simply have $\left\|f^{-1}(y_1) - f^{-1}(y_2)\right\| \leq 2\|y_1 - y_2\|$, showing that f^{-1} is continuous on W. Fix $y_0 = f(x_0) \in W$ for some $x_0 \in V$, $y_0 \in W$. We show differentiability of f^{-1} at y_0. Set $A = f'(x_0)$; naturally, one would expect the derivative to be A^{-1}; thus, we are to show

$$\lim_{y \to y_0} \frac{f^{-1}(y) - f^{-1}(y_0) - A^{-1}(y - y_0)}{\|y - y_0\|} = 0. \tag{1.116}$$

We know that $\lim_{h \to 0} \frac{\varphi(h)}{\|h\|} = 0$ where $\varphi(h) = f(x_0 + h) - f(x_0) - Ah$. Set $y = f(x_0 + h)$. Thus $A^{-1}(y - y_0) = A^{-1}(f(x_0 + h) - f(x_0)) = A^{-1}(Ah + \varphi(h)) = h + A^{-1}\varphi(h)$. Thus, the numerator of the above limit can be rewritten as $A^{-1}\varphi(f^{-1}(y) - f^{-1}(y_0))$, so we are left with proving

$$\lim_{y \to y_0} \frac{A^{-1}\varphi(f^{-1}(y) - f^{-1}(y_0))}{\|y - y_0\|} = 0 \Leftrightarrow \lim_{y \to y_0} \frac{\varphi(f^{-1}(y) - f^{-1}(y_0))}{\|y - y_0\|} = 0. \tag{1.117}$$

This limit is simple to see since

$$\lim_{y \to y_0} \frac{\left\|\varphi(f^{-1}(y) - f^{-1}(y_0))\right\|}{\|y - y_0\|} = \lim_{y \to y_0} \frac{\left\|\varphi(f^{-1}(y) - f^{-1}(y_0))\right\|}{\|f^{-1}(y) - f^{-1}(y_0)\|} \cdot \frac{\left\|f^{-1}(y) - f^{-1}(y_0)\right\|}{\|y - y_0\|} = 0 \tag{1.118}$$

since $\left\|f^{-1}(y_1) - f^{-1}(y_2)\right\| \leq 2\|y_1 - y_2\|$ for all $y_1, y_2 \in W$, and f^{-1} being continuous tells us $y \to y_0$ implies $f^{-1}(y) \to f^{-1}(y_0)$. ■

Remark 1.86. As f is a C^1-function, the partials $\frac{\partial f_i}{\partial x_j}$ are also continuous function. Now $Df^{-1} = (Df)^{-1}$ shows that f^{-1} is also a C^1-function. Moreover, if f is C^k, so is f^{-1}.

1.7.1 Implicit Function Theorem

We now talk about the derivative of a function $y = f(x)$ when we are given an 'implicit' equation $F(x, y) = 0$. Let us look at cases when one can extract $y = f(x)$ given $F(x, y) = 0$.

Example 1.87.

1. Let $F(x, y) = ax + by + c$ and suppose $F(x, y) = 0$. Then, if b is non-zero, one can simply write $y = -\frac{a}{b}x - \frac{c}{b}$. Note that here $b = \frac{\partial F}{\partial y} \neq 0$. This is somewhat of a global rewritting.
2. Suppose $F(x, y) = x^2 + y^2 - 1$ with $F(x, y) = 0$. This would then imply that $y = \pm\sqrt{1 - x^2}$. Thus, such an equation cannot be written. If, however, we are given that y is positive, one could take $y = f(x) = \sqrt{1 - x^2}$; similarly, y being negative gives $y = f(x) = -\sqrt{1 - x^2}$. We get different solutions based on different localities. Note that, again in this case, $\frac{\partial F}{\partial y} = 2y \neq 0$.
3. Let $F(x, y) = 0$ on O_n, a neighbourhood of (a, b). Let $y = f(x)$ be a solution of $F = 0$. Therefore $F(x, f(x)) = 0$ is a neighbourhood of $(a, f(a))$. This then implies that

$$\frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0 \implies \frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} \tag{1.119}$$

provided $\frac{\partial F}{\partial y}(a, b) \neq 0$.

September 1st.

Here's the setting; for $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, let $(x, y) \in \mathbb{R}^{n+m}$. Also suppose that $(a, b) \in O \subseteq \mathbb{R}^{n+m}$ where O is an open set, and suppose $F : O \to \mathbb{R}^m$ is a function with $F = (f_1, \ldots, f_m)$.

Theorem 1.88 (The *implicit function theorem*)**.** *Let $F \in C^1(O)$, with $F(a, b) = 0$. If*

$\det\left(\frac{\partial f_i}{\partial y_j}(a,b)\right)_{m\times m} \neq 0$, *then there exists an open set* U, *with* $a \in U \subseteq \mathbb{R}^n$, *and a function* $f : U \to \mathbb{R}^m$ *such that* $F(x, f(x)) = 0$ *for all* $x \in U$.

Proof. Define $\tilde{F} : O \to \mathbb{R}^{n+m}$ by $\tilde{F}(x,y) = (x, F(x,y))$. This is a C^1-function, and $J_{\tilde{F}}(a,b) = \begin{bmatrix} I_n & 0 \\ (\cdots) & J_f(a,b) \end{bmatrix}$. Now $\det J_{\tilde{F}}(a,b) = \det J_f(a,b) \neq 0$. Apply the inverse function theorem to get open boxes $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ such that $(a,b) \in A \times B$ and $\tilde{F} : A \times B \to \tilde{U} \subseteq \mathbb{R}^{n+m}$ is an invertible function with a C^1-inverse. Also, $\tilde{F}(a,b) = (a,0) \in \tilde{U}$. Call $U = A \ni a$. Note that $\tilde{F}^{-1}(x,y) = (x, g(x,y))$ for all $(x,y) \in \tilde{U}$ where $g : \tilde{U} \to \mathbb{R}^m$ is a C^1-function.

If Π_1 denote the projection of the first n variables, and Π_2 denote the projection of the remaining m variables, then $\Pi_2\tilde{F} = F$, and $\Pi_1\tilde{F}(A \times B) = A$. Therefore,

$$F(x, g(x,y)) = F \circ \tilde{F}^{-1}(x,y) = \Pi_2(x,y) = y \text{ for all } (x,y) \in \tilde{U}. \tag{1.120}$$

Pick $y = 0$. One then has $F(x, g(x,0)) = 0$ for all $x \in U$. Define $f : U \to \mathbb{R}^m$ by $f(x) = g(x,0)$ for all $x \in U$. Then $F(x, f(x)) = 0$ for all $x \in U$. ■

Remark 1.89. In particular for $n = m = 1$, let $f : \mathbb{R}^2 \to \mathbb{R}$ be a C^1-function. Suppose $f(a,b) = 0$ and $\frac{\partial f}{\partial y}(a,b) \neq 0$. Then by the implicit function theorem, there exists an open box $I \times J$ with $(a,b) \in I \times J$ such that for all $x \in I$, there exists a unique $y \in J$ such that $f(x,y) = 0$. We write $y = \varphi(x)$, that is, $f(x, \varphi(x)) = 0$ for all $x \in I$. Moreover, φ is a C^1-function and $\varphi'(x) = -\dfrac{\frac{\partial f}{\partial x}(x,\varphi(x))}{\frac{\partial f}{\partial y}(x,\varphi(x))}$ for all $x \in I$.

