Class of Excellence in Mathematics, HIT

LECTURE NOTES ON MATHEMATICAL ANALYSIS III

Class of Excellence in Mathematics, HIT

Version: Autumn Semester, 2025
Last compiled: September 8, 2026

LECTURE NOTES ON
MATHEMATICAL ANALYSIS III

Class of Excellence in Mathematics, HIT

CONTENTS

CHAPTER 1

DIFFERENTIAL CALCULUS

> Cette formulation «intrinsèque» du Calcul infinitésimal, parce qu'elle est plus abstraite et surtout parce qu'elle nécessite l'abandon des espaces de départ et la construction répétée des nouveaux «espaces fonctionelles» (en particulier lorsqu'on étudie les dérivées d'ordre supérieure) demande certainement un effort intellectuel qui contraste avec la routine machinale des formules classiques. Mais nous croyons que le résultat en vaut la peine, car il prépare l'étudiant à l'idée plus générale du Calcul différentiel sur une variété différentielle...
>
> *Éléments d'Analyse, page 149, Tome I.*
> Jean Dieudonné

We are going to develop differential calculus in the setting of Banach spaces, which are much more general than the usual Euclidean space $\mathbb{R}^n$. This approach, at the obvious cost of demanding the students to think more abstractly and conceptually than the traditional method of working through partial derivatives, will prove to be highly elegant and efficient, especially for our later treatment of the basics of the so-called "global analysis" on a differentiable manifold, which in turn, opens the gate to a huge chunk of modern mathematics and theoretical physics.

1.1. Review of Normed and Banach Spaces

Let X be a set, and $\mathfrak{B}(X)$ be the vector space of all *bounded* functions on X. For each $f \in \mathfrak{B}(X)$, observe that the number $\|f\| := \sup\{|f(x)| \mid x \in X\}$ is finite, and satisfies the property that for all $f, g \in \mathfrak{B}(X)$ and $c \in \mathbb{R}$:

(1) $\|f + g\| \leq \|f\| + \|g\|$;
(2) $\|cf\| = |c| \cdot \|f\|$;
(3) $\|f\| = 0$ if and only if $f = 0$.

The notion of a norm and normed space is a consequence of abstracting the above observation.

Definition 1.1.1. — A **norm** on a vector space V (not assuming to be finite dimensional) is a map $\|\cdot\| : V \to [0, +\infty[$, such that for all $x, y \in V$ and $c \in \mathbb{R}$, we have

(1) (the triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$;
(2) $\|cx\| = |c| \cdot \|x\|$;
(3) $\|x\| = 0$ if and only if $x = 0$.

A **normed space** is pair $(V, \|\cdot\|)$ where V is a vector space, and $\|\cdot\|$ is a norm on it. For ease of use, we often speak of a normed space V instead of the more cumbersome $(V, \|\cdot\|)$, omitting the norm $\|\cdot\|$ when it is obvious from context.

Given a normed space V, canonically, there is a metric d on V associated with its norm, namely, $d(x,y) := \|x-y\|$. One may then apply all the theory of metric spaces on V. The following notion is particularly important.

Definition 1.1.2. — A normed space V is called a **Banach space** if it is complete as a metric space.

Example 1.1.3. — (Exercise) The normed space $\mathfrak{B}(X)$ considered above is a Banach space.

Example 1.1.4. — (Exercise) Let $I = [a,b]$ be a closed interval and $k \in \mathbb{N}$. Define a norm $\|\cdot\|$ on $C^k(I)$ by

$$\|f\| := \sum_{l=0}^{k} \sup\Big\{\Big|f^{(l)}(x)\Big| \;\Big|\; x \in I\Big\}. \tag{1.1.1}$$

Then $\big(C^k(I), \|\cdot\|\big)$ is a Banach space.

Example 1.1.5. — (Exercise) On $\mathbb{R}^n$, there are uncountably different norms. For example, given $p \in [1,\infty]$, one defines the p-norm $\|\cdot\|_p$ on $\mathbb{R}^n$ as follows: if $p \in [1,+\infty[$, then set

$$\|(x_i)_{i=1}^n\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}; \tag{1.1.2}$$

whereas

$$\|(x_i)_{i=1}^n\|_\infty = \sup\big\{|x_i| \;\big|\; 1 \le i \le n\big\}. \tag{1.1.3}$$

Under each of these norms, $\mathbb{R}^n$ is a Banach space.

It is interesting to note for questions concerning the topology of $\mathbb{R}^n$, all of the norms in Example 1.1.5 make no essential difference. The precise formulation for this is given by the notion of equivalent norms.

Definition 1.1.6. — On a vector space V, two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are said to be **equivalent**, if there exists $c, C \in \,]0,+\infty[$, such that for all $x \in V$, we have

$$c\|x\|_1 \le \|x\|_2 \le C\|x\|_1. \tag{1.1.4}$$

Proposition 1.1.7. — *The following hold:*

(1) *Equivalent norms on the same vector space determine exactly the same open (resp. closed) sets;*
(2) *On a* finite dimensional *vector space, all norms are equivalent.*

Proof. — Exercise. □

The following corollary shall be important when we treat partial derivatives.

Corollary 1.1.8. — *Let $E_1, \ldots, E_n$ be normed spaces, and $E = \prod_{i=1}^n E_i$ (direct product of vector spaces). Denote the norm on E_i by $\|\cdot\|_i$ for each $1 \le i \le n$, and choose any norm α on $\mathbb{R}^n$ such that $\alpha(r_1,\ldots,r_n) \le \alpha(s_1,\ldots,s_n)$ whenever $0 \le r_i \le s_i$ for any $i \in \{1,\ldots,n\}$, then*

$$\|(x_1,\ldots,x_n)\|_\alpha := \alpha\big(\|x_1\|,\ldots,\|x_n\|\big) \tag{1.1.5}$$

is a well-defined norm on E. Moreover, up to equivalence, the norm $\|\cdot\|_\alpha$ on E does not depend on the choice of α; and E is a Banach space if and only if each E_i is a Banach space.

Proof. — Exercise. □

Once we have the notion of normed and Banach spaces, the next natural step is to study the "nice maps" among such spaces. For our current purpose of setting up the stage for doing differential calculus, the precise meaning of "nice maps" will be the ones that are both *linear*, i.e. respect the linear structure, and *continuous*, i.e. respect the topological structure. Linear maps between normed spaces are often referred as *linear operators*.

Now given two normed space E and F, and let $L(E,F)$ denote the space of all linear operators from E to F, which is still a vector space by setting $S+T \in L(E,F)$ and $cS \in L(E,F)$, for $S,T \in L(E,F)$ and $c \in \mathbb{R}$, to be the linear maps that are defined by $(S+T)x = Sx+Tx$ and $(cS)x = c(Sx)$. With each

$S \in L(E,F)$, one may associate a number $\|S\|$, which can be $+\infty$, that measures the size of S:

$$\begin{aligned}\|S\| &= \inf\Big\{c \in [0,+\infty] \;\Big|\; \forall x \in E,\ \|Sx\| \leq c\|x\|\Big\}\\ &= \sup\Big\{\|Sx\| \;\Big|\; \|x\| \leq 1\Big\}.\end{aligned} \tag{1.1.6}$$

We then have the following result characterizing whether a linear operator is "nice".

Proposition 1.1.9. — *Let E and F be normed spaces. Then the following hold.*

(1) *A linear operator $T \in L(E,F)$ is continuous if and only if $\|T\| < +\infty$, in which case we say T is* ***bounded****.*

(2) *Let $\mathcal{L}(E,F)$ denote the subset of $L(E,F)$ consisting of all bounded linear operators. Then $\mathcal{L}(E,F)$ is a subspace of $L(E,F)$, and $\|\cdot\|$ as defined in (1.1.6) is a norm on $\mathcal{L}(E,F)$.*

(3) *If E is not the zero space, then (1.1.6) can be defined equivalently as*

$$\|S\| = \sup\Big\{\|Sx\| \;\Big|\; x \in E \ \text{ and } \ \|x\| = 1\Big\} = \sup\left\{\frac{\|Sx\|}{\|x\|} \;\middle|\; x \in E \setminus \{0\}\right\}. \tag{1.1.7}$$

(4) *If E is* finite dimensional*, then $L(E,F) = \mathcal{L}(E,F)$, i.e. every linear operator from E is automatically bounded*[(1)]*.*

(5) *If F is a Banach space, then $\mathcal{L}(E,F)$ is a Banach space.*

Proof. — Exercise. □

Remark 1.1.10. — Conceptually, we can now already observe an advantage of working with normed spaces, and in particular, with Banach spaces: one may form the mapping space of continuous linear maps in a natural way, and still retain a natural normed (Banach) space structure. Once we have chosen a norm on $\mathbb{R}^n$ and $\mathbb{R}^m$ (with $m, n \in \mathbb{N}$), say both with the **Euclidean norm** $\|\cdot\|_2$, if we want to identify $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) = L(\mathbb{R}^n, \mathbb{R}^m)$ with $\mathbb{R}^{mn}$ (say via matrices), then it is not obvious which should be the "correct" norm on $\mathbb{R}^{mn}$, but adopting the normed space point of view, one sees immediately that the "correct norm" on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is exactly the operator norm as defined in (1.1.6) (which is different from the norm induced by the 2-norm on $\mathbb{R}^{mn}$ via the previous identification, unless $m = 0$ or $n = 0$). Granted, for finite dimensional spaces, this conceptual advantage plays no essential role as all norms will be equivalent (see Proposition 1.1.9). However, as is usually the case in mathematics, by clearly pointing out precisely what is essential, conceptual advantage often leads to deeper understanding and clearer proofs, as will be demonstrated later in these notes.

We now digress a little to introduce a process of transforming a two variable map into a map of a single variable, which is also known as *the currying*. For any two sets A, B, we use $\mathcal{F}(A,B)$ to denote the set of all maps from A to B. The currying process is captured in the following result.

Proposition 1.1.11. — *Let X, Y and Z be sets. Then there exists a natural*[(2)] *bijection*

$$\begin{aligned}\Phi_{X,Y;Z} : \mathcal{F}(X \times Y, Z) &\to \mathcal{F}\big(X, \mathcal{F}(Y,Z)\big)\\ \varphi &\mapsto \varphi^{\vee},\end{aligned}$$

where for each $x \in X$, $\varphi^{\vee}(x)$ is the map sending each $y \in Y$ to $\varphi(x,y) \in Z$; and the inverse of $\Phi_{X,Y;Z}$ is given by

$$\begin{aligned}\Psi_{X,Y;Z} : \mathcal{F}\big(X, \mathcal{F}(Y,Z)\big) &\to \mathcal{F}(X \times Y, Z)\\ \psi &\mapsto \psi^{\wedge},\end{aligned}$$

where $\psi^{\wedge}(x,y) = [\psi(x)](y) \in Z$ for each $(x,y) \in X \times Y$.

Proof. — That $\Phi_{X,Y;Z}$ and $\Psi_{X,Y;Z}$ are inverses of each other can be seen by a careful unwinding of the definition, hence they are both bijective. □

---

(1) However, this is no longer true as soon as E becomes infinite dimensional.

(2) For now, one may treat "natural" simply as an adjective. It actually has a precise meaning in a branch of mathematics called *category theory.*

The currying will be particularly useful when we consider the second derivative, and we now set up the preparation in the setting of normed spaces. Let E_1, E_2 and F be normed spaces, and use $B(E_1, E_2; F)$ to denote the set of all bilinear maps from $E_1 \times E_2$ to F. With any $\psi \in B(E_1, E_2; F)$, similar to (1.1.6), we may still associate a number $\|\psi\|$ (could be $+\infty$) that measures the size of ψ, which is given by

$$\begin{aligned} \|\psi\| &= \inf\Big\{c \in [0, +\infty] \;\Big|\; \forall x_1 \in E_1, \forall x_2 \in E_2,\ \|\psi(x_1, x_2)\| \le c\|x_1\| \cdot \|x_2\|\Big\} \\ &= \sup\Big\{\|\psi(x_1, x_2)\| \;\Big|\; x_1 \in E_1,\ x_2 \in E_2, \ \text{ and } \ \|x_1\| \le 1,\ \|x_2\| \le 1\Big\}. \end{aligned} \tag{1.1.8}$$

Again, we say ψ is **bounded** if $\|\psi\| < +\infty$.

Proposition 1.1.12. — *Let E_1, E_2 and F be normed spaces, and $\mathcal{B}(E_1, E_2; F)$ the subset of $B(E_1, E_2; F)$ consisting of all bounded bilinear maps. Then the following hold.*

(1) *A map $\varphi \in B(E_1, E_2; F)$ if and only if it is continuous at 0, which in turn is equivalent to $\|\varphi\| < +\infty$, i.e. $\varphi \in \mathcal{B}(E_1, E_2; F)$.*
(2) *The subset $\mathcal{B}(E_1, E_2; F)$ is a linear subspace of $B(E_1, E_2; F)$, and the map $\|\cdot\|$ defined in (1.1.8) is a norm on $\mathcal{B}(E_1, E_2; F)$.*
(3) *The currying map $\Phi_{E_1,E_2;F}$ in Proposition 1.1.11 restricts to a bijection from $\mathcal{B}(E_1, E_2; F)$ onto $\mathcal{L}\big(E_1, \mathcal{L}(E_2, F)\big)$ that is an isometric isomorphism of normed spaces.*
(4) *If F is a Banach space, so is the normed space $\mathcal{B}(E_1, E_2; F)$.*

This currying process can be generalized for an arbitrary finite number of sets, and the corresponding result, in the setting of normed (Banach) spaces, will be used ubiquitously in our later systematic treatment of higher derivatives.

Proposition 1.1.13. — *Let $X_1, \ldots, X_n$ and Y be sets. The map*

$$\begin{aligned} \Phi_{X_1,\ldots,X_n;Y} : \mathcal{F}\big(X_1 \times X_2 \times \cdots \times X_n, Y\big) &\to \mathcal{F}\Big(X_1, \mathcal{F}\big(X_2, \mathcal{F}(X_3, \cdots, \mathcal{F}(X_n, Y))\big)\Big) \\ \varphi &\mapsto \varphi^\vee \end{aligned}$$

is a natural bijection, where for each $x_1 \in X_1$, $\varphi^\vee(x_1)$ is a map on X_2, and for each $x_2 \in X_2$, $\varphi^\vee(x_1)(x_2)$ is a map on X_3, ..., and finally, for each $x_n \in X_n$, $\varphi^\vee(x_1)(x_2)(x_3)\cdots(x_n) \in Y$, which is defined to be

$$\varphi^\vee(x_1)(x_2)(x_3)\cdots(x_n) = \varphi(x_1, x_2, \ldots, x_n). \tag{1.1.9}$$

And the inverse of $\Phi_{X_1,\ldots,X_n;Y}$ is given by

$$\begin{aligned} \Psi_{X_1,\ldots,X_n;Y} : \mathcal{F}\Big(X_1, \mathcal{F}\big(X_2, \mathcal{F}(X_3, \cdots, \mathcal{F}(X_n, Y))\big)\Big) &\to \mathcal{F}(X_1 \times X_2 \times \cdots \times X_n; Y) \\ \psi &\mapsto \psi^\wedge, \end{aligned}$$

where

$$\psi^\wedge(x_1, \ldots, x_n) = \psi(x_1)(x_2)\cdots(x_n). \tag{1.1.10}$$

Remark 1.1.14. — This abstract result is the prototype in the so-called *lambda calculus*, which is important in computer science.

Let E_1, E_2, ..., E_n and F be normed spaces. We say a map

$$f : E_1 \times E_2 \times \cdots \times E_n \to F$$

is **multilinear**, if for each $1 \le i \le n$, and each fixed $v_j \in E_j$ for $j \ne i$, the map

$$\begin{aligned} f^{(i)}_{v_1,\ldots,v_{i-1},v_i,\ldots,v_n} : E_i &\to F \\ x_i &\mapsto f(v_1, \ldots, v_{i-1}, x_i, v_{i+1}, \ldots, v_n) \end{aligned}$$

is linear. Let $M(E_1, \ldots, E_n; F)$ denote the space of all multilinear maps from $E_1 \times E_2 \times \cdots \times E_n$ to F. With each $f \in M(E_1, \ldots, E_n; F)$, we again associate a number (possibly $+\infty$) $\|f\|$ to measure the size of f, namely

$$\begin{aligned} \|f\| &= \inf\big\{c \in [0, +\infty] \;\big|\; \forall 1 \le i \le n, \forall x_i \in E_i,\ \|f(x_1, \ldots, x_n)\| \le c\|x_1\| \cdots \|x_n\|\big\} \\ &= \sup\Big\{\|f(x_1, \ldots, x_n)\| \;\Big|\; \forall 1 \le i \le n,\ x_i \in E_i \ \text{ and } \ \|x_i\| \le 1\Big\}. \end{aligned} \tag{1.1.11}$$

Again, we say that f is **bounded** if $\|f\| < +\infty$.

We now have the following generalization of Proposition 1.1.12.

Proposition 1.1.15. — *Let $E_1, \ldots, E_n$ and F be normed spaces, and $\mathcal{M}(E_1, \ldots, E_n; F)$ the subset of $M(E_1, \ldots, E_n; F)$ consisting of all bounded bilinear maps. Then the following hold.*

(1) *A map $\varphi \in M(E_1, \ldots, E_n; F)$ if and only if it is continuous at 0, which in turn is equivalent to $\|\varphi\| < +\infty$, i.e. $\varphi \in \mathcal{M}(E_1, \ldots, E_n; F)$.*
(2) *The subset $\mathcal{M}(E_1, \ldots, E_n; F)$ is a linear subspace of $M(E_1, \ldots, E_n; F)$, and the map $\|\cdot\|$ defined in (1.1.11) is a norm on $\mathcal{M}(E_1, \ldots, E_n; F)$.*
(3) *The currying map $\Phi_{E_1,\ldots,E_n;F}$ in Proposition 1.1.13 restricts to a bijection from $\mathcal{M}(E_1, \ldots, E_n; F)$ onto $\mathcal{L}\Big(E_1, \mathcal{L}\big(E_2, \ldots, \mathcal{L}(E_n, F)\big)\Big)$ that is an isometric isomorphism of normed spaces.*
(4) *If F is a Banach space, so is the normed space $\mathcal{M}(E_1, \ldots, E_n; F)$.*

1.2. The Derivative

1.2.1. The infinitesimal. — Right in the beginning phase of calculus, one principle idea is that derivatives should be something that approximate the functions to be studied infinitesimally. But due to the lack of a precise notion of the limit, the original crude version of the infinitesimal calculus soon leads to some paradoxes, causing the so-called the second mathematical crisis. In dealing with this crisis, it takes the effort of several generations of mathematicians to make the idea of infinitesimal precise and rigorous, as a certain sort of limit.

Here is the precise formulation of the notion of an infinitesimal.

Definition 1.2.1. — Let E, F be normed spaces. An **infinitesimal** from E to F is a map f that is defined on some neighborhood $V \subseteq E$ of 0 into F, such that $f(0) = 0$ and f is continuous at 0, i.e. $\lim_{x\to 0} f(x) = f(0)$. The collection of all infinitesimals from E to F is denoted by $\mathfrak{I}(E, F)$.

We say f is a **locally defined Lipschitz map** (from E to F) at 0 if f is defined on a neighborhood V of 0 with values in F, such that there exists $c \geq 0$ with $\|f(x)\| \leq c\|x\|$ for all $x \in V$. The collection of locally defined Lipschitz map from E to F is denoted by $\mathfrak{O}(E, F)$.

We say f is an **infinitesimal of higher order** from E to F, if f is defined on some neighborhood V of 0 such that for any $\varepsilon > 0$, there exists a neighborhood W_ε of 0 with $W \subseteq V$, such that

$$\forall x \in W_\varepsilon, \qquad \|f(x)\| \leq \varepsilon\|x\|.$$

The collection of all infinitesimals of higher order from E to F is denoted by $\mathfrak{o}(E, F)$.