Example 1.90. Suppose $x^2 + 2y^2 + z^2 + w = 6$ and $2x^3 + 4y^2 + z + w^2 = 9$. We try to write z and w as functions of x and y. Define $F : \mathbb{R}^4 \to \mathbb{R}^2$ by $F(x,y,z,w) = (f_1, f_2)$ where

$$f_1 = x^2 + 2y^2 + z^2 + w - 6, \quad f_2 = 2x^3 + 4y^2 + z + w^2 - 9. \tag{1.121}$$

F is then a C^1-function. Letting $\alpha = (1,-1,-1,2)$ gives $F(\alpha) = 0$. We claim that (z,w) is a C^1-function of (x,y). Here,

$$\det\begin{bmatrix} \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial w} \\ \frac{\partial f_2}{\partial z} & \frac{\partial f_2}{\partial w} \end{bmatrix} = \det\begin{bmatrix} 2z & 1 \\ 1 & 2w \end{bmatrix} = 4zw - 1 \implies \det\begin{bmatrix} \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial w} \\ \frac{\partial f_2}{\partial z} & \frac{\partial f_2}{\partial w} \end{bmatrix}(\alpha) = -9 \neq 0. \tag{1.122}$$

So there exists a C^1-function $\varphi : U \to \mathbb{R}^2$ with $(1,-1) \to \mathbb{R}^2$ such that $F(x,y,\varphi(x,y)) = 0$ for all $(x,y) \in U$. Thus, (z,w) can be expressed as a function of (x,y).

Chapter 2

INTEGRATION: RIEMANN AND DARBOUX

2.1 Intuition

September 15th.

For the one-dimensional integration, we studied how to compute the length of curves, the area under a curve, and even antiderivatives. Our aim, now, is to extend these concepts to higher dimensions; given a function $f : O_n \to \mathbb{R}$, we want to give meaning to $\int_{O_n} f d*$, and fill the $*$ too.

Definition 2.1. We term $B^n = \prod_{i=1}^n [a_i, b_i]$ a *closed box* in $\mathbb{R}^n$. One can show that the closure of an open box is the corresponding closed box. The *volume of a box* is defined as

$$v(B^n) = v\left(\prod_{i=1}^n [a_i, b_i]\right) = v\left(\prod_{i=1}^n (a_i, b_i)\right) = \prod_{i=1}^n (b_i - a_i). \tag{2.1}$$

We extend the concept of partitions to boxes in the natural way.

Definition 2.2. Let $B^n = \prod_{i=1}^n [a_i, b_i]$ be a closed box in $\mathbb{R}^n$. For each $i = 1, 2, \ldots, n$, consider a partition of $[a_i, b_i]$ give by $P_i : a_i = x_{i,0} < x_{i,1} < \cdots < x_{i,n_i} = n_i$. Set $P = P_1 \times P_2 \times \cdots \times P_n$, a *partition of the box*, and set

$$\Lambda(P) = \{(t_1, t_2, \ldots, t_n) : 0 \leq t_i \leq n_i\}, \tag{2.2}$$

the set of multi-indices. For $\alpha \in \Lambda(P)$, define the α^{th} *closed sub-box* as $B^n_\alpha = I_{1,t_1} \times I_{2,t_2} \times \cdots \times I_{n,t_n}$, where $I_{i,t_i} = [x_{i,t_i-1}, x_{i,t_i}]$. The set of all closed sub-boxes is called a *partition of the box* B^n. We get

$$B^n = \bigcup_{\alpha \in \Lambda(P)} B^n_\alpha, \quad v(B^n) = \sum_{\alpha \in \Lambda(P)} v(B^n_\alpha). \tag{2.3}$$

Another unavoidable definition is that of refinement of partitions.

Definition 2.3. For a closed box B^n, a partition $\tilde{P}$ is termed a *refinement* of a partition P if $\tilde{P}_i$ is a refinement of P_i for each $i = 1, 2, \ldots, n$.

Given $f \in \mathcal{B}(B^n)$, a bounded function over B^n, and a partition P of B^n, we define

$$m_\alpha = \inf_{x \in B^n_\alpha} f(x), \quad M_\alpha = \sup_{x \in B^n_\alpha} f(x) \tag{2.4}$$

for all $\alpha \in \Lambda(P)$. We then define the *lower Darboux sum* and the *upper Darboux sum* of f with respect to P as

$$L(f,P) = \sum_{\alpha\in\Lambda(P)} m_\alpha v(B^n_\alpha), \quad U(f,P) = \sum_{\alpha\in\Lambda(P)} M_\alpha v(B^n_\alpha) \tag{2.5}$$

respectively. If $m = \inf_{x\in B^n} f(x)$ and $M = \sup_{x\in B^n} f(x)$, then it is easy to see that

$$mv(B^n) \leq L(f,P) \leq U(f,P) \leq Mv(B^n). \tag{2.6}$$

.

Theorem 2.4. *For $\tilde{P} \supset P$ partitions of B^n, and $f \in \mathcal{B}(B^n)$, we have*

$$L(f,P) \leq L(f,\tilde{P}) \leq U(f,\tilde{P}) \leq U(f,P). \tag{2.7}$$

Proof. The proof of this is left as an exercise to the reader; it is a straightforward generalization of the one-dimensional case. ■

We also denote $\mathcal{P}(B^n)$ as the set of all partitions of B^n.

Corollary 2.5. *For all $P,Q \in \mathcal{P}(B^n)$, $L(f,P) \leq U(f,Q)$.*

Proof. Easy to see since $P,Q \subset P \cup Q$. ■

Since refinement leads to an increase in the lower Darboux sum and a decrease in the upper Darboux sum, we can define as follows.

Definition 2.6. Given $f \in cB(B^n)$, we define the *lower Darboux integral* and the *upper Darboux integral* of f over B^n as

$$\underline{\int_{B^n}} f = \sup_{P\in\mathcal{P}(B^n)} L(f,P), \quad \overline{\int_{B^n}} f = \inf_{P\in\mathcal{P}(B^n)} U(f,P) \tag{2.8}$$

respectively. If $\underline{\int_{B^n}} f = \overline{\int_{B^n}} f$, we say that f is *Riemann-Darboux integrable* over B^n, and denote the common value as the *Darboux integral* of f over B^n, denoted as

$$\int_{B^n} f = \underline{\int_{B^n}} f = \overline{\int_{B^n}} f. \tag{2.9}$$

We may write

$$\int_{B^n} f = \int_{B^n} f dv = \int_{B^n} f(x_1,\ldots,x_n)dx_1dx_2\cdots dx_n \tag{2.10}$$

where dv denotes the infinitesimal volume element in n-dimensions, and with no intention of commuting the dx_i's; it is simply notation.

Also, we denote $\mathcal{R}(B^n) := \{f \in \mathcal{B}(B^n) : f \text{ is Riemann-Darboux integrable over } B^n\}$.