Remark 1.2.2. — (Exercise) It can be seen that the collections $\mathfrak{o}(E, F)$, $\mathfrak{O}(E, F)$ and $\mathfrak{I}(E, F)$ are *unchanged* after changing the norm on E (resp. F) to an equivalent one.

Notation 1.2.3. — *For $f, g \in \mathfrak{I}(E, F)$, we define the sum $f + g$ to be the map*

$$(f \restriction_{\mathrm{dom}(f)\cap\mathrm{dom}(g)}) + (g \restriction_{\mathrm{dom}(f)\cap\mathrm{dom}(g)}),$$

which is defined on the neighborhood $\mathrm{dom}(f)\cap\mathrm{dom}(g)$ of 0 in E. Similarly, we define scalar multiplication of infinitesimal functions: for $f \in \mathfrak{I}(E, F)$, $c \in \mathbb{R}$, we set cf to be the map $x \in \mathrm{dom}(f) \to cf(x) \in F$. One now checks that $\mathfrak{I}(E, F)$ is stable under taking sums and scalar multiplications. We shall denote $(-1)f$ by $-f$ and writing $f - g$ for $f + (-g)$.

Remark 1.2.4. — It is interesting to note that $\mathfrak{I}(E, F)$ is *not* a vector space. It does satisfy most of the axioms for a vector space. The axiom that fails is the existence of the opposite vector. More precisely, it is clear that the zero map $0 : E \to F$, $x \mapsto 0$ that is defined everywhere is the unique neutral element under taking sums, but then for any map f in $\mathfrak{I}(E, F)$ that is not defined everywhere, there does *not* exist any $g \in \mathfrak{I}(E, F)$ with $f + g = 0$, for $f + g$ is not defined outside $\mathrm{dom}(f)$. Note, however, we do have $f + (-f) = 0_{\mathrm{dom}(f)}$, where $-f := (-1)f$ and $0_{\mathrm{dom}\, f}$ is the zero map defined on $\mathrm{dom}(f)$. The way to remedy this, i.e. to extract a vector space out of $\mathfrak{I}(E, F)$, is to introduce the notion of *germs* of functions, i.e. we consider the quotient $\mathfrak{I}(E, F)/\sim$ where the equivalence relation $\sim$ is given by letting $f \sim g$ if and only if f and g restricts to the same map on some neighborhood of 0. One checks that the sum and scalar multiplication on $\mathfrak{I}(E,F)$ then induces a well-defined sum and scalar multiplication on $\mathfrak{I}(E,F)/\sim$, making the latter a vector space. For an $f \in \mathfrak{I}(E,F)$, the equivalence class $[f] \in \mathfrak{I}(E,F)/\sim$ is called the *germ* of f. When considering infinitesimals, usually it is the germ of the infinitesimal, rather than the infinitesimal itself, that is essential.

Proposition 1.2.5. — *Let E, F, G be normed spaces.*

(1) *We have*

$$\mathfrak{o}(E,F) \subseteq \mathfrak{O}(E,F) \subseteq \mathfrak{I}(E,F), \tag{1.2.1}$$

and both $\mathfrak{o}(E,F)$ and $\mathfrak{O}(E,F)$ are stable under taking sums and scalar multiplications.

(2) *Let $f \in \mathfrak{I}(E,F)$, $g \in \mathfrak{I}(F,G)$, then $g \circ f := g \circ \left(f \restriction_{f^{-1}(\mathrm{dom}(g))}\right)$ is a well-defined infinitesimal from E to F. Moreover, if either of the following holds*

(2.a) *$f \in \mathfrak{O}(E,F)$ and $g \in \mathfrak{o}(E,F)$;*

(2.b) *$f \in \mathfrak{o}(E,F)$ and $g \in \mathfrak{O}(E,F)$;*

then $f \circ g \in \mathfrak{o}(E,F)$.

(3) *We have $\mathfrak{o}(E,F) \cap \mathcal{L}(E,F) = \{0\}$.*

Proof. — (1) is a routine check and is left as an exercise.

(2). Since f is continuous at 0, the domain $\mathrm{dom}(g \circ f) = f^{-1}(\mathrm{dom}(g))$ is a neighborhood of 0. Since both g and $f \restriction_{f^{-1}(\mathrm{dom}(g))}$ are continuous at 0 and $f(0) = 0$, the composite $g \circ f$ is continuous at 0, and indeed $g \circ f \in \mathfrak{I}(E,G)$.

Suppose (2.a) holds. Define $h : \mathrm{dom}(g \circ f) \to G$ by setting

$$h(\xi) = \begin{cases} \dfrac{g(f(\xi))}{\|f(\xi)\|}, & \text{if } f(\xi) \neq 0 \\ 0, & \text{if } f(\xi) = 0. \end{cases} \tag{1.2.2}$$

Then for any $\xi \neq 0$ and $\xi \in \mathrm{dom}(g \circ f) = \mathrm{dom}(h)$, we have

$$0 \leq \frac{\|g(f(\xi))\|}{\|\xi\|} = \frac{\|f(\xi)\|}{\|\xi\|}\|h(\xi)\| \leq c\|h(\xi)\|, \tag{1.2.3}$$

where $c \geq 0$ is a constant with $\|f(\xi)\| \leq c\|\xi\|$ for all $\xi \in \mathrm{dom}(f)$, whose existence follows from the assumption that $f \in \mathfrak{O}(E,F)$. By (1.2.3), we will have $g \circ f \in \mathfrak{o}(E,G)$ as desired if we can show that $h \in \mathfrak{I}(E,G)$, i.e. if $\lim_{\xi \to 0} h(\xi) = 0$, which we now establish. Take any $\varepsilon > 0$. Since $g \in \mathfrak{o}(F,G)$, we can find a neighborhood V of 0 in F, such that

$$\forall \eta \in V, \qquad \|g(\eta)\| \leq \varepsilon\|\eta\|. \tag{1.2.4}$$

Then by continuity of f at 0 (note $f \in \mathfrak{O}(E,F) \subseteq \mathfrak{I}(E,F)$ by (1)), the set $U := f^{-1}(V) \subseteq \mathrm{dom}(f)$ is a neighborhood of 0 in E. For any $\xi \in U$, either $f(\xi) = 0$, in which case $h(\xi) = 0$; or $f(\xi) \neq 0$, in which case $\|h(\xi)\| \leq \varepsilon$ by setting $\eta = f(\xi)$ in (1.2.4). Thus $\|h(\xi)\| \leq \varepsilon$ for any $\xi \in U$, and we indeed have $\lim_{\xi \to 0} h(\xi) = 0$. This proves $g \circ f \in \mathfrak{o}(E,G)$ provided (2.a) holds.

Suppose (2.b) holds. Then for any $\varepsilon > 0$, there exists, by $f \in \mathfrak{o}(E,F)$, a neighborhood U of 0 in E, such that

$$\forall \xi \in U, \qquad \|f(\xi)\| \leq \varepsilon\|\xi\|. \tag{1.2.5}$$

By $g \in \mathfrak{O}(F,G)$, we also have a constant $C \geq 0$ with

$$\forall \eta \in \mathrm{dom}(g), \qquad \|g(\eta)\| \leq C\|\eta\|. \tag{1.2.6}$$

Combining (1.2.5) and (1.2.6), we obtain

$$\forall \xi \in \mathrm{dom}(g \circ f) = f^{-1}(\mathrm{dom}(g)), \qquad \|g(f(\xi))\| \leq C\|f(\xi)\| \leq C\varepsilon\|\xi\|. \tag{1.2.7}$$

Since $\varepsilon > 0$ is arbitrary, (1.2.7) implies $g \circ f \in \mathfrak{o}(E,G)$, as desired.

(3). Suppose $A \in \mathfrak{o}(E,F) \cap \mathcal{L}(E,F)$. It suffices to show that $A\xi_0 = 0$ for any nonzero $\xi_0 \in E$. Take any nonzero scalar $c \in \mathbb{R}$, we have

$$\frac{A\xi_0}{\|\xi_0\|} = \frac{A(c\xi_0)}{\|c\xi_0\|}. \tag{1.2.8}$$

By letting $c \to 0$ on the right side of (1.2.8) and noting that $A \in \mathfrak{o}(E,F)$, we obtain

$$\frac{A\xi_0}{\|\xi_0\|} = \lim_{c\to 0} \frac{A(c\xi_0)}{\|c\xi_0\|} = \lim_{\xi\to 0} \frac{A\xi}{\|\xi\|} = 0. \tag{1.2.9}$$

Thus $A\xi_0 = 0$, and the proof is complete. □

Given $f \in \mathfrak{I}(E,F)$ and a function g from a neighborhood W of 0 to $\mathbb{R}$, we may also define gf as the map $x \in \mathrm{dom}(f) \cap \mathrm{dom}(g) \to g(x)f(x) \in F$ with the neighborhood $\mathrm{dom}(f) \cap \mathrm{dom}(g)$ of 0 as domain.

Proposition 1.2.6. — *Using the above notation, the following hold:*

(1) *If $f \in \mathfrak{I}(E,F)$ and g is bounded, then $gf \in \mathfrak{I}(E,F)$.*
(2) *If $f \in \mathfrak{O}(E,F)$ (resp. $f \in \mathfrak{o}(E,F)$) and g is bounded, then $gf \in \mathfrak{o}(E,F)$.*
(3) *If $f \in \mathfrak{O}(E,F)$ and $g \in \mathfrak{O}(E,\mathbb{R})$, then $gf \in \mathfrak{o}(E,F)$.*

Proof. — Exercise. □

1.2.2. The derivative as a continuous linear map. — Let E, F be normed spaces, $U \subseteq E$ an open set in E and $x_0 \in U$. Intuitively, a map $f : U \to F$ is differentiable at x_0 should mean that near x_0, up to a negligible relative error as one gets nearer and nearer x_0, the map f looks like a continuous affine map (or equivalently, a continuous linear map if we rebase the origin at x_0). The precise definition is the following.

Definition 1.2.7. — Let E, F be normed spaces, $U \subseteq E$ an open set in E and $x_0 \in U$. A map $f : U \to F$ is **differentiable at** x_0, if there exists an $A \in \mathcal{L}(E,F)$ such that the "relative error map"

$$\begin{aligned} \mathcal{E}_{r,f,A} &: W \to F \\ \xi &\mapsto \begin{cases} \dfrac{(\Delta_{x_0} f)(\xi) - A\xi}{\|\xi\|}, & \text{if } \xi \neq 0 \\ 0, & \text{if } \xi = 0 \end{cases} \end{aligned} \tag{1.2.10}$$

is an infinitesimal from E to F, where W is a neighborhood of 0 such that $x_0 + W \subseteq U$, and $\Delta_{x_0} f : W \to F$ is the "incremental map" defined as

$$\begin{aligned} \Delta_{x_0} f &: W \to F \\ \xi &\mapsto f(x_0+\xi) - f(x_0). \end{aligned} \tag{1.2.11}$$

Equivalently, this means that the "error map"

$$\begin{aligned} (\Delta_{x_0} f - A) &: W \to F \\ \xi &\mapsto (\Delta_{x_0} f)(\xi) - A\xi = f(x_0+\xi) - f(x_0) - A\xi \end{aligned} \tag{1.2.12}$$

is an infinitesimal of higher order, i.e. $\Delta_{x_0} f - A \in \mathfrak{o}(E,F)$.

Note that in this case, the continuous linear operator A is unique, for if $B \in \mathcal{L}(E,F)$ and $\Delta_{x_0} f - B \in \mathfrak{o}(E,F)$, then (see Proposition 1.2.5 (3))

$$A - B = (\Delta_{x_0} f - B) - (\Delta_{x_0} f - A) \in \mathfrak{o}(E,F) \cap \mathcal{L}(E,F) = \{0\},$$

and we must have $B = A$. We thus call A **the derivative of** f at x_0.

Notation 1.2.8. — *Some common notation for the derivative of f at x_0 are*

$$(\mathrm{d}f)(x_0), \quad \mathrm{d}_{x_0} f, \quad (\mathrm{D}\, f)(x_0) \quad \textit{or} \quad \mathrm{D}_{x_0} f.$$

Remark 1.2.9. — In consistent with Remark 1.2.4, the differentiability of f at x_0 is really a notion about the germ of f at x_0 and so is the determination of the derivative $(\mathrm{d}f)(x_0)$, i.e. the existence and the value of $(\mathrm{d}f)(x_0)$ only depends on the value of f on a neighborhood of x_0, as can be easily checked by manipulating the notion of open sets.

Remark 1.2.10. — (Exercise) We will see in the exercise session that the differentiability of f at x_0, as well as the derivative $(\mathrm{d}f)(x_0)$ as a continuous linear map, does not depend on the choices of the equivalent norms on E and F.

Example 1.2.11. — It is clear that if $f = A\restriction_U$ for some $A \in \mathcal{L}(E, F)$, then f is differentiable at any x_0 with $(\mathrm{d}f)(x_0) = A$. In this case, the error map is identically zero: the derivative A is not merely the best approximation of f by continuous linear maps near x_0, it coincides with f near x_0.

Proposition 1.2.12. — *Using the setting in Definition 1.2.7, if f is differentiable at x_0, then f is continuous at x_0.*

Proof. — Continuity of f at x_0 is equivalent to the continuity of $\Delta_{x_0} f$ at 0. But from

$$\Delta_{x_0} f - A \in \mathfrak{o}(E, F) \subseteq \mathfrak{O}(E, F) \subseteq \mathfrak{I}(E, F) \tag{1.2.13}$$

and

$$A \in \mathfrak{O}(E, F) \subseteq \mathfrak{I}(E, F), \tag{1.2.14}$$

we deduce that

$$\Delta_{x_0} f = \big(\Delta_{x_0} f - A\big) + A \in \mathfrak{O}(E, F) \subseteq \mathfrak{I}(E, F). \tag{1.2.15}$$

and $\Delta_{x_0} f$ is indeed continuous at 0. □

1.2.3. The chain rule. — Let E, F, G be normed spaces, U (resp. V) is an open set in E (resp. F), and $x_0 \in U$ (resp. $y_0 \in V$). Given maps $f : U \to V \subseteq F$ with $f(x_0) = y_0$ and $g : V \to G$, we have $g \circ f : U \to G$. Assume f is differentiable at x_0 and g differentiable at y_0, and set $A = (\mathrm{d}f)(x_0)$, $B = (\mathrm{d}g)(y_0)$. We then have, for ξ in a small neighborhood of 0 in E, that $(\Delta_{x_0} f)(\xi) = f(x_0+\xi) - f(x_0)$ is an infinitesimal from F to G, and that (recall (1.2.10))

$$\begin{aligned}\big(\Delta_{x_0}(g \circ f)\big)(\xi) &= g\big(f(x_0 + \xi)\big) - g\big(f(x_0)\big)\\ &= g\big(f(x_0) + (\Delta_{x_0} f)(\xi)\big) - f(x_0)\\ &= g\big(y_0 + (\Delta_{x_0} f)(\xi)\big) - g(y_0).\end{aligned} \tag{1.2.16}$$

To estimate the right side of (1.2.16) using $B = (\mathrm{d}g)(y_0)$, we consider the infinitesimal $\varepsilon_1 \in \mathfrak{o}(F, G)$ defined by

$$\varepsilon_1(\eta) = g(y_0 + \eta) - g(\eta) - B\eta. \tag{1.2.17}$$

Then (1.2.16) and (1.2.17) together yields

$$\big(\Delta_{x_0}(g \circ f)\big)(\xi) - B(\Delta_{x_0} f)(\xi) = \varepsilon_1\big((\Delta_{x_0} f)(\xi)\big). \tag{1.2.18}$$

Note that $\Delta_{x_0} f \in \mathfrak{O}(E, F)$ by (1.2.15), hence $\varepsilon_1 \circ (\Delta_{x_0} f) \in \mathfrak{o}(E, G)$ by Proposition 1.2.5, thus (1.2.18) now says that

$$\Delta_{x_0}(g \circ f) - B \circ (\Delta_{x_0} f) \in \mathfrak{o}(E, G). \tag{1.2.19}$$

On the other hand, by $A = (\mathrm{d}f)(x_0)$, we have

$$\Delta_{x_0} f - A \in \mathfrak{o}(E, F). \tag{1.2.20}$$

It follows from $B \in \mathcal{L}(F, G) \subseteq \mathfrak{O}(F, G)$, (1.2.20) and Proposition 1.2.5 that

$$B \circ (\Delta_{x_0} f - A) \in \mathfrak{o}(E, G). \tag{1.2.21}$$

Now taking the sum of (1.2.19) and (1.2.21), we get

$$\Delta_{x_0}(g \circ f) - BA \in \mathfrak{o}(E, G), \tag{1.2.22}$$

which means precisely that $g \circ f$ is differentiable at x_0 with

$$\mathrm{d}(g \circ f)(x_0) = (\mathrm{d}_{f(x_0)} g) \circ (\mathrm{d}_{x_0} f). \tag{1.2.23}$$

To sum up, we have established the following result, known as **the chain rule** for calculating the derivative of composite maps.

Proposition 1.2.13. — *Let E, F, G be normed spaces, U (resp. V) is an open set in E (resp. F), and $x_0 \in U$ (resp. $y_0 \in V$). Given maps $f : U \to V \subseteq F$ with $f(x_0) = y_0$ and $g : V \to G$, we have $g \circ f : U \to G$. Assume f is differentiable at x_0 and g differentiable at y_0. Then $g \circ f$ is differentiable at x_0 with $\mathrm{d}(f \circ g)(x_0)$ given by (1.2.23).* □

1.2.4. The one-dimensional case and the directional derivative. — Let E, F be normed spaces, U an open set of E, and $f : U \to F$ a map that is differentiable at some point $x_0 \in U$.

In the special case where $\dim F = 1$, we can identify F with $\mathbb{R}$ by the isometric isomorphism $\varphi : \mathbb{R} \to F$, $t \mapsto ty_0$, where $y_0 \in F$ is such that $\|y_0\| = 1$. Via this identification, we have

$$(\mathrm{d}f)(x_0) \in \mathcal{L}(E, F) = \mathcal{L}(E, \mathbb{R}) = E'. \tag{1.2.24}$$

Recall that $E' = \mathcal{L}(E, \mathbb{R})$ is called *the topological dual* of E. If f is differentiable at every point in U, then we obtain a map $\mathrm{d}f : U \to E'$, $x \mapsto (\mathrm{d}f)(x)$. This $\mathrm{d}f$ is an example of the so-called 1-form, which will be treated later when we study differential forms.

In the special case where $\dim E = 1$ (F being arbitrary), we can again identify E with $\mathbb{R}$ as in the previous paragraph. In this case, $(\mathrm{d}f)(x_0) \in \mathcal{L}(\mathbb{R}, F)$. But we have the canonical isometric isomorphism

$$\begin{aligned} \mathcal{L}(\mathbb{R}, F) &\to F \\ \varphi &\mapsto \varphi(1) \end{aligned} \tag{1.2.25}$$

whose inverse is

$$\begin{aligned} F &\to \mathcal{L}(\mathbb{R}, F) \\ \xi &\mapsto \{t \mapsto t\xi\}. \end{aligned} \tag{1.2.26}$$

Proposition 1.2.14. — *Let U be an open set of $\mathbb{R}$ and F a normed space. Let $f : U \to F$ be a map, $A \in \mathcal{L}(\mathbb{R}, F)$ and $x_0 \in U$. Set $\xi = A(1) \in F$. Then the following are equivalent.*

(1) *The map f is differentiable at x_0 with $(\mathrm{d}f)(x_0) = A$.*
(2) *The limit*

$$\lim_{t \to 0} \frac{f(x_0 + t) - f(x_0)}{t} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{1.2.27}$$

exists and this limit is exactly $\xi \in F$.