Theorem 2.7. *Let $f \in \mathcal{B}(B^n)$. Then $f \in \mathcal{R}(B^n)$ if and only if for every $\varepsilon > 0$, there exists a partition $P \in \mathcal{P}(B^n)$ such that $U(f,P) - L(f,P) < \varepsilon$.*

Proof. Let us assume f is Riemann-Darboux integrable. Then

$$\inf_P U(f,P) - \sup_P L(f,P) = \inf_P (U(f,P) - L(f,P)) = 0 \tag{2.11}$$

showing the necessary result. For the converse, let $\varepsilon > 0$ be given. By hypothesis, there exists $P \in \mathcal{P}(B^n)$ such that $U(f,P) - L(f,P) < \varepsilon$. Simply by the definition of infimum and supremum, we have

$$0 \leq \overline{\int_{B^n}} f - \underline{\int_{B^n}} f \leq U(f,P) - L(f,P) < \varepsilon \implies f \in \mathcal{R}(B^n). \tag{2.12}$$

■

Note that $\mathcal{R}(B^n)$ is a vector space over $\mathbb{R}$, satisfying $\int_{B_n}(f+\alpha g) = \int_{B_n} f + \alpha \int_{B_n} g$ for all $f, g \in \mathcal{R}(B^n)$ and $\alpha \in \mathbb{R}$. Also, if $f, g \in \mathcal{R}(B^n)$ and $f(x) \leq g(x)$ for all $x \in B^n$, then $\int_{B^n} f \leq \int_{B^n} g$.

If $f \in \mathcal{R}(B^n)$, then $|f| \in \mathcal{R}(B^n)$ and $\left|\int_{B^n} f\right| \leq \int_{B^n} |f|$. Also, if $f, g \in \mathcal{R}(B^n)$, then $fg \in \mathcal{R}(B^n)$, making $\mathcal{R}(B^n)$ an algebra over $\mathbb{R}$.

Definition 2.8. The *diameter of a box* B^n is simply defined as the largest diagonal of the box, $\operatorname{diam}(B^n)$.

Using the above, the concept of mesh can be extended.

Definition 2.9. The *norm of a partition* or *mesh of a partition* P of a box B^n is defined as

$$\|P\| = \max_{\alpha \in \Lambda(P)} \operatorname{diam}(B^n_\alpha). \tag{2.13}$$

Theorem 2.10. *$\mathcal{C}(B^n) \subsetneq \mathcal{R}(B^n)$.*

Proof. Pick $f : B^n \to \mathbb{R}$ continuous. Given $\varepsilon > 0$, by uniform continuity, there exists $\delta > 0$ such that $|f(x) - f(y)| < \frac{\varepsilon}{2v(B^n)}$ whenever $\|x - y\| < \delta$. Now, choose a partition P of B^n such that $\|P\| < \delta$. Then, for each $\alpha \in \Lambda(P)$, $a_\alpha \in B^n_\alpha$. Thus for all $x \in B^n_\alpha$, we have

$$|f(x) - f(a_\alpha)| < \frac{\varepsilon}{2v(B^n)} = \tilde{\varepsilon} \implies f(a_\alpha) - \tilde{\varepsilon} < f(x) < f(a_\alpha) + \tilde{\varepsilon} \text{ for all } x \in B^n_\alpha. \tag{2.14}$$

Inserting the infimum and supremum, we get

$$f(a_\alpha) - \tilde{\varepsilon} < m_\alpha \leq M_\alpha < f(a_\alpha) + \tilde{\varepsilon} \implies c(P) - \frac{\varepsilon}{2} \leq L(f,P) \leq U(f,P) \leq c(P) + \frac{\varepsilon}{2} \tag{2.15}$$

where $c(P) = \sum_{\alpha \in \Lambda(P)} f(a_\alpha) v(B^n_\alpha)$. Thus $U(f,P) - L(f,P) < \varepsilon$, showing that $f \in \mathcal{R}(B^n)$. ■

September 19th.

We question whether the use of one-variable integration can help us in evaluating $\int_{B^n} f dv$ for $f \in \mathcal{R}(B^n)$. The answer is yes, and we will see how. Consider $B^2 = [a_1, b_1] \times [a_2, b_2] \subseteq \mathbb{R}^2$.Given $f : B^2 \to \mathbb{R}$, we consider slice functions $f_x : [a_2, b_2] \to \mathbb{R}$ defined as $f_x(y) = f(x,y)$ for each fixed $x \in [a_1, b_1]$. Similarly, one defines $f_y(x) = f(x,y)$ over $[a_2, b_2]$. Let P be a partition of $B^2 = B^1_1 \times B^1_2$, where $P = P_1 \times P_2$, and P_i is a partition of B^1_i for $i = 1, 2$. Moreover, $\Lambda(P) = \Lambda(P_1) \times \Lambda(P_2)$.

Example 2.11. Look at $B^2 = [0,1] \times [0,1]$ and $f(x,y) = \mathbf{1}_{1/2}(x) \cdot \mathbf{1}_{\mathbb{Q}}(y)$. If we consider the slice function $f_x : [0,1] \to \mathbb{R}$, then for $x = \frac{1}{2}$, $f_{\frac{1}{2}}$ is the Dirichlet function, which is not Riemann integrable. Thus, $\int_0^1 \int_0^1 f(x,y) dy dx$ is not defined. However, if we consider $f_y : [0,1] \to \mathbb{R}$, then for each fixed $y \in [0,1]$, $f_y(x)$ is 0 except a single point, where it is 1. Thus, f_y is Riemann integrable for each $y \in [0,1]$, and $\int_0^1 f_y(x) dx = 0$ along with $\int_0^1 \int_0^1 f(x,y) dx dy = 0$. Thus, the order of integration does not commute.
We, however, will show that $f \in \mathcal{R}(B^2)$ and $\int_{B^2} f dv = 0$. Pick $\varepsilon > 0$. Pick partition $\Lambda(P_1) = \{0, \frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon, 1\}$ and $\Lambda(P_2) = \{0, 1\}$. Then, $P = P_1 \times P_2$ is a partition of B^2. The sub-boxes are $B_1 = [0, \frac{1}{2} - \varepsilon] \times [0,1]$, $B_2 = [\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon] \times [0,1]$, and $B_3 = [\frac{1}{2} + \varepsilon, 1] \times [0,1]$. The supremum over these boxes are $M_1 = 0$, $M_2 = 1$, and $M_3 = 0$. The infimum over these boxes are $m_1 = 0$, $m_2 = 0$,

and $m_3 = 0$. Thus, we have

$$U(f,P) - L(f,P) = \sum_{j=1}^{3}(M_j - m_j)v(B_j) = 2\varepsilon \to 0. \tag{2.16}$$

This tells us that $f \in \mathcal{R}(B^2)$and $\int_{B^2} f dv = 0$.

Now suppose $B^m \times B^n = B^{m+n} \subseteq \mathbb{R}^{m+n}$, and let $x \in B^m$ and $y \in B^n$. Fix $P \in \mathcal{P}(B^{m+n})$ and let $P = P^m \times P^n$, where $P^m \in \mathcal{P}(B^m)$ and $P^n \in \mathcal{P}(B^n)$. Subsequently, $\Lambda(P) = \Lambda(P^m) \times \Lambda(P^n)$ and for all $\alpha(P) \in \Lambda(P)$, $\alpha = (\alpha(P^m), \alpha(P^n))$. The sub-boxes are $B^{m+n}_{\alpha(P)} = B^m_{\alpha(P^m)} \times B^n_{\alpha(P^n)}$. Let $f : B^{m+n} \to \mathbb{R}$ be a bounded function, and let $f_x : B^n \to \mathbb{R}$ be the slice function in x. We then define

$$\underline{f}(x) := \underline{\int_{B^n}} f_x(y)dv(y) \ \text{ and } \ \overline{f}(x) := \overline{\int_{B^n}} f_x(y)dv(y) \ \text{ for all } x \in B^m. \tag{2.17}$$

which are both functions as $\underline{f}, \overline{f} : B^m \to \mathbb{R}$.