Proof. — It is clear that $A(t) = A(t \cdot 1) = tA(1) = t\xi$. Now this equivalence can be easily checked using

$$\left(\frac{f(x_0 + t) - f(x_0)}{t}\right) - \xi = \frac{(\Delta_{x_0} f)(t) - A(t)}{t}, \tag{1.2.28}$$

and the definitions of limit and $(\mathrm{d}f)(x_0)$. □

Notation 1.2.15. — *Because of Proposition 1.2.14 (2), when $E = \mathbb{R}$, we often write $[(\mathrm{d}f)(x_0)](1)$ simply as $f'(x_0)$.*

Now we get back to the general setting.

Proposition 1.2.16. — *Let E, F be normed spaces, $U \subseteq E$ an open set, $x_0 \in U$, and $f : U \to F$ a map that is differentiable at x_0. Let $\varepsilon > 0$ be so small that $\alpha_\xi(t) = x_0 + t\xi$ defines a map $\alpha_\xi :]-\varepsilon, \varepsilon[\to U$ with $\alpha_\xi(0) = x_0$. Then the limit*

$$\lim_{t \to 0} \frac{f(x_0 + t\xi) - f(x_0)}{t} \tag{1.2.29}$$

exists and is precisely $[(\mathrm{d}f)(x_0)](\xi)$.

Proof. — We have, for all $t \in]-\varepsilon, \varepsilon[\setminus \{0\}$, that

$$\frac{\alpha_\xi(t) - \alpha_\xi(0)}{t} = \xi. \tag{1.2.30}$$

Hence $\alpha'_\xi(0) = \xi$. By Proposition 1.2.16, α_ξ is differentiable at 0 with $(\mathrm{d}\alpha_\xi)(0) = R_\xi$, where $R_\xi : \mathbb{R} \to E$ is the map sending $t \in \mathbb{R}$ to $t\xi$. Now consider the composite map $g := f \circ a_\xi$ from $]-\varepsilon, \varepsilon[$ to F and note that $\alpha_\xi(0) = x_0$. It follows from the chain rule (Proposition 1.2.13) that g is differentiable at 0 with

$$(\mathrm{d}g)(0) = (\mathrm{d}f)(x_0) \circ R_\xi. \tag{1.2.31}$$

Thus by Proposition 1.2.16 again, the limit

$$\lim_{t \to 0} \frac{g(t) - g(0)}{t} = \lim_{t \to 0} \frac{f(x_0 + t\xi) - f(x_0)}{t} \tag{1.2.32}$$

exists and equals $[(\mathrm{d}g)(0)](1) = [(\mathrm{d}f)(x_0)](\xi)$. □

Definition 1.2.17. — Let E, F be normed spaces, $U \subseteq E$ an open set, $x_0 \in U$, and $f : U \to F$ a map. Pick any $\xi \in E$. We say that f **is differentiable along the direction** $\xi \in E$ **at** x_0, if the limit in (1.2.29) exists, and we call this limit **the directional derivative of** f **along the direction** ξ **at** x_0.

Using the notion of directional derivatives, Proposition 1.2.16 can be reformulated as the following corollary.

Corollary 1.2.18. — *Let E, F be normed spaces, $U \subseteq E$ an open set, $x_0 \in U$, and $f : U \to F$ a map. If f is differentiable at x_0, then the directional derivative of f at x_0 exists along any direction $\xi \in E$, and equals $[(Df)(x_0)](\xi)$.*

Remark 1.2.19. — We see in the exercise that the converse of Corollary 1.2.18 is *false*: there exists a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that the directional derivative along any direction exists at 0 but the function is not even continuous at 0.

1.3. Some Examples of Derivatives

Example 1.3.1 **(Derivative of continuous affine maps).** — Let E and F be normed spaces. Consider a **continuous affine map** α from E to F, i.e. α is of the form $\alpha(x) = Ax + \eta$ for some $A \in \mathcal{L}(E, F)$ and $\eta \in F$. Then for any $x_0, \xi \in E$, we have

$$(\Delta_{x_0}\alpha)(\xi) = \alpha(x_0 + \xi) - \alpha(x_0) = A\xi, \tag{1.3.1}$$

which means $\Delta_{x_0}\alpha - A$ is a zero map that is trivially in $\mathfrak{o}(E, F)$. Hence α is differentiable at any $x_0 \in E$, and $(\mathrm{d}f)(x_0) = A$.

We now establish a lemma as a preparation for calculating the derivative of a continuous bilinear map.

Lemma 1.3.2. — *Let E_1, E_2, F and G be normed spaces, and $\beta \in \mathcal{B}(E_1 \times E_2, F)$ (i.e. β is continuous and bilinear). Suppose $f_i \in \mathfrak{O}(G, E_i)$ for $i = 1, 2$, then the map*

$$\beta \circ (f_1, f_2) : z \in \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2) \mapsto \beta\big(f_1(z), f_2(z)\big)$$

is an infinitesimal of higher order, i.e.

$$\beta \circ (f_1, f_2) \in \mathfrak{o}(G, F). \tag{1.3.2}$$

Proof. — For $i = 1, 2$, let $c_i \geq 0$ be a constant such that $\|f_i(x_i)\| \leq c_i\|x_i\|$ for any $x_i \in \mathrm{dom}(f_i)$. Then using the norm $\|(x_1, x_2)\| = \max(\|x_1\|, \|x_2\|)$ on $E_1 \times E_2$ (see Corollary 1.1.8 and Remark 1.2.2), we see, for any $z \in \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$, that

$$\big\|\beta\big(f_1(z), f_2(z)\big)\big\| \leq \|\beta\| \cdot \|f_1(z)\| \cdot \|f_2(z)\| \leq c_1c_2\|\beta\|\|z\|^2. \tag{1.3.3}$$

Now (1.3.2) follows from (1.3.3). □

Example 1.3.3 **(Derivative of continuous bilinear maps).** — Let E_1, E_2 and F be normed spaces and $\beta \in \mathcal{B}(E_1 \times E_2, F)$. For any $a = (a_1, a_2) \in E_1 \times E_2$, we have

$$(\Delta_a\beta)(\xi_1, \xi_2) = \beta(a_1 + \xi_1, a_2 + \xi_2) - \beta(a_1, a_2) = \beta(a_1, \xi_2) + \beta(\xi_1, a_2) + \beta(\xi_1, \xi_2). \tag{1.3.4}$$

For $i = 1, 2$, let $\pi_i : E_1 \times E_2 \to E_i$ be the i-th projection, and $u_{i,a} : E_i \to E_1 \times E_2$ the insertion such that $u_{i,a}(x_i)$ has x_i in the i-th component and a_j in the j-th component for $j \neq i$. Since β is continuous and bilinear, the composite $\beta_{i,a} := \beta \circ u_{i,a} \in \mathcal{L}(E_i, F)$. Then (1.3.4) implies that

$$\left\{(\Delta_a\beta) - \sum_{i=1}^{2} \beta_{i,a} \circ \pi_i\right\}(\xi_1, \xi_2) = \big(\beta \circ (\pi_1, \pi_2)\big)(\xi_1, \xi_2). \tag{1.3.5}$$

By Lemma 1.3.2, we see that $\beta \circ (\pi_1, \pi_2) \in \mathfrak{o}(E_1 \times E_2, F)$. Hence by (1.3.5), we see that β is differentiable at $a = (a_1, a_2)$ with

$$(\mathrm{d}\beta)(a) = \sum_{i=1}^{2} \beta_{i,a} \circ \pi_i \in \mathcal{L}(E_1 \times E_2, F). \tag{1.3.6}$$

More specifically, for all $(\xi_1, \xi_2) \in E_1 \times E_2$, we have

$$[(\mathrm{d}\beta)(a)](\xi_1, \xi_2) = \beta(a_1, \xi_2) + \beta(\xi_1, a_2). \tag{1.3.7}$$

Remark 1.3.4. — (Exercise) Lemma 1.3.2 and Example 1.3.3 can be generalized to continuous multilinear maps.

***Example 1.3.5* (Derivative of the inverse map).** — Let E, F be *Banach* spaces, and $A \in \mathcal{L}(E, F)$ an isomorphism in the sense that there exists $B \in \mathcal{L}(F, E)$ with $AB = \mathrm{Id}_E$ and $BA = \mathrm{Id}_F$, and we write $B = A^{-1}$. For any $X \in \mathcal{L}(E, F)$ with $\|X\| < \left\|A^{-1}\right\|^{-1}$, we have

$$A + X = A(\mathrm{Id}_E + A^{-1}X). \tag{1.3.8}$$

Since

$$\left\|A^{-1}X\right\| \leq \left\|A^{-1}\right\| \cdot \|X\| < 1, \tag{1.3.9}$$

the series

$$\sum_{k\geq 0}(-A^{-1}X)^k = \mathrm{Id}_E + \sum_{k\geq 1}(-A^{-1}X)^k$$

converges in the Banach space[(3)] $\mathcal{L}(E)$, and is easily checked to be the inverse of $\mathrm{Id}_E + A^{-1}X$, i.e.

$$(\mathrm{Id}_E + A^{-1}X)^{-1} = \sum_{k\geq 0}(-A^{-1}X)^k \in \mathcal{L}(E). \tag{1.3.10}$$

It follows from (1.3.8) and (1.3.10) that $A + X \in \mathcal{L}(E, F)$ is also an isomorphism with

$$(A + X)^{-1} = (\mathrm{Id}_E + A^{-1}X)^{-1}A^{-1} = \sum_{k\geq 0}(-A^{-1}X)^k A^{-1} \in \mathcal{L}(F, E). \tag{1.3.11}$$

Denote by $\mathrm{Iso}(E, F)$ the subset of $\mathcal{L}(E, F)$ consisting of isomorphisms from E onto F. For any $A \in \mathrm{Iso}(E, F)$, by the above, we see that the open ball centered at A with radius $\left\|A^{-1}\right\|^{-1} > 0$ lies entirely inside $\mathrm{Iso}(E, F)$. This means that $\mathrm{Iso}(E, F)$ is an open set in $\mathcal{L}(E, F)$.

Now we consider the inverse map

$$\begin{aligned} \nu : \mathrm{Iso}(E, F) &\to \mathcal{L}(F, E) \\ A &\mapsto A^{-1}. \end{aligned} \tag{1.3.12}$$

For any fixed $A \in \mathrm{Iso}(E, F)$, it follows from (1.3.11) that for any $X \in \mathcal{L}(E, F)$ with $\|X\| < \left\|A^{-1}\right\|^{-1}$, we have

$$(\Delta_A \nu)(X) + A^{-1}XA^{-1} = \sum_{k\geq 2}(-A^{-1}X)^k A^{-1}. \tag{1.3.13}$$

Note the inequality (1.3.9), we have

$$\begin{aligned} \left\|\sum_{k\geq 2}(-A^{-1}X)^k A^{-1}\right\| &\leq \sum_{k\geq 2}\left\|(A^{-1}X)^k A^{-1}\right\| \\ &\leq \left\|A^{-1}\right\| \sum_{k\geq 2}\left\|A^{-1}X\right\|^k \\ &= \left\|A^{-1}\right\| \cdot \left\|A^{-1}X\right\|^2 \sum_{k\geq 0}\frac{1}{1 - \|A^{-1}X\|} \\ &\leq \|X\|^2 \cdot \frac{\left\|A^{-1}\right\|^3}{1 - \|A^{-1}\| \cdot \|X\|}. \end{aligned} \tag{1.3.14}$$

It follows from (1.3.14) and (1.3.13) that

$$\Delta_A \nu - \left(-A^{-1}(\cdot)A^{-1}\right) \in \mathfrak{o}\big(\mathcal{L}(E, F), \mathcal{L}(F, E)\big). \tag{1.3.15}$$

Hence ν is differentiable at A with

$$(\mathrm{d}\nu)(A) = -A^{-1}(\cdot)A^{-1} \in \mathcal{L}\Big(\mathcal{L}(E, F), \mathcal{L}(F, E)\Big). \tag{1.3.16}$$

[(3)] This $\mathcal{L}(E)$ is actually a typical example of a Banach algebra.

1.4. The Mean Value Theorem and Some of Its Applications

Let $[a,b]$ be a compact interval, F a normed space and $f:[a,b]\to F$ a map. For any $t\in[a,b[$, we define the **right derivative** of f at t to be

$$f'_r(t) := \lim_{s\to 0^+}\frac{f(t+s)-f(t)}{s}, \tag{1.4.1}$$

provided this limit exists. Similarly, for $t\in\,]a,b]$, we define the **left derivative** of f at t to be

$$f'_l(t) := \lim_{s\to 0^-}\frac{f(t+s)-f(s)}{s}, \tag{1.4.2}$$

provided this limit exists. It is clear that for $t\in\,]a,b[$, the existence of $f'(t)$ implies the existence of $f'_l(t)$ and $f'_r(t)$ with $f'(t)=f'_l(t)=f'_r(t)$.

The following theorem is also known as the finite-increment theorem.

***Theorem 1.4.1* (The mean value theorem).** — *Let $[a,b]$ be a compact interval, $D\subseteq[a,b]$ a countable set, F a normed space, $f:[a,b]\to F$ a* continuous *map and $\varphi:[a,b]\to\mathbb{R}$ a* continuous *real-valued function. Suppose for each $t\in\,]a,b[\,\setminus D$, the left (resp. right) derivatives of f and φ both exists and $\|f'_l(t)\|\le\varphi'_l(t)$ (resp. $\|f'_r(t)\|\le\varphi'_r(t)$), then*

$$\|f(b)-f(a)\|\le\varphi(b)-\varphi(a). \tag{1.4.3}$$

Proof. — We treat the case of the right derivatives first. Let $(\rho_n)_{n\in\mathbb{N}}$ be an enumeration of D. For any $\varepsilon>0$, we claim that

$$\forall t\in[a,b],\qquad \|f(t)-f(a)\|\le\varphi(t)-\varphi(a)+\varepsilon(t-a)+\varepsilon\sum_{\rho_n<t}2^{-n}. \tag{1.4.4}$$

Set

$$A=\big\{s\in[a,b]\;\big|\;\forall t\in[a,s[,\ \text{ the inequality in (1.4.4) holds}\big\}.$$

One checks immediately that $a\in A$ so that $A\neq\emptyset$ and that $s_0:=\sup A\in A$. If $s_0=b$, then $b\in A$ and (1.4.4) is proved. We now show that this must be the case.

Suppose otherwise, then $s_0<b$. By continuity, we see that (1.4.4) holds not only for all $t\in[a,s_0[$ but also for $t=s_0$, i.e.

$$\|f(s_0)-f(a)\|\le\varphi(s_0)-\varphi(a)+\varepsilon(s_0-a)+\varepsilon\sum_{\rho_n<s_0}2^{-n}. \tag{1.4.5}$$

Now either $s_0\in D$ or $s_0\in\,]a,b[\,\setminus D$.

If $s_0\in D$, then $s_0=\rho_{n_0}$. By continuity of f and φ at s_0, there exists $s_1\in\,]s_0,b[$ such that for any $t\in[s_0,s_1[$, we have

$$\|f(t)-f(s_0)\|\le\frac{\varepsilon}{2^{-(n_0+1)}}, \tag{1.4.6}$$

and

$$\varphi(t)-\varphi(s_0)\le\frac{\varepsilon}{2^{-(n_0+1)}}. \tag{1.4.7}$$

Hence

$$\begin{aligned}\forall t\in[s_0,s_1[,\qquad &\|f(t)-f(a)\|\le\|f(s_0)-f(a)\|+\|f(t)-f(s_0)\|\\ &\le\varphi(s_0)-\varphi(a)+\varepsilon(s_0-a)+\varepsilon\sum_{\rho_n<s_0}2^{-n}+\frac{\varepsilon}{2^{-(n_0+1)}}\\ &=\varphi(s_0)-\varphi(a)+\Big(\varphi(t)-\varphi(s_0)-\frac{\varepsilon}{2^{-(n_0+1)}}\Big)\\ &\qquad+\varepsilon(s_0-a)+\varepsilon\left(\frac{1}{2^{-n_0}}+\sum_{\rho_n<s_0}2^{-n}\right)\\ &\le\varphi(t)-\varphi(a)+\varepsilon(t-a)+\varepsilon\sum_{\rho_n<t}2^{-n}.\end{aligned} \tag{1.4.8}$$

Now (1.4.5) and (1.4.8), together with $s_0\in A$, imply that $s_1\in A$, contradicting $s_0=\sup A$.

If $s_0 \in \left]a, b\right[\setminus D$, then both $f'_r(s_0)$ and $\varphi'_r(s_0)$ exist. Hence there exists $s_2 \in \left]s_0, b\right[$ such that for all $t \in \left]s_0, s_2\right[$

$$\left\| \frac{f(t) - f(s_0)}{t - s_0} - f'_r(s_0) \right\| \leq \frac{\varepsilon}{2}, \tag{1.4.9}$$

and

$$\frac{\varphi(t) - \varphi(s_0)}{t - s_0} - \varphi'_r(s_0) \geq -\frac{\varepsilon}{2}. \tag{1.4.10}$$

Thus for all $t \in \left]s_0, s_2\right[$, we have, by (1.4.10), that

$$\|f(t) - f(s_0)\| \leq \|f'_r(s_0)\|(t - s_0) + \frac{\varepsilon}{2}(t - s_0) \leq \varphi'_r(s_0)(t - s_0) + \frac{\varepsilon}{2}(t - s_0), \tag{1.4.11}$$

and by (1.4.10), that

$$\varphi'_r(s_0)(t - s_0) \leq \varphi(t) - \varphi(s_0) + \frac{\varepsilon}{2}(t - s_0). \tag{1.4.12}$$

It follows from (1.4.10) and (1.4.11) that

$$\forall t \in \left[s_0, s_2\right[, \qquad \|f(t) - f(s_0)\| \leq \varphi(t) - \varphi(s_0) + \varepsilon(t - s_0), \tag{1.4.13}$$

which, together with $s_0 \in A$, implies that $s_2 \in A$, again contradicting $s_0 = \sup A$.

Therefore, we have shown that we shall always have a contraction if $s_0 \neq b$, which forces $b = s_0 \in A$. Now (1.4.5) together with $b = s_0 \in A$ imply (1.4.4), as claimed. Since $\sum_{n \geq 0} 2^{-n} = 2$, it follows from (1.4.4) that

$$\|f(b) - f(a)\| \leq \varphi(b) - \varphi(a) + \varepsilon(b - a + 2). \tag{1.4.14}$$

Since $\varepsilon > 0$ can taken arbitrarily in (1.4.14), we obtain (1.4.3), which proves the theorem in the case of existence of right derivatives.

For the case of left derivatives, by replacing f with $g(t) := f(a + b - t)$; φ with $\psi(t) := -\varphi(a + b - t)$; and D with $D' = a + b - D$, one checks easily that both g and ψ are right differentiable at any point in $t \in \left]a, b\right[\setminus D'$ with $\|g'_r(t)\| \leq \psi'_r(t)$. By the above case for right derivatives, we still have

$$\|f(a) - f(b)\| = \|g(b) - g(a)\| \leq \psi(b) - \psi(a) = \varphi(b) - \varphi(a), \tag{1.4.15}$$

and the proof is complete. □

Definition 1.4.2. — Let E, F be normed spaces, $U \subseteq E$ a *connected* open set, $f : U \to E$ a map. We say that f is

- **differentiable** if f is differentiable at every $x_0 \in U$;
- **continuously differentiable**, or **of class** C^1, if f is differentiable, and the map

$$\begin{aligned} \mathrm{d}f : U &\to \mathcal{L}(E, F) \\ x &\mapsto \mathrm{d}f(x_0) \end{aligned} \tag{1.4.16}$$

is continuous.

Corollary 1.4.3. — *Let E, F be normed spaces, $U \subseteq E$ an open set, $f : U \to E$ a map, and $a, b \in U$ such that the segment $\{a + t(b - a) \mid t \in [0, 1]\}$ lie in U and the points on this segment at which f is not differentiable is at most countable. Then*

$$\|f(b) - f(a)\| \leq \sup_{0 < t < 1} \big\|\mathrm{d}f\big(a + t(b - a)\big)\big\| \cdot \|b - a\|. \tag{1.4.17}$$

Proof. — Apply Theorem 1.4.1 to the composite map $g : [0, 1] \to E$, $t \mapsto f\big(a + t(b - a)\big)$, by noting that $g'(t) = \big[\mathrm{d}f\big(a + t(b - a)\big)\big](b - a)$ whenever f is differentiable at $a + t(b - a)$. □

Before stating the following corollary, we need the following result on topology[4].

Lemma 1.4.4. — *Let E be a normed space and $U \subseteq E$ open. Then each connected component of U is still open and path-connected.*

[4] We will see in the topology course that the result is still true for all *locally path-connected* spaces.

Proof. — Take any $x_0 \in U$, it sufficed to show that the connected component containing x_0 is open and path-connected. Set

$$U_0 = \Big\{x \in U \;\Big|\; \text{There exists a continuous path from } x_0 \text{ to } x\Big\}. \tag{1.4.18}$$

For each $x \in U_0 \subseteq U$, since U is open, there exists an $r > 0$ with the open ball

$$B(x,r) := \{y \in E \mid \|x-y\| < r\}$$

lying entirely inside U. For each $y \in B(x,r)$, the line segment map $\gamma_{x,y} : [0,1] \to E$, $t \mapsto x + t(y-x)$, which is clearly continuous[5], has its image inside $B(x,r)$, for whenever $t \in [0,1]$, we have

$$\Big\|\Big(x + t(y-x)\Big) - x\Big\| = \|t(y-x)\| = t\|y-x\| < tr \leq r. \tag{1.4.19}$$

By $x \in U_0$ and the definition of U_0, there exists a continuous map $\gamma_x : [0,1] \to U$ with $\gamma_x(0) = x_0$ and $\gamma_x(1) = x$. Now we define $\gamma : [0,1] \to U$ by

$$\gamma(t) = \begin{cases} \gamma_x(2t), & \text{if } t \in [0, \frac{1}{2}]; \\ \gamma_{x_0,x}(2t-1), & \text{if } t \in]\frac{1}{2}, 1]. \end{cases}$$

One checks (check it!) that γ is continuous with $\gamma(0) = x_0$ and $\gamma(1) = y$. This means that $B(x,r) \subseteq U_0$ and U_0 is open. On the other hand, if $z \in U \setminus U_0$, but $B(z,s) \cap U_0 \neq \emptyset$ for some $s > 0$. Then we can take some $w \in B(z,s) \cap U_0$ and connect w with x_0 by some continuous path γ_w. Similarly, we can concatenate γ_w with the line segment map $\gamma_{w,z}$ to obtain a continuous path from x_0 to z, contradicting to $z \notin U_0$. Thus $B(z,s) \subseteq U \setminus U_0$ for some $s > 0$ and $U \setminus U_0$ is also open. Clearly, U_0 is path-connected, hence connected. Let U_0' be the unique connected component containing U_0, we then see that U_0 and $U' \setminus U_0$ is a partition of U' into its open subsets. By the connectedness of U' and $U_0 \neq \emptyset$ (since $x_0 \in U_0$, why?), we must have $U' \setminus U_0 = \emptyset$ and $U' = U_0$, which is open and path-connected. □

Corollary 1.4.5. — *Let E, F be normed spaces, $U \subseteq E$ a non-empty open set, $f : U \to E$ a differentiable map such that $\mathrm{d}f(x_0) = 0$ for each $x_0 \in U$, then f is locally constant, i.e. f is constant on each connected component of U.*

Proof. — Note that each connected component of U is open (Lemma 1.4.4), we may therefore assume that U itself is open. Take $x_0 \in U$, and consider the subset $U_0' := f^{-1}(f(x_0))$ of U. Since f is continuous and the singleton $\{f(x_0)\}$ is closed, U_0' is closed in U. On the other hand, for any $x \in U_0'$, since U is open, $B(x,r) \subseteq U$ for some $r > 0$ and Corollary 1.4.3 implies that $f(y) = f(x) = f(x_0)$ for any $y \in B(x,r)$, hence $B(x,r) \subseteq U_0'$ and U_0' is also open. Since U is connected and $x_0 \in U_0'$, we must have $U_0' = U$. □

As another application of the mean value theorem, we now state an important result relating the uniform convergence and derivatives for a sequence of maps.

Theorem 1.4.6. — *Let E be a normed spaces, F a Banach space, $U \subseteq E$ a connected open set, and (f_n) a sequence of differentiable maps from U to E. Suppose the following condition hold.*

1. *There exists an $x_0 \in U$ such that $(f_n(x_0))$ converges.*
2. *The sequence $(\mathrm{d}f_n)$ of maps from U to $\mathcal{L}(E,F)$ converges locally uniformly.*

Then (f_n) converges locally uniformly to a map $f : U \to F$, and f is differentiable with $(\mathrm{d}f_n)$ converges locally uniformly to $\mathrm{d}f$.

Proof. — Let A be the subset of U consisting of $a \in U$ such that $(f_n(a))$ converges. Clearly $x_0 \in A$ so that $A \neq \emptyset$. Now for any $a \in U$, there exists $r > 0$, such that $(\mathrm{d}f_n)$ converges uniformly on the open

[5] this follows from the continuity of the norm function.

ball $B(a,r) \subseteq U$. By Corollary 1.4.3 apply to $f_n - f_m$, we see that

$$\tag{1.4.20} \begin{aligned} \forall x \in B(a,r), \qquad & \big\| f_n(x) - f_n(a) - \big(f_m(x) - f_m(a)\big) \big\| \\ & \le \sup_{t \in]0,1[} \big\| (\mathrm{d}f_n - \mathrm{d}f_m)\big(a + t(x-a)\big) \big\| \cdot \|x - a\| \\ & \le \left(\sup_{y \in B(a,r)} \|\mathrm{d}f_n(y) - \mathrm{d}f_m(y)\| \right) \|x-a\| \\ & \le r \cdot \sup_{y \in B(a,r)} \|\mathrm{d}f_n(y) - \mathrm{d}f_m(y)\|. \end{aligned}$$

Note that since $(\mathrm{d}f_n)$ converges uniformly on $B(x,a)$, it follows from (1.4.20) that $(f_n(x))$ is a Cauchy sequence if and only if $(f_n(a))$ is a Cauchy sequence. Since $x \in B(a,r)$ is chosen arbitrarily, it follows that either $B(a,r) \subseteq A$ or $B(a,r) \cap A = \emptyset$. Hence A is both open and closed in U. Since U is connected and $A \neq \emptyset$ (recall $x_0 \in A$), we must have $U = A$, i.e. every point $(f_n(a))$ converges for any $a \in U$. Use (1.4.20) again, we see that (f_n) converges uniformly on $B(a,r)$. Hence (f_n) converges locally uniformly to some map $f : U \to F$.

It remains to show that f is differentiable and that $(\mathrm{d}f_n)$ converges locally uniformly to $\mathrm{d}f$. Since (df_n) already converges locally uniformly, it suffices to show that $\mathrm{d}f(a) = \lim_{n\to+\infty} \mathrm{d}f_n(a)$ for any $a \in U$. Let $r > 0$ be as above and take an arbitrary $\varepsilon > 0$. Since $(\mathrm{d}f_n)$ converges uniformly on $B(x,a)$, there exists $m \in \mathbb{N}$, such that

$$\tag{1.4.21} \forall n \ge m, \forall x \in B(a,r), \qquad \|\mathrm{d}f_n(x) - \mathrm{d}f_m(x)\| \le \varepsilon.$$

Taking $n \to +\infty$ in (1.4.20), by (1.4.21), we see that

$$\tag{1.4.22} \forall x \in B(a,r), \qquad \big\| f(x) - f(a) - \big(f_m(x) - f_m(a)\big) \big\| \le \varepsilon \|x - a\|.$$

On the other hand, since f_m is differentiable at a, there exists $r' \in]0, r[$, such that

$$\tag{1.4.23} \forall x \in B(a,r'), \qquad \|f_m(x) - f_m(a) - \mathrm{d}f_m(a)\cdot(x-a)\| \le \varepsilon\|x-a\|.$$

Let $g(a) = \lim_{n\to+\infty} \mathrm{d}f_n(a) \in \mathcal{L}(E,F)$. Then taking $n \to +\infty$ and $x = a$ in (1.4.21), we see that

$$\tag{1.4.24} \|g(a) - \mathrm{d}f_m(a)\| \le \varepsilon.$$

Apply $g(a) - \mathrm{d}f_m(a) \in \mathcal{L}(E,F)$ to $x - a \in E$, we see that

$$\tag{1.4.25} \forall x \in E, \qquad \|g(a)\cdot(x-a) - \mathrm{d}f_m(a)\cdot(x-a)\| \le \varepsilon\|x-a\|.$$

Combining (1.4.22), (1.4.23) and (1.4.25), we have

$$\tag{1.4.26} \forall x \in B(a,r'), \qquad \|f(x) - f(a) - g(a)\cdot(x-a)\| \le 3\varepsilon\|x-a\|.$$

Since $\varepsilon > 0$ is arbitrary, (1.4.26) implies that $\Delta_a f - g \in \mathfrak{o}(E,F)$, i.e. f is indeed differentiable at a with $\mathrm{d}f(a) = g(a) = \lim_{n\to+\infty} \mathrm{d}f_n(a)$, as desired. □

1.5. The Second and Higher Derivatives

Definition 1.5.1. — Let E, F be normed spaces, $U \subseteq E$ an open set, $x_0 \in U$ and $f : U \to E$ a map (see Definition 1.4.2). If the map f is differentiable on a neighborhood V of x_0 inside U, and the derived map $\mathrm{d}f : V \to \mathcal{L}(E,F)$ is differentiable at x_0, then we say **the second derivative**, or **the derivative of second order** of f at x_0 exists, or that f is **twice differentiable** at x_0, and we write $\mathrm{d}^2 f(x_0)$ for $\mathrm{d}(\mathrm{d}f)(x_0)$.

Remark 1.5.2. — Note that $\mathrm{d}^2 f(x_0) \in \mathcal{L}\big(E, \mathcal{L}(E,F)\big)$, so by currying (Proposition 1.1.12), we may and often do regard $\mathrm{d}^2 f(x_0)$ as the continuous bilinear map from $E \times E$ to F that sends (ξ_1, ξ_2) to $(\mathrm{d}^2 f)(x_0)(\xi_1)\cdot\xi_2 = \{[(\mathrm{d}^2 f)(x_0)](\xi_1)\}(\xi_2)$.

Proposition 1.5.3. — *Let E, F be normed spaces, $U \subseteq E$ an open set and $x_0 \in U$. If a map $f : U \to E$ a map that is twice differentiable at x_0. Then as a bilinear map, $\mathrm{d}^2 f(x_0)$ is symmetric, i.e. for all $\xi_1, \xi_2 \in E$, we have*

$$\mathrm{d}^2 f(x_0)(\xi_1) \cdot \xi_2 = \mathrm{d}^2 f(x_0)(\xi_2) \cdot \xi_1. \tag{1.5.1}$$

Proof. — Since U is open, there exists $r > 0$ such that $B(x_0, 2r) \subseteq U$. For any $\xi, \eta \in B(x_0, r)$, consider the expression

$$h(\xi, \eta) = f(x_0 + \xi + \eta) - f(x_0 + \xi) - f(x_0 + \eta) + f(x_0), \tag{1.5.2}$$

which is well-defined and symmetric, i.e.

$$h(\xi, \eta) = h(\eta, \xi). \tag{1.5.3}$$

For $s \in [0, 1]$, set

$$\alpha_{\xi,\eta}(s) = h(s\xi, \eta) - \mathrm{d}^2 f(x_0)(\eta)(s\xi). \tag{1.5.4}$$

Then by the chain rule, for any $s \in]0, 1[$, we have

$$\alpha'_{\xi,\eta}(s) = \Big(\mathrm{d}f(x_0 + s\xi + \eta) - \mathrm{d}f(x_0 + s\xi) - \mathrm{d}^2 f(x_0)(\eta)\Big) \cdot \xi. \tag{1.5.5}$$

Hence,

$$\begin{aligned} \forall \xi, \eta \in B(0, r), \qquad & \big\|h(\xi, \eta) - \mathrm{d}^2 f(x_0)(\eta) \cdot \xi\big\| = \|\alpha(1) - \alpha(0)\| \\ & \leq \sup_{s \in]0,1[} \Big\|\Big(\mathrm{d}f(x_0 + s\xi + \eta) - \mathrm{d}f(x_0 + s\xi) - \mathrm{d}^2 f(x_0)(\eta)\Big) \cdot \xi\Big\| \\ & \leq \|\xi\| \sup_{s \in]0,1[} \Big\|\Big(\mathrm{d}f(x_0 + s\xi + \eta) - \mathrm{d}f(x_0 + s\xi) - \mathrm{d}^2 f(x_0)(\eta)\Big)\Big\|. \end{aligned} \tag{1.5.6}$$

Now take any $\varepsilon > 0$. Since f is twice differentiable at x_0, there exists $r_1 \in]0, r[$, such that f is differentiable in $B(x_0, 2r_1)$, and that

$$\forall \zeta \in B(0, 2r_1), \qquad \big\|\mathrm{d}f(x_0 + \zeta) - \mathrm{d}f(x_0) - \mathrm{d}^2 f(x_0)(\zeta)\big\| \leq \varepsilon\|\zeta\|. \tag{1.5.7}$$

If $\xi, \eta \in B(x_0, r_1)$, then setting $\zeta = s\xi + \eta$ and $\zeta = s\xi$ in (1.5.7), one obtains respectively, that

$$\forall \xi, \eta \in B(0, r_1),\ s \in]0, 1[, \qquad \big\|\mathrm{d}f(x_0 + s\xi + \eta) - \mathrm{d}f(x_0) - \mathrm{d}^2 f(x_0)(s\xi + \eta)\big\| \leq \varepsilon\|s\xi + \eta\|, \tag{1.5.8}$$

and

$$\forall \xi, \eta \in B(0, r_1),\ s \in]0, 1[, \qquad \big\|\mathrm{d}f(x_0 + s\xi) - \mathrm{d}f(x_0) - \mathrm{d}^2 f(x_0)(s\xi)\big\| \leq \varepsilon\|s\xi\|. \tag{1.5.9}$$

By the triangle inequality and the linearity of the derivatives, we deduce from (1.5.8) and (1.5.9) that

$$\begin{aligned} \forall \xi, \eta \in B(0, r_1),\ s \in]0, 1[, \qquad & \big\|\mathrm{d}f(x_0 + s\xi + \eta) - \mathrm{d}f(x_0 + s\xi) - \mathrm{d}^2 f(x_0)(\eta)\big\| \\ & \leq \varepsilon\left(\|s\xi + \eta\| + \|s\xi\|\right) \leq \varepsilon\left(2\|\xi\| + \|\eta\|\right) \leq 2\varepsilon(\|\xi\| + \|\eta\|). \end{aligned} \tag{1.5.10}$$

Combining (1.5.10) with (1.5.6), we obtain

$$\forall \xi, \eta \in B(0, r_1), \qquad \big\|h(\xi, \eta) - \mathrm{d}^2 f(x_0)(\eta) \cdot \xi\big\| \leq \varepsilon\|\xi\| \cdot (2\|\xi\| + \|\eta\|) \leq 2\varepsilon\|\xi\| \cdot (\|\xi\| + \|\eta\|). \tag{1.5.11}$$

Switching ξ and η in (1.5.11), we obtain

$$\forall \xi, \eta \in B(0, r_1), \qquad \big\|h(\eta, \xi) - \mathrm{d}^2 f(x_0)(\xi) \cdot \eta\big\| \leq 2\varepsilon\|\eta\| \cdot (\|\eta\| + \|\xi\|). \tag{1.5.12}$$

By the triangle inequality and $h(\xi, \eta) = h(\eta, \xi)$, we deduce from (1.5.11) and (1.5.12) that

$$\forall \xi, \eta \in B(0, r_1), \qquad \big\|\mathrm{d}^2 f(x_0)(\eta) \cdot \xi - \mathrm{d}^2 f(x_0)(\xi) \cdot \eta\big\| \leq 2\varepsilon\big(\|\xi\| + \|\eta\|\big)^2. \tag{1.5.13}$$

For any $\xi_1, \eta_1 \in B(0, 1)$, we have $r_1\xi_1, r_1\eta_1 \in B(0, r_1)$. Substituting $\xi = r_1\xi_1$ and $\eta = r_1\eta_1$ into (1.5.13), we obtain

$$\forall \xi_1, \eta_1 \in B(0, 1), \qquad \big\|\mathrm{d}^2 f(x_0)(\eta_1) \cdot \xi_1 - \mathrm{d}^2 f(x_0)(\xi_1) \cdot \eta_1\big\| \leq 2\varepsilon\big\|\|\xi_1\| + \|\eta_1\|\big\|^2 \leq 8\varepsilon. \tag{1.5.14}$$

Let $\beta \in \mathcal{B}(E \times E, F)$ be the bilinear map defined by $\beta(u, v) = \mathrm{d}^2 f(x_0)(v) \cdot u - D^2 f(x_0)(u) \cdot v$, then (1.5.14) implies that $\|\beta\| \leq 8\varepsilon$. Since $\varepsilon > 0$ is arbitrary, we must have $\beta = 0$ and (1.5.1) follows. $\square$

Similarly, one can treat derivatives of arbitrary order. The following is a rigorous inductive definition for this, and the case $k = 1$ is already covered in Definition 1.4.2 and Definition 1.2.7.

Definition 1.5.4. — Let E, F be normed spaces, $U \subseteq E$ an open set, $x_0 \in U$ and $f : U \to E$ a map. For each $k \in \mathbb{N}^+$, we say that f is **differentiable up to order** k, or k**-times differentiable** at x_0, if f is $(k-1)$-times differentiable on a neighborhood of x_0, and the map $\mathrm{d}^{(k-1)} f$ is differentiable at x_0. In this case, we write $\mathrm{d}(\mathrm{d}^{k-1} f)(x_0)$ as $\mathrm{d}^k f(x_0)$, and call it the k**-th derivative of** f at x_0. We say f is k**-times differentiable**, or **differentiable up to order** k, on U, if it is so at each point in U, in which case, we write the map $x_0 \in U \to \mathrm{d}^k f(x_0)$ as $\mathrm{d}^k f$.

Remark 1.5.5. — It is clear that in the above setting, $\mathrm{d} f(x_0) \in \mathcal{L}(E, F)$, $\mathrm{d}^2 f(x_0) \in \mathcal{L}\big(E, \mathcal{L}(E, F)\big)$ et cetera. We adopt the common convention that $f^{(0)}(x_0) := f(x_0)$. In general, setting $\mathcal{L}_0(E, F) := F$, and $\mathcal{L}_k(E, F) := \mathcal{L}\big(E, \mathcal{L}_{k-1}(E, F)\big)$, we have $\mathrm{d}^k f(x_0) \in \mathcal{L}_k(E, F)$ for all $k \in \mathbb{N}$, even for $k = 0$.

Proposition 1.5.6. — *Let $k \in \mathbb{N}^+$, E, F be normed spaces, $U \subseteq E$ an open set and $x_0 \in U$. If a map $f : U \to E$ a map that is k-times differentiable at x_0, then $\mathrm{d}^k f(x_0)$, as a k-linear map from $E^k := E \times \cdots \times E$ (product of k copies of E) to F through the currying process in Proposition 1.1.15, is symmetric.*

Proof. — This follows from Proposition 1.5.3 and the induction on k. □

Definition 1.5.7. — Let E, F be normed spaces, $U \subseteq E$ open. We call a map $f : U \to F$ **smooth**, if it is k-times differentiable for all $k \in \mathbb{N}^+$.

Example 1.5.8. — We will calculate the higher derivatives of various examples in the exercise session. In particular, we will see that continuous (multi)linear maps, as well as the map ν as in Example 1.3.5, are all smooth.