Theorem 2.12 (*Fubini's theorem*)**.** *Let $f \in \mathcal{R}(B^{m+n})$. Then $\underline{f}, \overline{f} \in \mathcal{R}(B^m)$. Moreover,*

$$\int_{B^{m+n}} f dv = \int_{B^m} \underline{f}(x)dv(x) = \int_{B^m} \overline{f}(x)dv(x). \tag{2.18}$$

Proof. Fix $P \in \mathcal{P}(B^m \times B^n)$ and let $P = P^m \times P^n$, where $P^m \in \mathcal{P}(B^m)$ and $P^n \in \mathcal{P}(B^n)$. Now

$$L(f,P) = \sum_{\alpha \in \Lambda(P)} m_\alpha v(B^{m+n}_\alpha) = \sum_{\alpha(P^m) \in \Lambda(P^m)} \left(\sum_{\alpha(P^n) \in \Lambda(P^n)} m_{(\alpha(P^m), \alpha(P^n))} v(B^n_{\alpha(P^n)}) \right) v(B^m_{\alpha(P^m)}) \tag{2.19}$$

where we broke down $v(B^{m+n}_\alpha)$ as $v(B^m_{\alpha(P^m)})v(B^n_{\alpha(P^n)})$. Let us call the summation inside the parenthesis as $(*)$. For each $x \in B^m$, $\alpha(P^n) \in \Lambda(P^n)$, set $m_{\alpha(P^n)}(x) = \inf_{y \in B^n_{\alpha(P^n)}} f_x(y)$ Thus,

$$m_{\alpha(P^n)}(x) \geq m_{(\alpha(P^m), \alpha(P^n))} \text{ for all } x \in B^m_{\alpha(P^m)} \tag{2.20}$$

$$\implies (*) \leq \sum_{\alpha(P^n) \in \Lambda(P^n)} m_{\alpha(P^n)}(x) v(B^n_{\alpha(P^n)}) = L(f_x, P^n) \leq \underline{\int_{B^n}} f_x dv(x) = \underline{f}(x) \text{ for all } x \in B^m_{\alpha(P^m)}. \tag{2.21}$$

Note, that $\underline{m}_{B^m_{\alpha(P^m)}} = \inf_{x \in B^m_{\alpha(P^m)}} \underline{f}(x)$ is also an upper bound for the above. Therefore, we obtain

$$L(f,P) \leq \sum_{\alpha(P^m) \in \Lambda(P^m)} \underline{m}_{B^m_{\alpha(P^m)}} v(B^m_{\alpha(P^m)}) = L(\underline{f}, P^m). \tag{2.22}$$

Similarly, one obtains

$$U(f,P) \geq U(\underline{f}, P^m). \tag{2.23}$$

Finally, we obtain $\underline{f} \in \mathcal{R}(B^m)$. One may do the same work to obtain $\overline{f} \in \mathcal{R}(B^m)$. The equality in the theorem also follows. ∎

Corollary 2.13. *Let $f \in \mathcal{R}(B^{m+n})$. If $f_x \in \mathcal{R}(B^n)$ for all $x \in B^m$ then*

$$\int_{B^{m+n}} f dv = \int \int f(x,y)dv(y)dv(x). \tag{2.24}$$

Corollary 2.14. *Let $f \in \mathcal{C}(B^n)$. Then*

$$\int_{B^n} f dv = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n) dx_n \cdots dx_1 \tag{2.25}$$

where $B^n = \prod_{i=1}^n [a_i, b_i]$, and the order of integration can be interchanged.

September 22nd.

Example 2.15. Let us integrate $f(x,y) = xy$ over $B^2 = [0,1] \times [0,1]$. Since f is continuous, we have

$$\int_{[0,1]^2} f(x,y) dv = \int_0^1 x \left(\int_0^1 y dy \right) dx = \int_0^1 x \cdot \frac{1}{2} dx = \frac{1}{4}. \tag{2.26}$$

2.1.1 Generalization

So far, we have only discussed integration over boxes. We now extend this to more general sets; if $\Omega \subseteq \mathbb{R}^n$ is bounded, we wish to define $\int_\Omega f dv$ for $f \in \mathcal{B}(\Omega)$. Here onwards, most proofs will be omitted since they involve measure theoretic arguments beyond the scope of this course. Where an idea of a proof is needed, we will provide it.

Definition 2.16. Given $f \in \mathcal{B}(\Omega)$, for some bounded $\Omega \subseteq \mathbb{R}^n$, take a box $B^n \supseteq \Omega$ and define $\tilde{f} : B^n \to \mathbb{R}$ as

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \Omega \\ 0 & \text{if } x \in B^n \setminus \Omega. \end{cases} \tag{2.27}$$

We say $f \in \mathcal{R}(\Omega)$ if $\tilde{f} \in \mathcal{R}(B^n)$, and define the integral as $\int_\Omega f dv := \int_{B^n} \tilde{f} dv$.

Theorem 2.17. *Let $B_1^n, B_2^n \supseteq \Omega \subseteq \mathbb{R}^n$ be two boxes, with Ω bounded. Let $f \in \mathcal{B}(\Omega)$ and define $\tilde{f}_i : B_i^n \to \mathbb{R}$ as above for each i. Then $\tilde{f}_1 \in \mathcal{R}(B_1^n)$ if and only if $\tilde{f}_2 \in \mathcal{R}(B_2^n)$, and in that case, $\int_{B_1^n} \tilde{f}_1 dv = \int_{B_2^n} \tilde{f}_2 dv$.*

Definition 2.18. A set $S \subseteq \mathbb{R}^n$ is said to be of *content zero* if for every $\varepsilon > 0$ there exists a finite collection of boxes $\{B_j^n\}_{j=1}^m$ such that $S \subseteq \bigcup_{j=1}^m B_j^n$ and $\sum_{j=1}^m v(B_j^n) < \varepsilon$.

Note that content zero sets are *not* the same as measure zero sets. For example, $\mathbb{Q} \cap [0,1]$ is a measure zero set but not a content zero set in $\mathbb{R}^1$. Before we discuss the consequences, let us look at two special domains.

1. A set $\Omega \subseteq \mathbb{R}^2$ is y-simple (type-I) if there exist $\varphi_1, \varphi_2 \in \mathcal{B}[a,b]$ such that $\Omega = \{(x,y) \mid \varphi_1(x) \leq y \leq \varphi_2(x),\ a \leq x \leq b\}$.
2. Similarly, Ω is x-simple (type-II) if there exist $\psi_1, \psi_2 \in \mathcal{B}[c,d]$ such that $\Omega = \{(x,y) \mid \psi_1(y) \leq x \leq \psi_2(y),\ c \leq y \leq d\}$.

Theorem 2.19. *Let $f \in \mathcal{R}(\Omega)$, where $\Omega = \{(x,y) \mid \varphi_1(x) \leq y \leq \varphi_2(x),\ a \leq x \leq b\}$ for some $\varphi_1, \varphi_2 \in \mathcal{B}[a,b]$. Suppose $\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y) dy$ exists for all $x \in [a,b]$. Then*

$$\int_\Omega f(x,y) dv = \int_a^b \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y) dy dx. \tag{2.28}$$

September 29th.