1.6. Products and Differentials, Partial Derivatives

1.6.1. More on Products of Normed Spaces. — Let E be a normed space and $F = F_1 \times \cdots \times F_n$ a finite product of normed spaces $F_1, \ldots, F_n$, normed by (see Corollary 1.1.8 and Remark 1.2.10)

$$\|(y_1, \ldots, y_n)\| = \max_{1 \le i \le n} \|y_i\|. \tag{1.6.1}$$

Let $U \subseteq E$ be open, and $f : U \to F$ a map. For each $i = 1, \ldots, n$, denote by $\pi_i : F \to F_i$ the i-th projection, and set $f_i := \pi_i f$; denote by $\gamma_i : F_i \to F$ the i-th insertion that sends y_i to $(0, \ldots, 0, y_i, 0, \ldots, 0) \in F = F_1 \times \cdots \times F_n$ where the i-th component is y_i and all the other components are 0.

The product structure on F also induces a product structure on $\mathcal{L}(E, F)$. More precisely, we have the following result.

Lemma 1.6.1. — *Using the above notation, for each $i \in \{1, \ldots, n\}$, we have $\gamma_i : F_i \to F$ is an isometric embedding, and $\pi_i : F \to F_i$ a surjective contraction which is of norm 1 unless $F_i = 0$, such that*

$$\forall i \in \{1, \ldots, n\}, \qquad \pi_i \circ \gamma_i = \mathrm{Id}_{F_i}, \tag{1.6.2}$$

and

$$\sum_{i=1}^{n} \gamma_i \circ \pi_i = \mathrm{Id}_F\,. \tag{1.6.3}$$

Moreover, the map

$$\begin{aligned} \pi : \mathcal{L}(E, F) &\to \mathcal{L}(E, F_1) \times \cdots \times \mathcal{L}(E, F_n) \\ A &\to (\pi_1 A, \ldots, \pi_n A) \end{aligned} \tag{1.6.4}$$

is an isometric isomorphism of Banach spaces, where the norm on the right side of (1.6.4) is given by

$$\|(A_1, \ldots, A_n)\| = \max_{1 \le i \le n} \|A_i\|. \tag{1.6.5}$$

Moreover, the inverse of π is given by

$$\pi^{-1}(A_1, \dots, A_n) = \sum_{i=1}^{n} \gamma_i A_i. \tag{1.6.6}$$

Proof. — This is a routine check and the details are left as an exercise. □

Notation 1.6.2. — *From now on, unless stated otherwise, when $F = F_1 \times \cdots \times F_n$ with each F_i being a normed space, we equip F with the norm defined by (1.6.1); and for any normed space, we equip $\mathcal{L}(E, F_1) \times \cdots \times \mathcal{L}(E, F_n)$ with the norm defined by (1.6.5), and identify the resulting normed space with $\mathcal{L}(E, F)$ via the isometric isomorphism π in Lemma 1.6.1.*

1.6.2. Product Structure on the Codomain and the Differential. — Roughly speaking, the following result says that a map is differentiable if and only if each of its component is so. This is an easy consequence of the chain rule and the preparatory Lemma 1.6.1.

Proposition 1.6.3. — *Let E, F_1, $\dots F_n$ be normed spaces, $F = F_1 \times \cdots \times F_n$. Given $U \subseteq E$ open and a map $f : U \to F$. For each $i \in \{1, \dots, n\}$, let $\pi_i : F \to F_i$ be the i-th projection and set $f_i : U \to F_i$ by $f_i = \pi_i \circ f$. Then for any $x_0 \in U$, then f is differentiable at x_0 if and only if f_i is differentiable at x_0 for each $i \in \{1, \dots, n\}$. In this case, we have*

$$\forall i \in \{1, \dots, n\}, \qquad \mathrm{d}f_i(x_0) = \pi_i \circ \mathrm{d}f(x_0), \tag{1.6.7}$$

and

$$\mathrm{d}f(x_0) = \sum_{i=1}^{n} \gamma_i \circ \mathrm{d}f_i(x_0). \tag{1.6.8}$$

In other words, using the identification made in Notation 1.6.2, we have

$$\mathrm{d}f(x_0) = \big(\mathrm{d}f_1(x_0), \dots, \mathrm{d}f_n(x_0)\big). \tag{1.6.9}$$

Proof. — Suppose first f is differentiable at x_0. We know $\pi_i \in \mathcal{L}(F, F_i)$ (Lemma 1.6.1), by Example 1.3.1 and the chain rule (Proposition 1.2.13), we see that for each $i \in \{1, \dots, n\}$, the composite map $f_i = \pi_i \circ f$ is differentiable at x_0 and (1.6.7) holds.

Conversely, suppose each f_i is differentiable at x_0. We have, for each $x \in U$, that

$$\sum_{i=1}^{n} (\gamma_i \circ f_i)(x) = \sum_{i=1}^{n} (\gamma_i \circ \pi_i \circ f)(x) = \left(\sum_{i=1}^{n} \gamma_i \circ \pi_i \right) f(x) = f(x), \tag{1.6.10}$$

where the last equality follows from (1.6.3). It follows from (1.6.10) that as maps from U to F, we have

$$f = \sum_{i=1}^{n} \gamma_i \circ f_i. \tag{1.6.11}$$

Again by the chain rule, each $\gamma_i \circ f_i$ is differentiable at x_0 with $\mathrm{d}(\gamma_i \circ f_i)(x_0) = \gamma_i \circ (\mathrm{d}f_i)(x_0)$, and the differentiability of f at x_0, as well as (1.6.8), now follows from (1.6.11). □

1.6.3. Product on the Domain and the Differential. —

Definition 1.6.4. — Let $E = E_1 \times \cdots \times E_n$ with each E_i being a normed space and let F be a normed space. Suppose $U \subseteq E$ is open and $a \in U$. For each $i \in \{1, \dots, n\}$, denote by $\gamma_{a,i} : E_i \to E$ the isometric map sending $x_i \in E_i$ to $(b_1, \dots, b_n)$ with $b_i = x_i$ and $b_j = a_j$ when $j \neq i$ (check $\gamma_{a,i}$ indeed preserves distance), and set $U_i := \gamma_{a,i}^{-1}(U)$, which is open in E_i and $a_i \in U_i$. We say the ***i*-th partial derivative of** f exists at a if the composite map $f \circ \gamma_{a,i} \restriction_{U_i} : U_i \to F$ is differentiable at a_i, and in this case, we denote by $\partial_i f(a)$ the differential $\mathrm{d}(f \circ \gamma_{a,i})(a_i) \in \mathcal{L}(E_i, F)$, and call it the ***i*-th partial derivative of** f at a.

Notation 1.6.5. — *The partial derivative $\partial_i f(a)$ is often denoted alternatively by $\mathrm{D}_i\, f(a)$.*

Remark 1.6.6. — Intuitively, the i-th partial derivative are the derivative that we get by fixing all but the i-th variable. Note that $\gamma_{a,i}(x_i) = \gamma_i(x_i) + a - \gamma_i(a_i)$ for each $x_i \in E$, where $\gamma_i : E_i \to E$ is the i-th insertion defined at the beginning of § 1.6.1. Hence $\gamma_{a,i} : E_i \to E$ is a continuous affine map, and by Example 1.3.1, we have

$$\forall x_i \in E_i, \qquad \mathrm{d}\gamma_{a,i}(x_i) = \gamma_i. \tag{1.6.12}$$

Note that $\gamma_{a,i}(a_i) = a$. It follows from the chain rule that if f is differentiable at a, then for each $i \in \{1, \ldots, n\}$, the i-th partial derivative $\partial_i f(a)$ at a exists, and

$$\partial_i f(a) = \mathrm{d}f(a) \circ \gamma_i. \tag{1.6.13}$$

Moreover, by (1.6.13) and Lemma 1.6.1, we have

$$\sum_{i=1}^{n} \partial_i f(a) \circ \pi_i = \sum_{i=1}^{n} \mathrm{d}f(a) \circ \gamma_i \circ \pi_i = \mathrm{d}f(a) \left(\sum_{i=1}^{n} \gamma_i \circ \pi_i \right) = \mathrm{d}f(a). \tag{1.6.14}$$

However, conversely, we can *not* deduce the differentiability of f at a from the existence of each partial derivatives of f at a. This can be illustrated by the following counter-example. Consider $g : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g(x_1, x_2) = \begin{cases} \dfrac{2x_1x_2}{x_1^2 + x_2^2}, & \text{if } (x_1, x_2) \neq (0, 0); \\ 0. & \text{otherwise }. \end{cases} \tag{1.6.15}$$

Then by definition, $g(x_1, 0) = g(0, x_2) = 0$ for any $x_1, x_2 \in \mathbb{R}$. Hence both $\partial_1 g(0)$ and $\partial_2 g(0)$ exist, yet g is not even continuous at 0, as $g(t, t) = 1$ for all $t \in \mathbb{R} \setminus \{0\}$. Hence g can not be differentiable at 0 (see Proposition 1.2.12)!

Proposition 1.6.7. — *Using the above setting, the following hold.*

(1) *If f is differentiable at a, then for each $i \in \{1, \ldots, n\}$, the partial derivative $\partial_i f(a)$ exists. Moreover,*

$$\forall i \in \{1, \ldots, n\}, \qquad \partial_i f(a) = \mathrm{d}f(a) \circ \gamma_i, \tag{1.6.16}$$

and

$$\mathrm{d}f(a) = \sum_{i=1}^{n} \partial_i f(a) \circ \pi_i. \tag{1.6.17}$$

(2) *The following are equivalent.*
   (2.a) *We have $f : U \to F$ is of class C^1 (Definition 1.4.2).*
   (2.b) *For each $i \in \{1, \ldots, n\}$ and each $x_0 \in U$, the i-th partial derivative $\partial_i f(x_0)$ at x_0 exists, and the map*

$$\begin{aligned} \partial_i f : U &\to \mathcal{L}(E_i, F) \\ x &\mapsto \partial_i f(x) \end{aligned} \tag{1.6.18}$$

is continuous.

Proof. — We have already established (1) in Remark 1.6.6, and we show (2) in the following.

(2.a) implies (2.b). The existence part of (2.b) already follows from (1) and the assumption (2.a). For the continuity part, note that for each $i \in \{1, \ldots, n\}$, the linear map

$$\begin{aligned} \mathcal{R}_{\gamma_i} : \mathcal{L}(E, F) &\to \mathcal{L}(E_i, F) \\ A &\mapsto A \circ \gamma_i \end{aligned} \tag{1.6.19}$$

is bounded since $\|A \circ \gamma_i\| \leq \|\gamma_i\| \cdot \|A\|$ (so $\|\mathcal{R}_{\gamma_i}\| \leq \|\gamma_i\|$), hence continuous. Using (1.6.16), we see that $\partial_i f = \mathcal{R}_{\gamma_i} \circ (\mathrm{d}f)$, hence is continuous.

(2.b) implies (2.a). Assuming (2.b), we aim to establish (2.a). To simplify the notation in the proof, we first do this in the special case of $n = 2$. In this case, $E = E_1 \times E_2$. We first establish the differentiability of f at an arbitrary fixed point $a = (a_1, a_2) \in U \subseteq E = E_1 \times E_2$. Sine U_1 and U_2 are open with $a_1 \in U_1$ and $a_2 \in U_2$, we may find an $r > 0$, such that $B(a_1, r) \subseteq U_1$ and $B(a_2, r) \subseteq U_2$. We now have

$$
\begin{aligned}
&\forall \xi = (\xi_1, \xi_2) \in B(a_1, r) \times B(a_2, r) = B(a, r),\\
&\quad f(a+\xi) - f(a)\\
&= f(a_1 + \xi_1, a_2 + \xi_2) - f(a_1 + \xi_1, a_2)\\
&\qquad\qquad + f(a_1 + \xi_1, a_2) - f(a_1, a_2).
\end{aligned}
\tag{1.6.20}
$$

This motivate us to consider the maps $\alpha, \beta : [0,1] \to F$ defined by

$$\alpha(s) = f(a_1 + \xi_1, a_2 + s\xi_2) \tag{1.6.21}$$

and

$$\beta(s) = f(a_1 + s\xi_1, a_2) \tag{1.6.22}$$

respectively. Denote by $\tau_i : [0,1] \to E_i$ the continuous affine map $t \mapsto a_2 + t\xi_i$ $(i = 1, 2)$. Then it follows from the definition of the corresponding maps that

$$\alpha = f \circ \gamma_{(a_1+\xi_1, a_2), 2} \circ \tau_2 \tag{1.6.23}$$

and

$$\beta = f \circ \gamma_{(a_1, a_2+\xi_2), 1} \circ \tau_1. \tag{1.6.24}$$

Note that the map $\gamma_{(a_1+\xi_1, a_2), 2} : E_2 \to F$ is independent of a_2. For each $s \in]0,1[$, we have

$$\gamma_{(a_1+\xi_1, a_2), 2} = \gamma_{(a_1+\xi_1, a_2+s\xi_2), 2} \qquad \text{and} \qquad f \circ \gamma_{(a_1+\xi_1, a_2), 2} = f \circ \gamma_{(a_1+\xi_1, a_2+s\xi_2), 2}. \tag{1.6.25}$$

By assumption (2.b), we know that $f \circ \gamma_{(a_1+\xi_1, a_2+s\xi_2), 2}$ is differentiable at the point $\tau_2(s) = a_2 + s\xi_2$, it follows from (1.6.23), (1.6.25) and the chain rule that α is differentiable at s, and

$$\forall s \in]0,1[, \qquad \mathrm{d}\alpha(s) = \Big(\mathrm{d}(f \circ \gamma_{(a_1+\xi_1, a_2+s\xi_2)})\big(\tau_2(s)\big)\Big) \circ \mathrm{d}\tau_2(s) = (\partial_2 f)(a_1 + \xi_1, a_2 + s\xi_2) \circ R_{\xi_2}, \tag{1.6.26}$$

where $R_{\xi_2} \in \mathcal{L}(R, E_2)$ is the map $t \in \mathbb{R} \mapsto t\xi_2 \in E_2$ with $\|R_{\xi_2}\| = \|\xi_2\|$. Similarly, we have

$$\forall s \in]0,1[, \qquad \mathrm{d}\beta(s) = (\partial_1 f)(a_1 + s\xi_1, a_2) \circ R_{\xi_1}. \tag{1.6.27}$$

For each $s \in [0,1]$, we set

$$\alpha_0(s) = \alpha(s) - (\partial_2 f)(a_1, a_2)(s\xi_2), \tag{1.6.28}$$

and

$$\beta_0(s) = \beta(s) - (\partial_1 f)(a_1, a_2)(s\xi_1). \tag{1.6.29}$$

If follows from (1.6.26) and (1.6.28) that

$$\forall s \in]0,1[, \qquad \mathrm{d}\alpha_0(s) = \Big[(\partial_2 f)(a_1 + \xi_1, a_2 + s\xi_2) - (\partial_2 f)(a_1, a_2)\Big] \circ R_{\xi_2}. \tag{1.6.30}$$

Hence by the mean value theorem (Theorem 1.4.1), we see that

$$
\begin{aligned}
&\|f(a_1 + \xi_1, a_2 + \xi_2) - f(a_1 + \xi_1, a_2) - (\partial_2 f)(a_1, a_2)\xi_2\|\\
&= \|\alpha_0(1) - \alpha_0(0)\| \leq \sup_{s \in]0,1[} \|\mathrm{d}\alpha_0(s)\|\\
&\leq \|\xi_2\| \sup_{s \in]0,1[} \Big\|(\partial_2 f)(a_1 + \xi_1, a_2 + s\xi_2) - (\partial_2 f)(a_1, a_2)\Big\|\\
&\leq \|\xi\| \sup_{s \in]0,1[} \Big\|(\partial_2 f)(a_1 + \xi_1, a_2 + s\xi_2) - (\partial_2 f)(a_1, a_2)\Big\|.
\end{aligned}
\tag{1.6.31}
$$

Similarly, by (1.6.27) and (1.6.29), we have

$$\forall s \in]0,1[, \qquad \mathrm{d}\beta_0(s) = \Big[(\partial_1 f)(a_1 + s\xi_1, a_2) - (\partial_1 f)(a_1, a_2)\Big] \circ R_{\xi_1}, \tag{1.6.32}$$

and together with the mean value theorem, we have

$$
\begin{aligned}
&\|f(a_1+\xi_1,a_2)-f(a_1,a_2)\| \\
&=\|\beta_0(1)-\beta_0(0)\| \le \sup_{s\in]0,1[}\|\mathrm{d}\beta_0(s)\| \\
&\le \|\xi_1\| \sup_{s\in]0,1[}\Big\|(\partial_1 f)(a_1+s\xi_1,a_2)-(\partial_1 f)(a_1,a_2)\Big\| \\
&\le \|\xi\| \sup_{s\in]0,1[}\Big\|(\partial_1 f)(a_1+s\xi_1,a_2)-(\partial_1 f)(a_1,a_2)\Big\|.
\end{aligned} \tag{1.6.33}
$$

Now by (2.b), $\partial_1 f$ and $\partial_2 f$ are continuous. Hence by shrinking $r>0$ if necessary, we have

$$\forall \zeta=(\zeta_1,\zeta_2)\in B(0,r)\subseteq E=E_1\times E_2,\ \forall i=1,2, \qquad \|\partial_i f(a+\zeta)-\partial_i f(a)\|\le \frac{\varepsilon}{2}. \tag{1.6.34}$$

By (1.6.31) and (1.6.34), we have

$$\|f(a_1+\xi_1,a_2+\xi_2)-f(a_1+\xi_1,a_2)-(\partial_2 f)(a_1,a_2)\xi_2\|\le \frac{\varepsilon}{2}\|\xi\|, \tag{1.6.35}$$

while (1.6.33) and (1.6.34) together imply that

$$\|f(a_1+\xi_1,a_2)-f(a_1,a_2)-(\partial_1 f)(a_1,a_2)\xi_1\|\le \frac{\varepsilon}{2}\|\xi\|. \tag{1.6.36}$$

It follows from (1.6.35) and (1.6.36) that

$$\forall \xi\in B(0,r)\subseteq E, \qquad \|\Delta_a f(\xi)-(\partial_1 f)(a).\xi_1-(\partial_2 f)(a).\xi_2\|\le \varepsilon\|\xi\|. \tag{1.6.37}$$

This means that f is indeed differentiable at a, with

$$\mathrm{d}f(a).\xi=(\partial_1 f)(a).\xi_1+(\partial_2 f)(a).\xi_2. \tag{1.6.38}$$

This being so for every $a\in U$, we see as maps from U to $\mathcal{L}(E,F)$, we have

$$\mathrm{d}f=\mathcal{R}_{\pi_1}\circ(\partial_1 f)+\mathcal{R}_{\pi_2}\circ(\partial_2 f) \tag{1.6.39}$$

Now from (1.6.38), the continuity of $\mathrm{d}f$ follows from the continuity of the partial derivatives and the continuity of $\mathcal{R}_{\pi_i}$ $(i=1,2)$, the latter being defined in the obvious way as in (1.6.19). This finishes the proof in the special case of $n=2$.

For general n, we can finish the proof by induction on n by noting that we may canonically identify $E_1\times\cdots\times E_n$ with $(E_1\times\cdots\times E_{n-1})\times E_n$. The details are left as an exercise. □

1.7. Differential of Maps between Euclidean Spaces, Mixed Derivatives

We now explain how the general theory we have developed so far applies to the special case of maps between Euclidean spaces.

Recall that an Euclidean space is a finite dimensional inner product space that is isometrically isomorphic to some $\mathbb{R}^n$ equipped with its canonical inner product

$$(x,y)=\sum_{i=1}^n x_i y_i, \qquad x=(x_1,\dots,x_n),\ y=(y_1,\dots,y_n)\in\mathbb{R}^n. \tag{1.7.1}$$

The canonical inner product on $\mathbb{R}^n$ induces the 2-norm as described in Example 1.1.5.

Given $m,n\in\mathbb{N}^+$, $U\subseteq\mathbb{R}^n$ open, and $f:U\to\mathbb{R}^m$. Most classical treatment of differential calculus studies the differentiability (resp. higher differentiability) of f in terms of the partial derivatives (resp. mixed derivatives), which we now describe in more detail.

Since each vector in $\mathbb{R}^m$ has m-components, we may rewrite f as a row of m maps $(f_1,\dots,f_m)$, with each f_i being a map from $U\to\mathbb{R}^m$, such that $f(x)=(f_1(x),\dots,f_m(x))$. Fix $a\in U$ and $j=1,\dots,n$, and let $x\in U$ vary near a with the restriction that $x_k=a_k$, $k\ne j$, i.e. only the j-th component x_j can vary. For each f_i, $i=1,\dots,m$, we write $\frac{\partial f_i}{\partial x_j}(a)$ as the derivative $g_a'(a_j)$ of the function g at a_j, if the latter exists, where g is defined in a small neighborhood of $a_j\in\mathbb{R}$ by the formula

$$g(t)=f_i(a_1,\dots,a_{j-1},t,a_{j+1},\dots,a_n).$$

Note that $\frac{\partial f_i}{\partial x_j}(a)$ as described above is really the derivative of a real-valued function of a single real variable, with which we are already familiar.

Definition 1.7.1. — In the above setting, we call the number $\frac{\partial f_i}{\partial x_j}(a)$ the **partial derivative** of the i-th component of f with respect to the j-th variable at a. In a notation that more reflects the partial differential operation, this partial derivative is often also denoted by

$$\frac{\partial}{\partial x_j}\Big|_a f_i. \tag{1.7.2}$$

Notation 1.7.2. — *To facilitate discussion in the rest of this section, we adopt the following convention of notation. For each $\mathbb{R}^k$, we use $e_{i,k}$, $(i = 1, \ldots, k)$ to denote the i-th vector in the standard basis, i.e. $e_{i,k} \in \mathbb{R}^k$ with the l-th coordinate of $e_{i,k}$ is $\delta_{i,l}$. Correspondingly, we use $\omega_{i,k} : \mathbb{R}^k \to \mathbb{R}$ to denote the i-th projection. When the dimension k is clear from the context, we often omit it from the subscript and write simply e_i and ω_i.*

Unwinding the definitions, we immediately obtain the following result relating the partial derivative $\frac{\partial f_i}{\partial x_j}(a)$ to the already familiar notion of the directional derivative (Definition 1.2.17).

Proposition 1.7.3. — *In the above setting, for each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, and for any $a \in U$, the partial derivative $\frac{\partial f_i}{\partial x_j}(a)$ of the i-th component of f with respect to the j-th variable at a is exactly the directional derivative*

$$\lim_{t\to 0} \frac{f_i(a + te_j) - f_i(a)}{t}$$

of the i-th component of f along the direction $e_j \in \mathbb{R}^n$. □

Corollary 1.7.4. — *In the above setting, if f is differentiable at a, then for all $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$, the partial derivatives $\frac{\partial f_i}{\partial x_j}(a)$ exists, and the matrix*

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} \tag{1.7.3}$$

is exactly the matrix for the linear map $\mathrm{d}f(a) : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard basis of $\mathbb{R}^m$ and $\mathbb{R}^n$.

Proof. — The existence part follows directly from Proposition 1.6.3, Proposition 1.2.16 and the above Proposition 1.7.3. For the matrix part, just note that the (i, j)-entry for the matrix of $\mathrm{d}f(a)$ is precisely the number

$$\omega_{i,n}\Big(\mathrm{d}f(a).e_{j,m}\Big) = \mathrm{d}f_i(a).e_{j,m} = \frac{\partial f_i}{\partial x_j}(a),$$

where the first equality follows from the chain rule and Proposition 1.2.16 again, and the second from Proposition 1.7.3. □

Definition 1.7.5. — When all the partial derivatives of f at a exist, the matrix (1.7.3) is called the **Jacobian matrix** of f at a, after the German mathematician Carl Jacobi. This matrix is often denoted by

$$Jf(a) = \frac{\partial(f_1, \ldots, f_m)}{\partial(x_1, \ldots, x_n)} = \left(\frac{\partial f_i}{\partial x_j}\right)_{m\times n}, \tag{1.7.4}$$

or simply by $\left(\frac{\partial f_i}{\partial x_j}\right)$. When $m = n$, the determinant of the Jacobian matrix $Jf(a)$ is called **the Jacobian** of f at a, and is often denoted by

$$|Jf(a)| = |\partial(f_1, \ldots, f_n)|\partial(x_1, \ldots, x_n) = \left|\left(\frac{\partial f_i}{\partial x_j}\right)\right|.$$

Remark 1.7.6. — It follows from Corollary 1.7.4 that the matrix under the standard basis for $\mathrm{d}f(a) : \mathbb{R}^n \to \mathbb{R}^m$ is exactly the Jacobian matrix $Jf(a)$. But conversely, if the Jacobian matrix $Jf(a)$ exists, i.e. if all partial derivatives of f exist at certain $a \in U$ does *not* imply the existence of the total derivative $\mathrm{d}f(a)$. As a counter-example, see the map $g : \mathbb{R}^2 \to R$ in (1.6.15) at the point $0 = (0, 0) \in \mathbb{R}^2$.

Nevertheless, if $\frac{\partial f_i}{\partial x_j}(a)$ exists for each i, j, and $a \in U$, and if each $\frac{\partial f_i}{\partial x_j}(a)$ is continuous as a function of $a \in U$ on U, then $f : U \to \mathbb{R}^m$ is of class C^1 according to Proposition 1.6.7.

In the special case of $m = 1$, the above $f : U \to \mathbb{R}$ has only one component functions, and the Jacobian matrix $Jf(a)$ becomes the row vector

$$\left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right) \in \mathbb{R}^n. \tag{1.7.5}$$

Definition 1.7.7. — In vector analysis, the row vector (1.7.5) is also called **the gradient** of f at a, and is often denoted by $\operatorname{grad} f(a)$, or by $\nabla f(a)$. The symbol ∇ is known as the **nabla operator**.

Definition 1.7.8. — Keep the above general setting $f : U \to \mathbb{R}^m$ with $U \subseteq \mathbb{R}^n$ open and $a \in U$. Suppose $\frac{\partial f_i}{\partial x_j}(x)$ exists for all x in a neighborhood of a, then $\frac{\partial f_i}{\partial x_j}$ is a well-defined map from this neighborhood of a to $\mathbb{R}^m$. If this latter map is partially differentiable with respect to the k-th variable, the number

$$\left.\frac{\partial}{\partial x_k}\right|_a \left(\frac{\partial f_i}{\partial x_j}\right) \tag{1.7.6}$$

is termed the **mixed derivative** (of order 2) of the i-th component of f first with respect to the j-th, then with respect to the k-th component at a. We often denote the mixed derivative (1.7.6) by

$$\frac{\partial^2 f_i}{\partial x_k \partial x_j}(a) \quad \text{or} \quad \left.\frac{\partial^2}{\partial x_k \partial x_j}\right|_a f_i. \tag{1.7.7}$$

One may continue the process to define **mixed derivatives of higher order** by the induction on the order, i.e. the number of times of taking partial derivatives, at a. More precisely, if the mixed derivatives of order $l-1$

$$\frac{\partial^{l-1} f_i}{\partial x_{j_{l-1}} \cdots \partial x_{j_1}}(x) = \left.\frac{\partial^{l-1}}{\partial x_{j_{l-1}} \cdots \partial x_{j_1}}\right|_a f_i \tag{1.7.8}$$

exist for all x in a neighborhood of a, and j_l an index in $\{1, \ldots, n\}$, then we call the number (if it exists)

$$\left.\frac{\partial}{\partial x_{j_l}}\right|_a \frac{\partial^{l-1} f_i}{\partial x_{j_{l-1}} \cdots \partial x_{j_1}} = \frac{\partial}{\partial x_{j_l}}\left(\frac{\partial^{l-1} f_i}{\partial x_{j_{l-1}} \cdots \partial x_{j_1}}\right)(a) \tag{1.7.9}$$

the **mixed derivative of order** l of the i-th component of f at a by taking partial derivatives with respect to $x_{j_1}, x_{j_2}, \ldots, x_{j_l}$ consecutively, and denote it by

$$\frac{\partial^l f_i}{\partial x_{j_l} \partial x_{j_{l-1}} \cdots \partial x_{j_1}}(x) = \left.\frac{\partial^l}{\partial x_{j_l} \partial x_{j_{l-1}} \cdots \partial x_{j_1}}\right|_a f_i. \tag{1.7.10}$$

The following result says that when taking mixed derivatives, as long as the function has enough regularity, the order with respect to which we take the partial derivatives is of minor importance.

Proposition 1.7.9. — *Let $U \subseteq \mathbb{R}^n$ be open, $a \in U$, $k \in \mathbb{N}^+$ and $l \in \mathbb{N}^+$ with $l \leq k$. If a map $f : U \to \mathbb{R}$ is differentiable up to order $k \in \mathbb{N}^+$, then every mixed derivatives of f of order l at a exists. Moreover, if $i_1, \ldots, i_l$ and $j_1, \ldots, j_l$ are two sequences of indices in $\{1, \ldots, n\}$ that are a rearrangement of each other, then*

$$\frac{\partial^l f}{\partial x_{i_l} \cdots \partial x_{i_1}}(a) = \frac{\partial^l f}{\partial x_{j_l} \cdots \partial x_{j_1}}(a). \tag{1.7.11}$$

Proof. — This is an easy consequence of various results we have obtained so far, with the key part being Proposition 1.5.6. The details are left as an exercise. □

***Notation 1.7.10* (The multi-index notation).** — *Given an n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ with each $\alpha_i \in \mathbb{N}$ (so it could happen that some $\alpha_i = 0$), we will write*

$$|\alpha| = \sum_{i=1}^{n} \alpha_i \tag{1.7.12}$$

*and call it the **order** of α. We also write*

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \tag{1.7.13}$$

and

$$\frac{\partial^{|\alpha|}}{\partial x^\alpha} = \frac{\partial^{\alpha_1}}{\partial x_1 \cdots \partial x_1} \circ \cdots \circ \frac{\partial^{\alpha_n}}{\partial x_n \cdots \partial x_n}., \tag{1.7.14}$$

where in (1.7.14)*, the part*

$$\frac{\partial^{\alpha_i}}{\partial x_i^{\alpha_i}} := \frac{\partial^{\alpha_i}}{\partial x_i \cdots \partial x_i} \tag{1.7.15}$$

means taking the operation $\frac{\partial}{\partial x_i}$ *repeatedly for* $|\alpha_i|$ *times, and by convention, if* $\alpha_i = 0$*,* (1.7.15) *is just the identity.*

Thus in the setting of Proposition 1.7.9*, if* $|\alpha| = l$*, and if the index* i *occurred in* $(i_1, \dots, i_l)$ *exactly* α_i *times, we will write the mixed derivative* (1.7.11) *simply as*

$$\frac{\partial^\alpha f}{\partial x^\alpha}(a), \qquad \frac{\partial^\alpha}{\partial x^\alpha}\Big|_a f \qquad \text{or just} \qquad \partial_\alpha f(x_0). \tag{1.7.16}$$

We will see in the exercises some basic properties and usage of this neat multi-index notation.

1.8. Banach Space Valued Integrals of a Map of One Real Variable

Fix a Banach space E and a compact interval $[a, b]$ in $\mathbb{R}$. We diverge to talk about E-valued integrals on $[a, b]$. We give a quick treatment for $\int_a^b f$ when $f : [a, b] \to E$ is continuous, using approximation by step functions. In the exercise session, we will see an alternative (but slightly more involved) treatment using the more familiar Riemann sums.

Definition 1.8.1. — A map $f : [a, b] \to E$ is called an E**-valued step function**, or simply a **step function**, if there exists a partition $a = t_0 < t_1 < \cdots < t_n = b$ of $[a, b]$, such that the restriction $f \restriction_{]t_{i-1}, t_i[}$ is constant for every $1 \leq i \leq n$. Let $c_i \in E$ be the constant value for $f \restriction_{]t_{i-1}, t_i[}$, the **(Riemann) integral** of f, denoted by $\int_a^b f(t)\mathrm{d}t$, or simply by $\int_a^b f$, is defined as

$$\int_a^b f(t)\mathrm{d}t = \sum_{i=1}^n c_i(t_i - t_{i-1}). \tag{1.8.1}$$

Remark 1.8.2. — It is clear that whenever we have a refinement $a = s_0 < s_1 < \cdots < s_m = b$ of the partition $(t_i)_{0 \leq i \leq n}$, we still have $f \restriction_{]s_{j-1}, s_j[}$, $1 \leq j \leq m$ is constant. Hence, we should, according to Definition 1.8.1, assign $\int_a^b f(t)\mathrm{d}t$ to a similar sum using the partition $(s_j)_{0 \leq j \leq m}$. We leave as an exercise to check that the integral in (1.8.1) actually does not depend on the choice of the partition $(t_i)_{0 \leq i \leq n}$ for which each $f \restriction_{]t_{i-1}, t_i[}$ is constant, hence the integral $\int_a^b f(t)\mathrm{d}t$ for a step function f is well-defined.

Notation 1.8.3. — *We use* $\mathsf{S}([a, b], E)$ *to denote the space of step functions from* $[a, b]$ *to* E*. It is clear that* $\mathsf{S}([a, b], E)$ *is stable under scalar multiplication, pointwise sum and pointwise product.*

Proposition 1.8.4. — *The following hold.*

(1) *(Linearity of* $\int_a^b$*) For all* $f, g \in \mathsf{S}([a, b], E)$*, we have* $\int_a^b (f + g) = \int_a^b f + \int_a^b g$*; and for all* $c \in \mathbb{R}$*,* $f \in \mathsf{S}([a, b], E)$*, we have* $\int_a^b (cf) = c \int_a^b f$*.*

(2) *For each* $f \in \mathsf{S}([a, b], E)$*, we have*

$$\left\| \int_a^b f(t)\mathrm{d}t \right\| \leq \int_a^b \|f(t)\|\mathrm{d}t. \tag{1.8.2}$$

(3) *For each* $f \in \mathsf{S}([a, b], E)$ *and each* $c \in \,]a, b[$*, we have*

$$\int_a^b f(t)\mathrm{d}t = \int_a^c f(t)\mathrm{d}t + \int_c^b f(t)\mathrm{d}t. \tag{1.8.3}$$

Proof. — (1). The additivity can be checked by using a common refinement of suitable partitions for f and g, while the homogeneity is obtained by directly applying (1.8.1).

(2). Using the notation in defining (1.8.1), we have

$$\left\|\int_a^b f\right\| = \left\|\sum_{i=1}^n c_i(t_i - t_{i-1})\right\| \leq \sum_{i=1}^n \|c_i\|(t_i - t_{i-1}) = \int_a^b \|f(t)\|\mathrm{d}t, \tag{1.8.4}$$

where the inequality follows from the triangle inequality of the norm.

(3). Let $(t_i)_{i=0}^n$ be a partition of $[a, b]$ such that f restricts to a constant on each interval $]t_{i-1}, t_i[$. Either $c = t_i$ for some i, or $c \in\,]t_{i-1}, t_i[$ for some i. Adding c to the partition (see Remark 1.8.2), we may always assume that $c = t_i$ for some i. Then (1.8.3) follows immediately using this partition. □

We now proceed to treat the E-valued Riemann integrals of $f : [a, b] \to E$ when f is continuous.

Proposition 1.8.5. — *Let E be a Banach space and $[a, b]$ a compact interval. Suppose $f : [a, b] \to E$ is continuous. For each $\varepsilon > 0$, we set*

$$\mathsf{S}_f(\varepsilon) := \left\{ g \in \mathsf{S}([a, b], E) \;\middle|\; \|f - g\|_\infty := \sup_{t\in[a,b]} |f(t) - g(t)| \leq \varepsilon \right\}. \tag{1.8.5}$$

Then the following hold.

(1) *The set $\mathsf{S}_f(\varepsilon) \neq \emptyset$.*

(2) *For all $g_1, g_2 \in \mathsf{S}_f(\varepsilon)$, we have*

$$\|g_1 - g_2\|_\infty \leq 2\varepsilon \qquad \text{and} \qquad \left\|\int_a^b g_1 - \int_a^b g_2\right\| \leq 2\varepsilon(b - a). \tag{1.8.6}$$

(3) *There exists a unique $\eta \in E$, such that*

$$\forall \varepsilon > 0, \qquad \left\{ \int_a^b g \;\middle|\; g \in S_f(\varepsilon) \right\} \subseteq \overline{B(\eta, 2\varepsilon|a - b|)}. \tag{1.8.7}$$

Proof. — (1). The key is the compactness of $[a, b]$. For each $t \in [a, b]$, by continuity of f at t, we may find an $r_t > 0$, such that

$$s \in\,]t - 2r_t, t + 2r_t[\, \cap [a, b] \implies |f(t) - f(s)| \leq \frac{\varepsilon}{2}. \tag{1.8.8}$$

Since $]t - r_t, t + r_t[$, $t \in [a, b]$ form an open cover of the compact $[a, b]$, this open cover admits a finite subcover, i.e. there exists a finite sequence $t_1, \ldots, t_n$ in $[a, b]$ with

$$[a, b] \subseteq \cup_{i=1}^n]t_i - r_{t_i}, t_i + r_{t_i}[. \tag{1.8.9}$$

Set $\delta := \min_{1\leq i\leq n} r_{t_i} > 0$. Now take any $s_1, s_2 \in [a, b]$ with $|s_1 - s_2| \leq \delta$, we have $s_1 \in\,]t_i - r_{t_i}, t + r_{t_i}[$, and $|s_1 - s_2| \leq \delta \leq r_{t_i}$ implies that

$$s_1, s_2 \in\,]t_i - 2r_{t_i}, t_i + 2r_{t_i}[\, \cap [a, b]. \tag{1.8.10}$$

It follows from (1.8.10) and (1.8.8) that

$$|f(s_1) - f(s_2)| \leq |f(s_1) - f(t_i)| + |f(s_2) - f(t_i)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{1.8.11}$$

Therefore, we have established

$$\forall s_1, s_2 \in [a, b],\ |s_1 - s_2| \leq \delta \implies |f(s_1) - f(s_2)| \leq \varepsilon. \tag{1.8.12}$$

Now choose an $N \in \mathbb{N}$ big enough so that $N^{-1}(b - a) < \delta$, and let $(s_i)_{i=0}^N$ be the partition of $[a, b]$ into N equal parts. Set $g_N \in \mathsf{S}([a, b], E)$ to be the step function defined by $g_N(s) = f(s_i)$ if $s \in [s_i, s_{i+1}[$ for some $1 \leq i \leq N - 1$ and $g_N(s_N) = f(b)$. It follows from the definition of g and (1.8.12) that $g \in \mathsf{S}_f(\varepsilon)$.

(2). Clearly

$$\|g_1 - g_2\|_\infty \leq \|g_1 - f\| + \|g_2 - f\| \leq 2\varepsilon. \tag{1.8.13}$$

Moreover, by Proposition 1.8.4, we have

$$\left\|\int_a^b g_1 - \int_a^b g_2\right\| \leq \int_a^b \|g_1(t) - g_2(t)\|\mathrm{d}t \leq \int_a^b \|g_1 - g_2\|_\infty \leq 2\varepsilon(b - a). \tag{1.8.14}$$

(3) follows from (2) by noting that in a complete metric space, the intersection of a decreasing sequence of closed balls whose radii converge to 0 is a (unique) singleton. The details are left as an exercise. ☐

Definition 1.8.6. — In the setting of Proposition 1.8.5, we call η the **Riemann integral** of the continuous map f, and write

$$\eta = \int_a^b f = \int_a^b f(t)\mathrm{d}t. \tag{1.8.15}$$

We have a parallel result for Proposition 1.8.4 concerning integrals of continuous maps.