Example 2.20. Let Ω be the region bounded by $y = 1$ and $y = x^2$. We ask to compute the integral $\int_\Omega x^2 y dA$. Here, we have

$$\int_\Omega x^2 y dA = \int_{-1}^{1} \int_{y=x^2}^{1} x^2 y dy dx = \int_{-1}^{1} \frac{1}{2}(x^2 - x^6) dx = \frac{4}{21}. \tag{2.29}$$

Example 2.21. Look at $\Omega = [0,1]^2$ and $f(x,y) = x$ if $y \leq x^2$ and y if $y > x^2$. We wish to compute $\int_\Omega f(x,y) dA$. We split apart region Ω into two parts, $\Omega_1 = \{(x,y) \mid 0 \leq x \leq 1, 0 \leq y \leq x^2\}$ and $\Omega_2 = \{(x,y) \mid 0 \leq x \leq 1, x^2 < y \leq 1\}$. Then, $f|_{\Omega_1}(x,y) = x$ and $f|_{\Omega_2}(x,y) = y$. Thus,

$$\int_\Omega f dA = \int_{\Omega_1} f + \int_{\Omega_2} f = \int_0^1 \int_{y=0}^{x^2} x dy dx + \int_0^1 \int_{y=x^2}^{1} y dy dx = \frac{13}{20}. \tag{2.30}$$

2.1.2 Change of Variables

Recall the $n = 1$ case. Let $\varphi : O \to \mathbb{R}$ be a C^1-function with $O \subseteq \mathbb{R}$ open. Also suppose $\varphi'(x) \neq 0$ for all $x \in O$. If we let $f \in C(\varphi[a,b])$, where $[a,b] \subseteq O$, then

$$\int_{\varphi[a,b]} f = \int_a^b f \circ \varphi \, |\varphi'| \tag{2.31}$$

We generalize this to higher dimensions.

Theorem 2.22. *Let $\varphi : O_n \to \mathbb{R}^n$ be an injective C^1-function. Suppose* $\det J_\varphi(x) \neq 0$ *for all $x \in O_n$. Let $\Omega \subseteq O_n$ be a bounded set such that $\Omega \cup \partial\Omega \subseteq O_n$ and $\partial\Omega$ is content zero. If $f \in \mathcal{R}(\varphi(\Omega))$, then*

$$\int_{\varphi(\Omega)} f = \int_\Omega f \circ \varphi \, |\det J_\varphi| \,. \tag{2.32}$$

The proof of this theorem is beyond the scope of the course.

Example 2.23. Let $(r,\theta) \in \mathbb{R}^2$. Set $x = r\cos\theta$ and $y = r\sin\theta$. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as $\varphi(r,\theta) = (x(r,\theta), y(r,\theta)) = (r\cos\theta, r\sin\theta)$. Here, the Jacobian is

$$J_\varphi = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}. \tag{2.33}$$

Thus, $\det J_\varphi = r$. Set $O_2 = \{(r,\theta) \mid r > 0, 0 < \theta < 2\pi\}$. Then, $\varphi : O_2 \to \mathbb{R}^2$ with the same definition as above is injective and C^1 with $\det J_\varphi(r,\theta) \neq 0$ for all $(r,\theta) \in O_2$. We shall apply this in the next example.

Example 2.24. We wish to compute $\int_{x^2+y^2<1} e^{-(x^2+y^2)} dA$. Set $\Omega = \{(x,y) \mid x^2 + y^2 < 1\}$ and $\tilde{\Omega} = \{(r,\theta) \mid 0 \leq r < 1, 0 \leq \theta < 2\pi\}$. Then, $\varphi(\tilde{\Omega}) = \Omega$. Since $\partial\tilde{\Omega}$ is content zero, we can apply the change of variables theorem. Thus,

$$\int_{x^2+y^2<1} e^{-(x^2+y^2)} dA = \int_{\theta=0}^{2\pi} \int_{r=0}^{1} e^{-r^2} |r| \, dr d\theta = \pi(1 - e^{-1}). \tag{2.34}$$

Example 2.25. We wish to compute the area of $\Omega = \{(x,y) \mid x^{2/3} + y^{2/3} \leq 1\}$. We use the transformation $x = s\cos^3 t$ and $y = s\sin^3 t$. Define $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ by $\varphi(s,t) = (s\cos^3 t, s\sin^3 t)$. Then $\Omega = \varphi([0,1] \times [0,2\pi])$. The determinant of the Jacobian is $\det J_\varphi(s,t) = 3s\cos^2 t \sin^2 t$. Thus, the area is simply

$$\int_{\varphi([0,1]\times[0,2\pi])} 1dA = \int_0^{2\pi}\int_0^1 3s\cos^2 t\sin^2 tdsdt = \frac{3\pi}{8}. \tag{2.35}$$

Here, even though $\det J_\varphi(s,t) = 0$ when $s = 0$ or $t = k\pi/2$ for some integer k, the change of variables theorem is still applicable since these points form a content zero set.

Example 2.26. We discuss spherical coordinates. In this case, we have $(x, y, z) = (r\sin\phi\cos\theta, r\sin\phi\sin\theta, r\cos\phi)$. We consider the set $O_3 = \{(r, \theta, \phi) \mid r > 0,\ 0 < \phi < \pi,\ 0 < \theta < 2\pi\}$. Define $\varphi : O_3 \to \mathbb{R}^3$ as $\varphi(r, \theta, \phi) = (r\sin\phi\cos\theta, r\sin\phi\sin\theta, r\cos\phi)$. This φ is a one-to-one C^1-function. The determinant of the Jacobian is $\det J_\varphi = r^2\sin\phi$, which is non-zero in O_3. Therefore, for $f \in C(\varphi([r_1, r_2] \times [\phi_1, \phi_2] \times [\theta_1, \theta_2])) = C(\varphi(O))$ where $0 < r_1 < r_2$, $0\phi_1 < \phi_2 < \pi$, and $0 < \theta_1 < \theta_2 < 2\pi$. We then have

$$\int_{\varphi(O)} f = \int_O f\circ\phi\cdot r^2\sin\phi drd\phi d\theta. \tag{2.36}$$

2.2 Curves and Surfaces

October 1st.

Definition 2.27. Given an interval $I = [a, b] \subseteq \mathbb{R}$, a parametrized *curve* in $\mathbb{R}^n$ is a continuous function $\gamma : I \to \mathbb{R}^n$. We call $\{\gamma(t) \mid t \in I\}$ to be the *path* of γ.

For example, one may define the curve $\gamma : [0, 2\pi] \to \mathbb{R}^2$ as $\gamma(t) = (\cos t, \sin t)$, which is the unit circle in $\mathbb{R}^2$. If we had chosen $\tilde{\gamma} : [0, 2\pi] \to \mathbb{R}^2$ as $\tilde{\gamma}(t) = (\cos 2t, \sin 2t)$, then the path of $\tilde{\gamma}$ is still the unit circle, but it is traversed twice as fast. Thus given the path of a curve, there are infinitely many parametrizations of the same path.

Definition 2.28. Let $\gamma : I \to \mathbb{R}^n$ be a curve. γ is termed a C^1-function if each component function $\gamma_i : I \to \mathbb{R}$ is a C^1-function for all $i = 1, 2, \ldots, n$. A differentiable γ is called *smooth* if γ is C^1 and $\gamma'(t) \neq 0$ for all $t \in I$.