Proposition 1.8.7. — *The following hold.*

(1) *(Linearity of $\int_a^b$) For all $f, g \in C([a,b],E)$, we have $\int_a^b(f+g) = \int_a^b f + \int_a^b g$; and for all $c \in \mathbb{R}$, $f \in C([a,b],E)$, we have $\int_a^b(cf) = c\int_a^b f$.*

(2) *For each $f \in C([a,b],E)$, we have*

$$\left\|\int_a^b f(t)\mathrm{d}t\right\| \le \int_a^b \|f(t)\|\mathrm{d}t. \tag{1.8.16}$$

(3) *For each $f \in C([a,b],E)$ and each $c \in \,]a,b[$, we have*

$$\int_a^b f(t)\mathrm{d}t = \int_a^c f(t)\mathrm{d}t + \int_c^b f(t)\mathrm{d}t. \tag{1.8.17}$$

Proof. — This follows from Proposition 1.8.4 by a routine approximation procedure. The details are left as an exercise. ☐

The following result is useful when we study differentiation under the integral sign.

Proposition 1.8.8. — *Let E be a normed space, F a Banach space, $[a,b]$ a compact interval, and $f \in C\big([a,b],\mathcal{L}(E,F)\big)$. Then $\int_a^b f(t)\mathrm{d}t \in \mathcal{L}(E,F)$ is the bounded linear operator sending each $\xi \in E$ to $\int_a^b \big(f(t)\xi\big)\mathrm{d}t$.*

Proof. — For $\xi = 0$, this is obvious. We thus suppose $\xi \neq 0$ in the sequel. The evaluation at ξ is a continuous map $\mathcal{L}(E,F) \to F$, hence $f_\xi(t) := f(t)\xi$ defines a continuous map $g \in C([a,b],F)$. Thus $\int_a^b (f(t)\xi)\,\mathrm{d}t = \int_a^b f_\xi$ makes sense. Let $\varepsilon > 0$ and choose a $g \in \mathsf{S}_f(\varepsilon)$. We have

$$\forall t \in [a,b], \qquad \|g_\xi(t) - f_\xi(t)\| \le \|g(t) - f(t)\| \cdot \|\xi\| \le \|\xi\|\varepsilon. \tag{1.8.18}$$

Hence $g_\xi \in \mathsf{S}_f(\|\xi\|\varepsilon)$. By Proposition 1.8.7, we have

$$\left\|\int_a^b f - \int_a^b g\right\| \le \varepsilon(b-a), \tag{1.8.19}$$

and

$$\left\|\int_a^b f_\xi - \int_a^b g_\xi\right\| \le \varepsilon(b-a)\|\xi\|. \tag{1.8.20}$$

It is clear that for the step function g, we have

$$\left(\int_a^b g\right)\xi = \int_a^b g_\xi = \int_a^b \big(g(t)\xi\big)\mathrm{d}t. \tag{1.8.21}$$

Applying the bounded linear operator $\int_a^b (f-g)$ to ξ, we have, by (1.8.19), that

$$\left\|\left(\int_a^b f - \int_a^b g\right)\xi\right\| \le \varepsilon(b-a)\|\xi\|. \tag{1.8.22}$$

It follows from (1.8.20), (1.8.21) and (1.8.22) that

$$\left\|\left(\int_a^b f\right)\xi - \int_a^b f(t)\xi\mathrm{d}t\right\| \le 2\varepsilon(b-a)\|\xi\|. \tag{1.8.23}$$

Since ε is arbitrary, (1.8.23) implies that

$$\left(\int_a^b f\right)\xi = \int_a^b f(t)\xi \mathrm{d}t \tag{1.8.24}$$

and the Proposition is proved. □

We will also need the following version of the fundamental theorem of calculus.

Proposition 1.8.9. — *Let E be a Banach space, $[a,b[$ an interval in $\mathbb{R}$, and $f : [a,b[\to E$ a continuous map. Define $F(t) := \int_a^t f(s)\mathrm{d}s$ for each $t \in]a,b[$, then $F :]a,b[\to E$ is differentiable with $F'(t) = f(t)$ for every $t \in]a,b[$.*

Proof. — Let $\int_\alpha^\beta = -\int_\beta^\alpha$ if $\beta < \alpha$. With this convention, we have, whenever $t \in]a,b[$ and $h \in \mathbb{R}$ with $t + h \in]a,b[$, that

$$F(t+h) - F(t) - hf(t) = \int_t^{t+h} \big(f(s) - f(t)\big)\mathrm{d}s. \tag{1.8.25}$$

The continuity of f at t implies that for each $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $|h| < \delta$, we have $[t - |h|, t + |h|] \subseteq]a,b[$, and $|f(s) - f(t)| \le \varepsilon$ for any $s \in [t - |h|, t + |h|]$. Thus $|f(s) - f(t)| \le \varepsilon$ for any s between t and $t + h$. A routine discussion of cases according to h being positive or negative shows that

$$\forall h \in]-\delta,\delta[, \qquad |F(t+h) - F(t) - hf(t)| \le \varepsilon |h|. \tag{1.8.26}$$

And (1.8.26) shows that F is differentiable at t with $F'(t) = f(t)$. □

The following result of evaluating a Riemann integral of a continuous map using its anti-derivate is useful when we study Taylor formula in § 1.10.

***Corollary 1.8.10* (Fundamental theorem of calculus, Banach space valued version)**

Let E be a Banach space and $[a,b]$ a compact interval, $\varphi : [a,b] \to E$ continuous. If $\Phi : [a,b] \to E$ is continuous and Φ is differentiable at any $t \in]a,b[$ with $\Phi'(t) = \varphi(t)$, then $\int_a^b \varphi(t)\mathrm{d}t = \Phi(b) - \Phi(a)$.

Proof. — Define $\Psi(t) = \int_a^t \varphi(s)\mathrm{d}s$. Then $\Psi : [a,b] \to E$ is continuous by Proposition 1.8.7. Moreover, $\Psi - \Phi$ is continuous, and $(\Psi - \Phi)'(t) = \varphi(t) - \varphi(t) = 0$ for any $t \in]a,b[$. Thus $\Psi - \Phi$ is constant on $]a,b[$, hence constant on $[a,b]$ by continuity. Consequently, we have

$$\int_a^b f(t)\mathrm{d}t = \Psi(b) - \Psi(a) = \Phi(b) - \Phi(a) \tag{1.8.27}$$

as desired. □

1.9. Integrals Depending on a Parameter

Integrals depending on a parameter is a very useful tool. We now describe some basic results in this topic. Deeper results require the theory of Lebesgue integrals and will be presented in later courses.

In these notes, we focus on real-valued functions so that the familiar theory of Riemann integrals can be applied. In the exercise session, we will see that these results can be generalized to Banach space valued maps without much difficulty.

To set the stage up, consider a normed space E, a Banach space F, an open set $U \subseteq E$, and a compact interval $[a,b]$ in $\mathbb{R}$. Note that $[a,b] \times U$ is a subset of the product normed space $\mathbb{R} \times E$, hence is a metric space with the metric induced from the norm on $\mathbb{R} \times E$. Let

$$f : [a,b] \times U \to F \tag{1.9.1}$$

be a map such that for each fixed $\xi \in U$, the map

$$\begin{aligned} f_\xi : [a,b] &\to F \\ t &\mapsto f(t,\xi) \end{aligned} \tag{1.9.2}$$

is continuous, hence we can take its Riemann integral as in § 1.8. Set

$$g(\xi) = \int_a^b f_\xi(t)\mathrm{d}t = \int_a^b f(t,\xi)\mathrm{d}t. \tag{1.9.3}$$

We regard $\xi \in U$ as a parameter for the integrals on the right side of (1.9.3).

Before we proceed further, we record a particular instance of the so-called "tube lemma".

Lemma 1.9.1. — *Using the above setting, suppose $\xi_0 \in U$ and W is an open set in $[a,b] \times U$. If $[a,b] \times \{\xi_0\} \subseteq W$, then there exists an open neighborhood. V of ξ_0, such that $[a,b] \times V \subseteq W$.*

Proof. — For each $t \in [a,b]$, since W is open in $[a,b] \times U$, there exists $r_t > 0$, such that

$$B\Big((t,\xi_0), r_t\Big) \cap \big([a,b] \times U\big) \subseteq W. \tag{1.9.4}$$

Note the convention of the norm on $\mathbb{R} \times E$ made in Notation 1.6.2, inside $\mathbb{R} \times E$, we have

$$B\Big((t,\xi_0), r_t\Big) = B(t, r_t) \times B(\xi_0, r_t). \tag{1.9.5}$$

By compactness of $[a,b]$ and the fact that $B(t,r_t)$, $t \in [a,b]$ form an open cover of $[a,b]$, there exists a finite sequence $t_1, \ldots, t_n$ in $[a,b]$, such that

$$[a,b] \subseteq \cup_{i=1}^n B\left(t_i, r_{t_i}\right). \tag{1.9.6}$$

Set $r := \min_{1\le i\le n} r_{t_i} > 0$. We claim that $V = B(\xi_0, t) \cap U$ is a desired open neighborhood of ξ_0. Indeed, for each $(t,\xi) \in [a,b] \times V$, by (1.9.6), we have $t \in B(t_i, r_{t_i})$ for some index i. On the other hand, for the same i, we also have $\xi \in V \subseteq B(\xi_0, r) \subseteq B(\xi_0, r_{t_i})$. It follows from (1.9.5) and (1.9.4) that

$$(t,\xi) \in B\Big((t,\xi_0), r_t\Big) \cap \big([a,b] \times U\big) \subseteq W, \tag{1.9.7}$$

which establishes $[a,b] \times V \subseteq W$. □

***Proposition 1.9.2* (Continuous dependence on the parameter)**

In the above setting, if $f : [a,b] \times U \to F$ is continuous, then the map $g : U \to F$, $\xi \mapsto g(\xi)$ is also continuous.

Proof. — Take any $\xi_0 \in U$, we wish to establish the continuity of g at ξ_0. Take an arbitrary $\varepsilon > 0$. By continuity of f, the set

$$W := \Big\{(t,\xi) \in [a,b] \times U \;\Big|\; \|f(t,\xi) - f(t,\xi_0)\| < \varepsilon\Big\} \tag{1.9.8}$$

is open in $[a,b] \times U$. Obviously, $[a,b] \times \{\xi_0\} \subseteq W$. Hence by Lemma 1.9.1, there exists an open neighborhood V of ξ_0, such that $[a,b] \times V \subseteq W$. Consequently,

$$\forall \xi \in V, \qquad \|g(\xi) - g(\xi_0)\| = \left\|\int_a^b \big(f(t,\xi_0) - f(t,\xi)\big)\mathrm{d}t\right\| \le \int_a^b \|f(t,\xi_0) - f(t,\xi)\|\mathrm{d}t \le \varepsilon(b-a). \tag{1.9.9}$$

Since V is an open neighborhood of ξ_0 and $\varepsilon > 0$ is arbitrary, (1.9.9) implies the desired continuity of g at ξ_0, which finishes the proof. □

With Proposition 1.9.2 as preparation, we now establish the following result concerning the validity of differentiation under the integral sign.

***Proposition 1.9.3* (Differentiation under the integral sign).** — *Using the above setting, assume further that $f(t,\xi)$ is partially differentiable with respect to ξ, and that*

1. *The map $f : [a,b] \times U \to F$ is continuous.*
2. *The partial derivative $\partial_2 f : [a,b] \times U \to \mathcal{L}(E,F)$ is continuous.*

Then g is differentiable and

$$\forall \xi \in U, \qquad \mathrm{d}g(\xi) = \int_a^b \partial_\xi f(t,\xi)\mathrm{d}t. \tag{1.9.10}$$

Proof. — Fix an arbitrary $\xi_0 \in U$. Since U is open, we may choose an $r > 0$ with $B(\xi_0, r) \subseteq U$. Given any $\varepsilon > 0$, it suffices to show that by shrinking $r > 0$ if necessary, we have

$$\forall \eta \in B(0,r) \subseteq E, \qquad \left\| g(\xi_0+\eta) - g(\xi_0) - \left(\int_a^b \partial_\xi f(t,\xi_0) \mathrm{d}t \right) \eta \right\| \leq \varepsilon(b-a)\|\eta\|. \tag{1.9.11}$$

By definition of g as well as Proposition 1.8.8, we have

$$g(\xi_0+\eta) - g(\xi_0) - \left(\int_a^b \partial_\xi f(t,\xi_0) \mathrm{d}t \right) \eta = \int_a^b \rho(t,\eta) \mathrm{d}t. \tag{1.9.12}$$

where

$$\rho(t,\eta) := f(t,\xi_0+\eta) - f(t,\xi_0) - \partial_2 f(t,\xi_0)\eta. \tag{1.9.13}$$

Apply the mean value theorem to $\alpha(s) := \rho(t, s\eta)$, we obtain

$$\begin{aligned} \|\rho(t,\eta)\| = \|\alpha(1)-\alpha(0)\| &\leq \sup_{s\in]0,1[} \left\| \Big(\partial_2 f(t,\xi_0+s\eta) - \partial_2 f(t,\xi_0) \Big) \eta \right\| \\ &\leq \|\eta\| \sup_{s\in]0,1[} \|\partial_2 f(t,\xi_0+s\eta) - \partial_2 f(t,\xi_0)\|. \end{aligned} \tag{1.9.14}$$

Now consider the map

$$\begin{aligned} \gamma : [a,b] \times B(0,r) &\to \mathcal{L}(E,F) \\ (t,\zeta) &\mapsto \partial_2 f(t,\xi_0+\zeta) - \partial_2 f(t,\xi_0). \end{aligned} \tag{1.9.15}$$

Then (2) implies that γ is continuous with $\gamma(\cdot,\zeta) = 0$. Hence $\gamma^{-1}\left(B_{\mathcal{L}(E,F)}(0,\varepsilon)\right)$ is an open subset of the metric space $[a,b] \times U$ containing $[a,b] \times \{0\}$. By Lemma 1.9.1, shrinking $r > 0$ if necessary, we may assume that

$$[a,b] \times B_E(0,r) \subseteq \gamma^{-1}\left(B_{\mathcal{L}(E,F)}(0,\varepsilon)\right). \tag{1.9.16}$$

It now follows from (1.9.16) and (1.9.16) that

$$\forall (t,\eta) \in [a,b] \times B(0,r), \qquad \|\rho(t,\eta)\| \leq \varepsilon\|\eta\|. \tag{1.9.17}$$

Hence

$$\forall \eta \in B(0,r), \qquad \left\| \int_a^b \rho(t,\eta) \mathrm{d}t \right\| \leq \int_a^b \|\rho(t,\eta)\| \mathrm{d}t \leq \varepsilon(b-a)\|\eta\|. \tag{1.9.18}$$

It follows from (1.9.18) and (1.9.18) that (1.9.11) holds, and the proof is complete. □

***Corollary 1.9.4* (Smooth dependence on the parameter).** — *Keep the same hypothesis as in Proposition 1.9.3 and assume further inductively that $\partial_2^{(0)} f = f$, and for each $k \in \mathbb{N}^+$, the map $\partial_2^{(k-1)} f : [a,b] \times U \to \mathcal{L}_{k-1}(E,F)$ (see Remark 1.5.5 for the latter notation) is partially differentiable with respect to the second variable, and $\partial_2^{(k)} f : [a,b] \times U \to \mathcal{L}_k(E,F)$ is continuous. Then g is smooth.*

Proof. — Just apply Proposition 1.9.3 repeatedly. □

1.10. Taylor's Formula and Local Extrema

Lemma 1.10.1. — *Let E, F, G be normed spaces, $I =]a,b[$ an open interval in $\mathbb{R}$ and $p \in \mathbb{N}^+$. Suppose $f \in C^p(I,E)$, $g \in C^p(I,F)$, and $[\cdot,\cdot] : E \times F \to G$ a bounded bilinear map. Then*

$$[f, \mathrm{d}^p g] - (-1)^p [\mathrm{d}^p f, g] = \mathrm{d}\left(\sum_{k=0}^{p-1} (-1)^k \left[\mathrm{d}^k f, \mathrm{d}^{p-1-k} g\right] \right). \tag{1.10.1}$$

Proof. — Note that by the chain rule and Example 1.3.3, for any $k \in \{1, \ldots, p-1\}$, we have

$$(-1)^k \mathrm{d}\left[\mathrm{d}^k f, \mathrm{d}^{p-1-k} g\right] = (-1)^k \mathrm{d}\left[\mathrm{d}^k f, \mathrm{d}^{p-k} g\right] - (-1)^k \left[\mathrm{d}^{k+1} f, \mathrm{d}^{p-1-k} g\right]. \tag{1.10.2}$$

Summing (1.10.2) over $k = 0, \ldots, p-1$ proves (1.10.1). □

Proposition 1.10.2. — *Let $I =]a,b[$ be an open interval, $t_0 \in]a,b[$ and E a Banach space. If $f \in C^p(I,E)$, then for any $t \in]a,b[$, we have*

$$f(t) = f(t_0) + \sum_{k=1}^{p-1} \frac{t^k}{k!} f^{(k)}(t_0)(t-t_0)^k + \int_{t_0}^{t} \frac{(t-s)^{p-1}}{(p-1)!} f^{(p)}(s)(t-t_0)^{(p-1)} \mathrm{d}s. \tag{1.10.3}$$

Proof. — Apply Lemma 1.10.1 by taking $g(s) = (-1)^p \frac{(t-s)^{p-1}}{(p-1)!}$ and $[\cdot,\cdot] : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ the multiplication $(a,b) \mapsto ab$, we obtain

$$\begin{aligned} \forall s \in]a,b[, \qquad & -\frac{(t-s)^{p-1}}{(p-1)!} f^{(p)}(s) \\ &= \left(\sum_{k=0}^{p-1} (-1)^{k+p+(p-1-k)} f^{(k)}(s) \frac{(t-s)^k}{k!} \right)' \\ &= -\left(\sum_{k=0}^{p-1} f^{(k)}(s) \frac{(t-s)^k}{k!} \right)', \end{aligned} \tag{1.10.4}$$

where all derivatives are taken against the variable s. Integrate (1.10.4) from t_0 to t and apply Corollary 1.8.10, we obtain (1.10.3). □

Definition 1.10.3. — Formula (1.10.3) is called **the Taylor formula** of f near t_0 of order $p-1$ with **integral remainder**. If f is smooth, we call the power series (on the variable t)

$$f(t_0) + \sum_{k=1}^{\infty} \frac{f^{(k)}(t_0)}{k!}(t-t_0)^k \tag{1.10.5}$$

the **Taylor series** of f at t_0.

Remark 1.10.4. — In general, the Taylor series does not converge, and even if it does, it does not necessarily converges to $f(t)$ for t near t_0. As a counter-example, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(t) = \begin{cases} 0, & \text{if } t \leq 0; \\ e^{-\frac{1}{t}}, & \text{if } t > 0. \end{cases} \tag{1.10.6}$$

We have seen in the exercise that f is smooth, and $f^{(k)}(0) = 0$ for any $k \in \mathbb{N}$, so the Taylor series for f at 0 is 0, nevertheless f is not the zero function on any neighborhood of 0.

Even though there are negative phenomenon as explained in Remark 1.10.4, the Taylor formula still enjoys some sort of "good approximation". We elaborate this point in a more general setting.