Definition 2.29. A parametrized curve $\gamma : I \to \mathbb{R}^n$ is said to be *piecewise smooth* if there exists a partition of $I = [a, b]$, say, $P : a = x_0 < x_1 < \cdots < x_n = b$ such that $\gamma|_{[x_{j-1},x_j]}$ is smooth for all $j = 1, 2, \ldots, n$.

Definition 2.30. Two parametrized curves $\gamma : [a, b] \to \mathbb{R}^n$ and $\tilde{\gamma} : [\tilde{a}, \tilde{b}] \to \mathbb{R}^n$ are equivalent if there exists a strictly increasing parametrized curve $\varphi : [\tilde{a}, \tilde{b}] \to [a, b]$, which is also surjective and continuously differentiable, such that $\tilde{\gamma} = \gamma \circ \varphi$.

Definition 2.31. Let $\gamma : I \to \mathbb{R}^n$ be a C^1-curve. The *speed of the curve* γ at $t \in I$ is defined as $\|\gamma'(t)\|$. The *length of the curve* γ between t_1 and t_2 is defined as

$$\int_{t_1}^{t_2} \|\gamma'(t)\|\, dt \tag{2.37}$$

for any $[t_1, t_2] \subseteq I$.

Most of these definitions seem more physical rather than mathematical. A more natural way to compute the length of a curve is as follows. Given a curve $\gamma : I \to \mathbb{R}^n$, we consider a partition

$P : a = t_0 < t_1 < t_2 < \cdots < t_n = b$ of I. Then, we define the length of γ with respect to P as

$$\ell(\gamma, P) = \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\| . \tag{2.38}$$

Definition 2.32. A curve $\gamma : I \to \mathbb{R}^n$ is said to be *rectifiable* if $\ell(\gamma) := \lim_{\|P\|\to 0} \ell(\gamma, P)$ exists.

Note that for a piecewise smooth curve $\gamma : [a, b] \to \mathbb{R}^n$, the curve is rectifiable and $\ell(\gamma) = \int_a^b \|\gamma'(t)\|\, dt$. However, a recetifiable curve need not be piecewise smooth. A popular example is the Cantor function.

October 6th.

Example 2.33. Note that the curves $\gamma(t) = (t^3, t^6)$ and $\tilde{\gamma}(t) = (t, t^2)$ for $t \in [0, 1]$ have the same trace. However, one may note that $\gamma(t)$ is not smooth since $\gamma'(0) = 0$, but $\tilde{\gamma}(t)$ is smooth. Note that we had seen $\gamma = \tilde{\gamma} \circ \varphi$ is the condition for equivalent curves. But here $\varphi(t) = t^{1/3}$ is not differentiable at $t = 0$. Thus, γ and $\tilde{\gamma}$ are not equivalent.

Example 2.34. Let $\gamma(t) = (\alpha t, \beta t - 16t^2)$ with $\alpha, \beta \in \mathbb{R} \setminus \{0\}$. In this case, $\gamma'(t) = (\alpha, \beta - 32t)$. Thus the length of γ is $\ell(\gamma) = \int \sqrt{\alpha^2 + (\beta - 32t)^2} dt$ with suitable limits.

Example 2.35. We look at the graph of a function $f \in \mathbb{C}^1[a, b]$. Define $\gamma : [a, b] \to \mathbb{R}^2$ as $\gamma(t) = (t, f(t))$. Then, $\gamma'(t) = (1, f'(t))$, and the length of the curve is $\ell(\gamma) = \int_a^b \sqrt{1 + (f'(t))^2} dt$.

2.2.1 Line Integrals

To begin, a *scalar field* is simply a function of the form $f : \mathbb{R}^n \to \mathbb{R}$ and a *vector field* is a function of hte form $f : \mathbb{R}^n \to \mathbb{R}^n$.

Let $\gamma : [a, b] \to \mathbb{R}^n$ be a smooth (or rectifiable) curve with C the trace of γ. Let $f \in \mathcal{B}(C)$ be a scalar field. Given a partition $P : a = t_0 < t_1 < \cdots < t_m = b$ of $[a, b]$, we define $s_i := \|\gamma(t_i) - \gamma(t_{i-1})\|$ for all $i = 1, 2, \ldots, m$. We also define $C_i := \gamma([t_{i-1}, t_i])$ and $M_i = \sup_{x\in C_i} f(x)$ and $m_i = \inf_{x\in C_i} f(x)$. We then define the upper and lower sums as

$$U(f, P) = \sum_{i=1}^{m} M_i s_i \quad \text{and} \quad L(f, P) = \sum_{i=1}^{m} m_i s_i. \tag{2.39}$$

Definition 2.36. We say that the above f is integrable if

$$\sup_{P\in\mathcal{P}([a,b])} L(f, P) = \inf_{P\in\mathcal{P}([a,b])} U(f, P) \quad \left(= \int_C f ds \right). \tag{2.40}$$

The last notation is termed the *line integral* of f over C with respect to arc length.

Theorem 2.37. *The following hold.*

1. *If f is a constant function, then f is integrable over C.*
2. *f is integrable over C if and only if $\lim_{\|P\|\to 0} \sum_{i\in\Lambda(P),\, \zeta_i\in C_i} f(\zeta_i) s_i$ exists. In such a case, the limit is always equal to $\int_C f ds$.*
3. *Let γ be a C^1-curve and $f \in \mathcal{R}(C)$. Then f is integrable over C and*

$$\int_C f ds = \int_a f(\gamma(t)) \|\gamma'(t)\| \, dt. \tag{2.41}$$

For two equivalent parametrized curves γ and $\tilde{\gamma} = \gamma \circ \varphi$,with $\varphi : [c, d] \to [a, b]$ differentiable, strictly increasing, and surjective, we have

$$\int_c^d f(\tilde{\gamma}(s)) \|\tilde{\gamma}'(s)\| \, ds = \int_c^d f(\gamma(\varphi(s))) \|\gamma'(\varphi(s))\| \, \varphi'(s) ds = \int_a^b f(\gamma(t)) \|\gamma'(t)\| \, dt \tag{2.42}$$

where we made the change of variable $\varphi(s) \mapsto t$. Thus, the line integral is independent of the (equivalent) parametrization of the curve. We may rewrite the line integral as $\int_\gamma f$, for any equivalent parametrization γ of C.

Let $\gamma : [a, b] \to \mathbb{R}^n$ be a piecewise smooth curve with C the trace of γ. Let f and g be continuous scalar fields over C. Then, the following hold.

1. $\int_\gamma f + \alpha g = \int_\gamma f + \alpha \int_\gamma g$ for all $\alpha \in \mathbb{R}$.
2. If $f \geq g$, then $\int_\gamma f \geq \int_\gamma g$.
3. Let $a < c < b$, and let $\gamma_1 = \gamma|_{[a,c]}$ and $\gamma_2 = \gamma|_{[c,b]}$. Then, $\int_\gamma f = \int_{\gamma_1} f + \int_{\gamma_2} f$.
4. $\left|\int_\gamma f\right| \leq \int_\gamma |f|$.
5. Given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 < t - \tilde{t} < \delta$, we have $\int_{\gamma|_{[\tilde{t},t]}} f < \varepsilon$.

October 8th.

Let $F : O_n \to \mathbb{R}^n$ be a vector field, where $O_n \subseteq \mathbb{R}^n$ is open, and let $\gamma : [a, b] \to O_n$ be a curve. Consider a partition $P : a = t_0 < t_1 < \cdots < t_m = b$ of $[a, b]$. For each i, set $\gamma_i = \gamma(t_i)$, and $\delta\gamma_i = \gamma(i) - \gamma_{i-1}$. We then define the Riemann sum as

$$R(F, P) = \sum_{i=1}^{m} F(\gamma_i) \cdots \delta\gamma_i. \tag{2.43}$$

Finally, the line integral of F over the curve C, the trace of γ, is defined as

$$\int_C F \cdot d\gamma = \lim_{\|P\| \to 0} R(F, P) \tag{2.44}$$

If γ is C^1 and $\int_C F \cdot d\gamma$ exists, then

$$\int_C F \cdot d\gamma = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt. \tag{2.45}$$

We now provide an analogue of the fundamental theorem of calculus for line integrals.

Theorem 2.38. *Let $f : O_n \to \mathbb{R}$ be a C^1-scalar field, and let γ be a piecewise smooth curve joining A and B, with C the trace of γ. Then,*

$$\int_C \nabla f \cdot d\gamma = f(B) - f(A). \tag{2.46}$$

Proof. We have

$$\int_C \nabla f \cdot d\gamma = \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t) dt = \int_a^b \frac{d}{dt}(f(\gamma(t))) dt = f(\gamma(b)) - f(\gamma(a)) = f(B) - f(A). \tag{2.47}$$

■

We move on to dealing with surfaces.

Definition 2.39. A *region* R in $\mathbb{R}^n$ is a bounded open subset such that ∂R is of content zero.

Definition 2.40. Let $R \subseteq \mathbb{R}^2$ be a region. A C^1-function $r : R \to \mathbb{R}^3$ is called a surface parametrization if

1. $\left\{\frac{\partial r}{\partial u}, \frac{\partial r}{\partial v}\right\}$ is a a bounded set on R,
2. r is one-to-one,
3. $r_u \times r_v|_{(u,v)} \neq 0$ for all $(u, v) \in R$.

The range S of r is termed a parametrized *surface* in $\mathbb{R}^3$.

A standard example is the equation of a plane in $\mathbb{R}^3$ through (x_0, y_0, z_0) with normal vector $N = (a, b, c)$. This equation is given by

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0. \tag{2.48}$$

If $c \neq 0$, then we may rewrite this as $z = z_0 - \frac{a}{c}(x - x_0) - \frac{b}{c}(y - y_0)$. Thus, we may parametrize this plane as $r(u, v) = (u, v, z_0 - \frac{a}{c}(u - x_0) - \frac{b}{c}(v - y_0))$ for $(u, v) \in \mathbb{R}^2$.

Let us talk about the third property in the definition os a surface. Fix $(u_0, v_0) \in R$. Since R is open, there exists $\varepsilon > 0$ such that

$$(u_0 - \varepsilon, u_0 + \varepsilon) \times \{v_0\},\ \{u_0\} \times (v_0 - \varepsilon, v_0 + \varepsilon) \subseteq R. \tag{2.49}$$

Thus, we can define a map over $\varphi : (-\varepsilon, \varepsilon) \to R$ by $t \mapsto (u_0 + t, v_0)$, and then look at the mapping $\gamma = r \circ \varphi$, which sends $t \in (-\varepsilon, \varepsilon)$ to $r(u_0 + t, v_0)$. This makes γ a curve. Therefore,

$$\gamma' = \frac{\partial r}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial r}{\partial v}\frac{\partial v}{\partial t} = r_u \tag{2.50}$$

and $\gamma'(t_0) = r_u|_{(u_0,v_0)}$. If we work similarly for r_v, we obtain $\gamma'(t_0) = r_v|_{(u_0,v_0)}$. Thus, r_u and r_v are tangent vectors to the surface at the point $r(u_0, v_0)$. Since $r_u \times r_v \neq 0$, we see that r_u and r_v are linearly independent, and hence span the tangent plane to the surface at $r(u_0, v_0)$. The third property ensures that a tangent plane (and a normal vector) exists at every point on the surface.

Example 2.41. Let $O_2 \subseteq \mathbb{R}^2$ be a region, and let $f \in C^1(O_2)$ with bounded partials. Consider the graph of f, which is the set $\mathcal{G}(f) = \{(x, y, f(x, y)) \mid (x, y) \in O_2\}$. Define $r : O_2 \to \mathbb{R}^3$ as $r(u, v) = (u, v, f(u, v))$. Then $r_u = (1, 0, f_u)$ and $r_v = (0, 1, f_v)$. Clearly, r_u and r_v are bounded on O_2, and r is one-to-one. Finally, $r_u \times r_v = (-f_u, -f_v, 1) \neq 0$ for all $(u, v) \in O_2$. Thus, r is a surface parametrization and $\mathcal{G}(f)$ is a parametrized surface.

Example 2.42. Look at the equation of the torus given as $(x^2 + y^2 - a^2)^2 + z^2 = b^2$ for some $a > b > 0$. It is left as an exercise to the reader to show that this is a parametrized surface.

October 13th.

Example 2.43. A general example of a surface is *surface of revolution*. Consider a C^1-curve $t \mapsto (0, f(t), g(t)) \in \mathbb{R}^3$ for $t \in [a, b]$. Rotate this curve about the z-axis. The resulting surface is given as

$$r(u, v) = (f(u) \cos v, f(u) \sin v, g(u)) \tag{2.51}$$

for all $(u, v) \in [a, b] \times [0, 2\pi)$. Here, $r_u \times r_v = f(u)(-g'(u) \cos v, -g'(u) \sin v, f'(u))$. To make this non-zero, one can impose several constraints; for example, $f(u) \neq 0$ for all $u \in [a, b]$.

Example 2.44. The surface of a finite cylinder is parametrized as $r(u, v) = (a \cos v, a \sin v, u)$ for $(u, v) \in [0, b] \times [0, 2\pi)$ for some $a, b > 0$. One can verify that $r_u \times r_v \neq 0$ for all (u, v). Note that this is a surface of revolution with $f(u) = a$ and $g(u) = u$.

2.2.2 Tangent Plane and Normal

Let $r : R \to \mathbb{R}^3$ be a parametrized surface and let S be the range of r. Fix a point $P = r(u_0, v_0) \in S$.

Definition 2.45. $T_P S$ is defined to be the vector space 'generated' by $r_u(P)$ and $r_v(P)$. This is termed the *tangent plane* of S at P. Formally,

$$T_P S = \{\vec{r}_0 + t_1 r_u(P) + t_2 r_v(P) \mid t_1, t_2 \in \mathbb{R}\} \tag{2.52}$$

Example 2.46. Consider the graph function $z = f(x, y)$ for $(x, y) \in O_2 \subseteq \mathbb{R}^2$ open. Then $r(u, v) = (u, v, f(u, v))$ is the graph function. Here, $r_u \times r_v = (-f_u, -f_v, 1)$. Fix $P = (a, b, f(a, b))$. Then the tangent plane at P is given by

$$-f_u(a, b)(x - a) - f_v(a, b)(y - b) + (z - f(a, b)) = 0 \tag{2.53}$$

and the equation of the normal is

$$\frac{x - a}{-f_u(a, b)} = \frac{y - b}{-f_v(a, b)} = \frac{z - f(a, b)}{1}. \tag{2.54}$$

Example 2.47. Equation of the tangent plane and normal to $z = \frac{2x}{y} - x^2$ at $(1, 1, 1)$; here the partials are

$$\frac{\partial z}{\partial x} = \frac{2}{y} - 2x, \quad \text{and} \quad \frac{\partial z}{\partial y} = -\frac{2x}{y^2}. \tag{2.55}$$

The normal vector at $(1, 1, 1)$ is $(0, 2, 1)$. Thus the equation of the normal is

$$\frac{x - 1}{0} = \frac{y - 1}{2} = \frac{z - 1}{1} \tag{2.56}$$

and the equation of the tangent plane is $2(y - 1) + (z - 1) = 0$.

Remark 2.48. Consider the graph function $\{(x, y, f(x, y)) \mid (x, y) \in O_2\}$. Recalling the tangent plane from before, we can approximate $f(x, y)$ near (a, b) as

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b). \tag{2.57}$$

Example 2.49. Let us approximate $(1.99)^2 - \frac{1.99}{1.01}$ using the above method. Clearly a suitable function is $z = x^2 - \frac{x}{y}$ and a suitable point is $(2, 1)$. Here, $f(2, 1) = 2$, $f_x(x, y) = 2x - \frac{1}{y}$ and $f_y(x, y) = \frac{x}{y^2}$. Thus, $f_x(2, 1) = 3$ and $f_y(2, 1) = 2$. Therefore,

$$f(x, y) \approx 2 + 3(x - 2) + 2(y - 1) \tag{2.58}$$

for points (x, y) near $(2, 1)$. Setting $x = 1.99$ and $y = 1.01$, we get $f(1.99, 1.01) \approx 1.99$. The actual value is 1.9898.

Our goal, now, is to compute the area of surfaces. In the most simple case, a parallelogram in $\mathbb{R}^3$ with sides $\vec{u}_0$ and $\vec{v}_0$. The area of such a surface is simply $\|\vec{u}_0 \times \vec{v}_0\|$. Now consider the surface $\{(x, y, ax+by+c) \mid (x, y) \in B^2\}$. The area of this surface is simply $\sqrt{1 + a^2 + b^2} \cdot \text{area}(B^2) = \|r_x \times r_y\| \cdot \text{area}(B^2)$. This is because the normal vector to this surface is $(-a, -b, 1)$, and hence the area of the parallelogram formed by $(1, 0, a)$ and $(0, 1, b)$ is $\sqrt{1 + a^2 + b^2}$.

Let S be the range of r where $r : B^2 \to \mathbb{R}^3$ is a parametrized surface. Write

$$B^2 = \bigcup_{\alpha \in \Lambda(P)} B^2_\alpha \tag{2.59}$$

where P is a partition of B_2 and B_α^2 are small rectangles. Pick $x_\alpha \in \operatorname{int}(B_\alpha^2)$ for all $\alpha \in \Lambda(P)$. Then the approximate are of S induced by P is given as

$$\sum_{\alpha \in \Lambda(P)} \|r_u(x_\alpha) \times r_v(x_\alpha)\| \cdot \operatorname{area}(B_\alpha^2). \tag{2.60}$$

The area of the surface S is then defined to be the limit of the above expression as $\|P\| \to 0$, provided the limit exists. Thus this limit, for a parametrized curve $r : R \to \mathbb{R}^3$, is equal to

$$\int_R \|r_u \times r_v\| \, dA. \tag{2.61}$$

Example 2.50. Let $f \in C^1(R)$, for $R \subseteq \mathbb{R}^2$ a region. Consider $r : R \to \mathbb{R}^3$ defined as $r(u, v) = (u, v, f(u, v))$. Therefore the area of the graph of f over R is given as

$$\int_R \sqrt{f_u^2 + f_v^2 + 1} dA. \tag{2.62}$$

October 14th.

Example 2.51. We compute the area of the surface parametrized by $r(x, y) = (\cos x, \sin x, y)$ on the region $0 \le x \le \pi/2$ and $0 \le y \le 1$. Here $r_x = (-\sin x, \cos x, 0)$ and $r_y = (0, 0, 1)$. Thus $r_x \times r_y = (\cos x, \sin x, 0)$ and $\|r_x \times r_y\| = 1$. Therefore the area is simply $\int_0^{\pi/2} \int_0^1 1 dy dx = \pi/2$.

Example 2.52. We compute the area of the unit hemisphere. In spherical coordinates, the region we're dealing with is $R = \{(u, v) \in \mathbb{R}^2 \mid 0 < u, v < \pi\}$ and the parametrization is $r(u, v) = (\sin u \cos v, \sin u \sin v, \cos u)$ for all $(u, v) \in R$. Here $\|r_u \times r_v\| = |\sin u|$. Thus the area of the surface is

$$\int_R \sin u dA = 2\pi. \tag{2.63}$$

So far, we have computed the line integral of both scalar and vector fields. We have also computed the area of surfaces. We now move on to computing the surface integral of scalar and vector fields.

Definition 2.53. Let $S \subseteq O_3$ be a surface in an open set $O_3 \subseteq \mathbb{R}^3$. Let $f \in C(O_3)$. As in the area computation,

$$\lim_{\|P\| \to 0} \sum_{\alpha \in \Lambda(P),\, x_\alpha \in B_\alpha^2} f(r(x_\alpha)) \|r_u(x_\alpha) \times r_v(x_\alpha)\| \cdot \operatorname{area}(B_\alpha^2) = \int_R f \circ r \|r_u \times r_v\| \, dA. \tag{2.64}$$

The *surface integral* of f over S is defined to be the above limit, provided it exists. The notation for this is

$$\int_S f ds = \int_R f \circ r \|r_u \times r_v\| \, dA. \tag{2.65}$$

One can verify that for a reparametrization $\tilde{r}$ such that $r = \tilde{r} \circ \varphi$ for a continuously differentiale bijective φ, the surface integral remains unchanged over the new region $\tilde{R}$ and

$$\int_R f \circ r \|r_u \times r_v\| \, dA = \int_{\tilde{R}} f \circ \tilde{r} \|\tilde{r}_u \times \tilde{r}_v\| \, dA. \tag{2.66}$$

Example 2.54. For a surface given by $z = \sqrt{x^2 + y^2}$ for $0 \le z \le 1$, let us compute $\int_S (x^2 + y^2 + z^2) ds$. A parametrization is $r(x, y) = (x, y, \sqrt{x^2 + y^2})$ for $(x, y) \in R = \{(x, y) \mid x^2 + y^2 \le 1\}$. Noting that

$r_u = \frac{x}{\sqrt{x^2+y^2}}$ and $r_v = \frac{y}{\sqrt{x^2+y^2}}$, we have

$$\int_S (x^2 + y^2 + z^2) ds = \int_R (x^2 + y^2 + (x^2 + y^2)) \left\| r_u \times r_v \right\| dA = 2\sqrt{2} \int_R \sqrt{x^2 + y^2} dA = \sqrt{2}\pi. \quad (2.67)$$

We now define the surface integral of a vector field.

Definition 2.55. A surface $S \subseteq \mathbb{R}^3$ is oriented if there exists a continuous vector field $\vec{n} : S \to \mathbb{R}^3$ such that $\vec{n}(x)$ is a unit normal vector to the tangent plane $T_x S$ for all $x \in S$.

As an example, the Möbis strip is not orientable, but the sphere is orientable.

Definition 2.56. The surface integral of a vector field $\vec{F} : S \to \mathbb{R}^3$ over an oriented surface S with orientatio $\vec{n}$ is defined as

$$\int_S \vec{F} \cdot d\vec{s} = \int_S \vec{F} \cdot \vec{n} ds \quad (2.68)$$

38

Index