***Proposition 1.10.5* (General Taylor's Formula with the Integral Remainder)**

Let E be a normed space, F a Banach spaces, $U \subseteq E$ open, $x_0 \in U$ and $f \in C^p(U,F)$ for some $p \in \mathbb{N}^+$. Then for any $r > 0$ with $B(x_0,r) \subseteq U$, we have for any $\xi \in B(0,r)$, that

$$f(x_0+\xi) = f(x_0) + \sum_{k=1}^{p-1} \frac{1}{k!} \mathrm{d}^k f(x_0).\xi^{(k)} + \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} \mathrm{d}^p f(x_0 + t(x-x_0)).\xi^{(p)} \mathrm{d}t, \tag{1.10.7}$$

where $\mathrm{d}^k f(x_0)$ is regarded as a symmetric k-linear map from E^k to F as explained in Proposition 1.5.6, and $\xi^{(k)}$ denotes the vector in E^k with each component being ξ. In particular, for any $\varepsilon > 0$, there exists $r(\varepsilon) > 0$, such that

$$\forall \xi \in B(0,r(\varepsilon)), \qquad \left\| f(x_0+\xi) - \left(f(x_0) + \sum_{k=1}^{p} \frac{1}{k!} \mathrm{d}^k f(x_0).\xi^{(k)} \right) \right\| \leq \varepsilon \|\xi\|^p. \tag{1.10.8}$$

Proof. — Formula (1.10.7) follows from Proposition 1.10.2 by considering the map $g(t) = f(x_0 + t\xi)$ defined on $[0,1]$. Note that

$$\int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} \mathrm{d}^p f(x_0).\xi^{(p)} \mathrm{d}t = \frac{1}{p!} \mathrm{d}^p f(x_0).\xi^{(p)} \tag{1.10.9}$$

by Corollary 1.8.10. It follows from (1.10.7) that

$$
\begin{aligned}
\forall \xi \in B(0,r), \qquad & f(x_0+\xi) - \sum_{k=0}^{p} \frac{1}{k!}\mathrm{d}^k f(x_0).\xi^{(k)} \\
&= \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\Big[\mathrm{d}^p f\big(x_0+t\xi\big) - \mathrm{d}^p f(x_0)\Big].\xi^{(p)}\mathrm{d}t.
\end{aligned}
\tag{1.10.10}
$$

Since $f \in C^p(U,E)$, there exists $r(\varepsilon) \in]0,r[$ such that

$$
\forall \eta \in B(0,r(\varepsilon)), \qquad \|d^p f(x_0+\eta) - \mathrm{d}^p f(x_0)\| \leq p!\varepsilon.
\tag{1.10.11}
$$

Thus it follows from (1.10.10), (1.10.11) and Proposition 1.8.7 that

$$
\begin{aligned}
\forall \xi \in B(0,r(\varepsilon)), \qquad & \left\| f(x_0+\xi) - \sum_{k=0}^{p} \frac{1}{k!}\mathrm{d}^k f(x_0).\xi^{(k)} \right\| \\
&\leq \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\|d^p f(x_0+t\xi) - \mathrm{d}^p f(x_0)\| \cdot \|\xi\|^p \mathrm{d}t \\
&\leq p!\varepsilon\|\xi\|^p \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\mathrm{d}t = \varepsilon\|\xi\|^p,
\end{aligned}
\tag{1.10.12}
$$

which finishes the proof. ☐

As an application of the Taylor formula, we now turn to the study of local extrema of functions.

Definition 1.10.6. — Let E be a normed space, $U \subseteq E$ open, $x_0 \in U$ and $f : U \to \mathbb{R}$ a function. We say that f achieves its **local maximum**(resp. **local minimum**) at x_0, if there exists a neighborhood V of x_0 in U, such that $f(x) \leq f(x_0)$ (resp. $f(x) \geq f(x_0)$) for any $x \in V$.

The following result is natural as we are already familiar with its one-dimensional case.

Proposition 1.10.7. — *Let E be a normed space, $U \subseteq E$ open, $x_0 \in U$ and $f : U \to \mathbb{R}$ a function that achieves its local maximum (resp. minimum) at x_0. For any $\xi \in E$, if the directional derivative of f at x_0 along ξ exists, then this directional derivative is 0. In particular, if $\mathrm{d}f(x_0) \in \mathcal{L}(E,\mathbb{R}) = E'$ exists, then $\mathrm{d}f(x_0) = 0$.*

Proof. — Just apply the one-dimensional result to $\alpha(t) := f(x_0+t\xi)$ defined in a neighborhood of 0: note that α achieves its local maximum (resp. minimum) at 0, hence $\alpha'(0)$, which exists since it is exactly the directional derivative of f at x_0 along the direction ξ, must be 0. The rest follows from Proposition 1.2.16. ☐

Proposition 1.10.7 gives a necessary condition for a function to achieve its local extrema. In general, it is however, not sufficient, as demonstrated by the one-dimensional example $f(x) = x^3$, $x \in \mathbb{R}$ near the point $x = 0$. Using our general version of Taylor's formula, we may derive a sufficient condition as follows.

Proposition 1.10.8. — *Let E be a normed space, $U \subseteq E$ open, $x_0 \in U$ and $f : U \to \mathbb{R}$ a C^2-function.*

(1) *If $\mathrm{d}f(x_0) = 0$ and*

$$
\inf_{\xi \in E,\, \|\xi\|=1} \mathrm{d}^2 f(x_0).\xi^{(2)} > 0,
\tag{1.10.13}
$$

then f achieves its local maximum at x_0, and there exists a neighborhood $V \subseteq U$ of x_0 such that $f^{-1}\big(f(x_0) = \{x_0\}\big)$.

(2) *If $\mathrm{d}f(x_0) = 0$ and*

$$
\sup_{\xi \in E,\, \|\xi\|=1} \mathrm{d}^2 f(x_0).\xi^{(2)} < 0,
\tag{1.10.14}
$$

then f achieves its local maximum at x_0, and there exists a neighborhood $V \subseteq U$ of x_0 such that $f^{-1}\big(f(x_0) = \{x_0\}\big)$.

Proof. — We only prove (1), as (2) then follows by symmetry or simply by considering $-f$. Set

$$
c := \min_{\xi \in E,\, \|\xi\|=1} \mathrm{d}^2 f(x_0).\xi^{(2)} > 0.
\tag{1.10.15}
$$

Then by Proposition 1.10.5, we see that for each $\varepsilon > 0$, there exists $r(\varepsilon) > 0$, such that $B(x_0, r(\varepsilon)) \subseteq U$ and

$$\xi \in B(0, r(\varepsilon)), \qquad \left| f(x_0 + \xi) - \mathrm{d}f(x_0).\xi - \frac{1}{2}\mathrm{d}^2 f(x_0).\xi^{(2)} \right| \leq \varepsilon \|\xi\|^2. \tag{1.10.16}$$

Since $\mathrm{d}f(x_0) = 0$, it follows from (1.10.16) that

$$\begin{aligned} \forall \xi \in B(0, r(\varepsilon)) \setminus \{0\}, \qquad f(x_0 + \xi) &\geq f(x_0) + \frac{1}{2}\mathrm{d}^2 f(x_0).\xi^{(2)} - \varepsilon \|\xi\|^2 \\ &\geq f(x_0) + \|\xi\|^2 \left(\frac{1}{2}\mathrm{d}^2 f(x_0). \left(\frac{\xi}{\|\xi\|} \right)^{(2)} - \varepsilon \right) \\ &\geq f(x_0) + \|\xi\|^2 \left(\frac{c}{2} - \varepsilon \right). \end{aligned} \tag{1.10.17}$$

Hence it suffices to take $V = B(x_0, r(\varepsilon))$ for any positive $\varepsilon < \frac{c}{2}$. $\square$

Definition 1.10.9. — In Proposition 1.10.8, when $E = \mathbb{R}^n$, it is easy to see that

$$\forall i, j \in \{1, \ldots, n\}, \qquad \mathrm{d}^2 f(x_0).(e_i, e_j) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0),$$

hence

$$\forall x, y \in \mathbb{R}^n, \qquad \mathrm{d}^2 f(x_0)(x, y) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) x_i y_j. \tag{1.10.18}$$

The symmetric (why?) matrix $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \right)_{1 \leq i,j \leq n}$ is called the **Hessian** (or **Hessian matrix**) of f at x_0, and is often denoted by $\operatorname{Hess} f(x_0)$ or simply by $Hf(x_0)$.

Remark 1.10.10. — It follows from Proposition 1.10.5 that a C^2 function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ can be locally approximated up to order 2 using the gradient and the Hessian of f, namely,

$$f(x_0 + \xi) = f(x_0) + \nabla f(x_0).\xi^\intercal + \xi Hf(x_0)\xi^\intercal + o(\|\xi\|^2), \tag{1.10.19}$$

where vectors in $\mathbb{R}^n$ are regarded as row vectors, and $x_0 \in U$, $\xi \in B(0, r)$ with $B(0, r) \subseteq U$ for some $r > 0$. This observation is one of the key idea in Morse Lemma (see § 1.15), which is the starting point of a beautiful branch of mathematics known as *the Morse theory.*

Corollary 1.10.11. — *Let $n \in \mathbb{N}^+$, $U \subseteq \mathbb{R}^n$, $x_0 \in U$ and $f \in C^2(U, \mathbb{R})$. If $\nabla f(x_0) = 0$ and if the Hessian matrix $Hf(x_0)$ is strictly positive (resp. strictly negative) definite, then there exists a neighborhood $V \subseteq U$ of x_0 such that x_0 is the unique point in V at which f assumes its minimum (resp. maximum) value on V.*

Proof. — Since the unit $(n-1)$-sphere $S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is compact in $\mathbb{R}^n$, the continuous map $\alpha : S^{n-1} \to \mathbb{R}$, $\xi \mapsto \mathrm{d}^2 f(x_0).\xi^{(2)} = \xi.Hf(x_0).\xi^\intercal$ assumes its minimum $\alpha(\xi_0)$ at some $\xi_0 \in S^{n-1}$. Since the Hessian matrix $Hf(x_0)$ is strictly positive (resp. strictly negative) definite, the corollary now follows from Proposition 1.10.8. $\square$

Remark 1.10.12. — More generally, in the setting of Definition 1.10.9 with $f \in C^p(U, \mathbb{R})$, for a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $|\alpha| \leq p$, using the multi-index notation (Notation 1.7.10), by an easy induction argument, we have

$$\mathrm{d}^{|\alpha|} f(x_0)(e_1, \ldots, e_1, \ldots, e_n, \ldots, e_n) = \partial_\alpha f(x_0), \tag{1.10.20}$$

where each e_i appears exactly α_i times on the left side of (1.10.20). In this case, it is easy to see (again by induction for example) that

$$\forall \xi = (\xi_i) \in \mathbb{R}^n, \qquad \mathrm{d}^k f(x_0).\xi^{(k)} = \sum_{\alpha,\, |\alpha| = k} \frac{k!}{\alpha!} \partial_\alpha f(x_0) x^\alpha, \tag{1.10.21}$$

where the sum is taken over all multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $|\alpha| = k$ and $\alpha! := \alpha_1! \cdots \alpha_n!$. The coefficient $\frac{k!}{\alpha!}$ has a combinatorial meaning, it is the coefficient of x^α in the expansion of $(x_1 + \cdots + x_n)^k$. We often rewrite Taylor's formula (1.10.7) using (1.10.21) in the case where $E = \mathbb{R}^n$ and $F = \mathbb{R}$.

1.11. The Contraction Mapping Principle

Definition 1.11.1. — A map $f : X \to X$ from a metric space (X, d) to itself is called a **contraction**, if there exists $k \in \,]0, 1[$, such that

$$\forall x_1, x_2 \in X, \qquad d\big(f(x_1), f(x_2)\big) \le kd(x_1, x_2). \tag{1.11.1}$$

It is obvious that each contraction is continuous.

Remark 1.11.2. — We will be mainly interested in the fixed point of a contraction. From the definition, it is immediate that the fixed point of a contraction f must be unique if it exists. Indeed, if $f(x_i) = x_i$, $i = 1, 2$, then

$$d(x_1, x_2) = d(f(x_1), f(x_2)) \le kd(x_1, x_2), \tag{1.11.2}$$

which forces $d(x_1, x_2) = 0$, hence $x_1 = x_2$.

The following result provides a key technical tool in establishing the inverse and implicit function theorem.

***Proposition 1.11.3* (The Contraction Mapping Principle).** — *Let (X, d) be a* complete[6] *metric space, and $f : X \to X$ a contraction, then f admits a unique fixed point.*

Proof. — The uniqueness part is already seen in Remark 1.11.2. To see the existence part, take any $x_0 \in X$, define a sequence (x_n) in X inductively by letting $x_n = f(x_{n-1})$ for all $n \in \mathbb{N}^+$. Then we have, for each $n \in \mathbb{N}^+$, that

$$k^{-(n+1)} d(x_{n+1}, x_{n+2}) = k^{-(n+1)} d(f(x_n), f(x_{n+1})) \le k^{-n} d(x_n, x_{n+1}) \le \cdots \le d(x_0, x_1), \tag{1.11.3}$$

thus

$$d(x_n, x_{n+1}) \le k^n d(x_0, x_1). \tag{1.11.4}$$

From (1.11.4) and the triangle inequality, we have the estimation

$$d(x_n, x_{n+m}) \le \sum_{i=1}^{m-1} d(x_{n+i-1}, x_{n+i}) \le \sum_{i=1}^{m-1} k^{n+i-1} d(x_0, x_1) \le \frac{k^n}{1-k} d(x_0, x_1), \tag{1.11.5}$$

where $k \in \,]0, 1[$ is a constant such that (1.11.1) holds. It follows from (1.11.5) that (x_n) is a Cauchy sequence in (X, d), hence convergent by completeness. Set $x_\infty = \lim_{n\to\infty} x_n \in X$, by the continuity of f, we have

$$x_\infty = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right) = f(x_\infty), \tag{1.11.6}$$

and x_∞ is a (hence the unique) fixed point of f. □

1.12. The Inverse and Implicit Mapping Theorem

***Theorem 1.12.1* (Inverse Mapping Theorem (IMT), aka. Inverse Function Theorem (IFT))**
Let E, F be Banach spaces, $U \subseteq E$ open, $x_0 \in U$, and $f : U \to F$ a map of class C^p ($p \in \mathbb{N}^+ \cup \{\infty\}$). If $A_0 := \mathrm{d}f(x_0)$ is an isomorphism from E onto F, then there exists an open neighborhood V of x_0 in U, such that

1. *$W := f(V)$ is open in F and contains $y_0 := f(x_0)$;*
2. *the restriction of f is a bijection from V onto W;*
3. *the inverse map $g : W \to V$, $f(x) \to x$ is also of class C^p;*
4. $\mathrm{d}g(y_0) = A_0^{-1}$.

[6] We tacitly require that complete metric space must be non-empty.

Proof. — Let $r > 0$ be so small that $B(x_0, 2r) \subseteq U$, and

$$\forall x \in B(x_0, r), \qquad \|\mathrm{d}f(x) - A_0\| \le \frac{1}{2}\|A_0^{-1}\|^{-1}, \tag{1.12.1}$$

which is possible since $\mathrm{d}f$ is continuous. Set $s = (2\|A_0\|)^{-1}r$ and $W_0 := B(y_0, s)$. For each $y \in B(y_0, s)$, define (see Remark 1.12.2 for the motivation)

$$\begin{aligned} S_y : U &\to F \\ x &\mapsto x + A_0^{-1}\big(y - f(x)\big). \end{aligned} \tag{1.12.2}$$

Clearly S_y is of class C^p, and $S_y(x) = x$ if and only if $y = f(x)$. Moreover, by (1.12.1),

$$\begin{aligned} \forall x \in B(x_0, r), \qquad \|\mathrm{d}S_y(x)\| &= \big\|\mathrm{Id} - A_0^{-1}\mathrm{d}f(x)\big\| = \big\|A_0^{-1}(A_0 - \mathrm{d}f(x))\big\| \\ &\le \big\|A_0^{-1}\big\| \cdot \|A_0 - \mathrm{d}f(x)\| \le \frac{1}{2}. \end{aligned} \tag{1.12.3}$$

We claim that S_y is contraction from $\overline{B(x_0, r)}$ to itself for any $y \in W_0$. Indeed, for all $x_1, x_2 \in \overline{B(x_0, r)}$ with $x_1 \neq x_2$, we have $x_1 + t(x_2 - x_1) \in B(x_0, r)$ for all $t \in]0, 1[$, hence by setting $\alpha(t) = S_y(x_1 + t(x_2 - x_1))$, $t \in [0, 1]$ and apply the mean value theorem, we obtain

$$\begin{aligned} \forall y \in W_0, \qquad \|S_y(x_1) - S_y(x_2)\| &= \|\alpha(1) - \alpha(0)\| \le \sup_{t \in]0,1[} \|\alpha'(t)\| \\ &= \sup_{t \in]0,1[} \|\mathrm{d}S_y(x_1 + t(x_2 - x_1)).(x_2 - x_1)\| \\ &\le \|x_2 - x_1\| \cdot \sup_{t \in]0,1[} \|\mathrm{d}S_y(x_1 + t(x_2 - x_1))\| \\ &\le \frac{1}{2}\|x_2 - x_1\|, \end{aligned} \tag{1.12.4}$$

where the last inequality follows from (1.12.3). On the other hand, for each $x \in \overline{B(x_0, r)}$, we have

$$\begin{aligned} \forall y \in B(y_0, s) \qquad \|S_y(x) - x_0\| &= \big\|S_y(x) - S_y(x_0) + A_0^{-1}(y - y_0)\big\| \\ &\le \|S_y(x) - S_y(x_0)\| + \big\|A_0^{-1}\big\| \cdot \|y - y_0\| \\ &\le \frac{1}{2}\|x - x_0\| + s\big\|A_0^{-1}\big\| \\ &\le \frac{r}{2} + \frac{r}{2} = r. \end{aligned} \tag{1.12.5}$$

Our claim that for any $y \in W_0$, the map S_y is a contraction from $\overline{B(x_0, r)}$ to itself now follows from (1.12.4) and (1.12.5). Now by the contraction mapping principle (Proposition 1.11.3), for each $y \in W_0$, there exists a unique $x \in \overline{B(x_0, r)}$, such that $S_y(x) = x$, and we define this x to be $g(y)$. Since f is continuous, the inverse image $V_0 := f^{-1}(W_0)$ is an open set. We now set $V := V_0 \cap B(x_0, r)$ and $W := f(V)$, then V is open and $x_0 \in V$. To see that W is also open is slightly more involved. First, note that $\mathrm{d}f(x)$ is invertible for each $x \in B(x_0, r)$, since

$$\mathrm{d}f(x) = A_0 - (A_0 - \mathrm{d}f(x)) = A_0\Big(\mathrm{Id} - A_0^{-1}(A_0 - \mathrm{d}f(x))\Big) \tag{1.12.6}$$

with $\big\|A_0^{-1}(A_0 - \mathrm{d}f(x))\big\| \le \frac{1}{2}$ (see (1.12.3), and Example 1.3.5). The above reasoning shows that

$$f\big(B(x_0, 2r)\big) \supseteq f\big(\overline{B(x_0, r)}\big) \supseteq W, \tag{1.12.7}$$

hence, we've established the fact that if $\mathrm{d}f(x_0)$ is invertible, then there exists a neighborhood of x_0 whose image under f is a neighborhood of $f(x_0)$. Apply this fact to $f|_{B(x_0,r)}$, for which the derivative is everywhere invertible, and we see that $f : B(x_0, r) \to F$ is an open map. In particular, $W = f(V)$ is open and $x_0 \in V$ implies $y_0 = f(x_0) \in W$. Clearly, the restriction $f : W \to V$ is bijective and the inverse $g : V \to W$ is continuous. So far, we have established (1) and (2).

To finish the proof, it suffices to establish (4), since then (3) follows from the chain rule and the smoothness of the inverse map $\nu : \mathrm{Iso}(E, F) \to \mathrm{Iso}(F, E)$ in Example 1.3.5 (see Example 1.5.8). For all

$y \in W$, we have, by definition of g that $S_y(g(y)) = g(y)$. Hence

$$
\begin{aligned}
g(y) - g(y_0) &= S_y(g(y)) - S_{y_0}(g(y)) + S_{y_0}(g(y)) - S_{y_0}(g(y_0)) \\
&= A_0^{-1}(y - y_0) + \Big(S_{y_0}(g(y)) - S_{y_0}(g(y_0))\Big),
\end{aligned} \tag{1.12.8}
$$

hence

$$
\forall y \in W, \qquad \big\|g(y) - g(y_0) - A_0^{-1}(y - y_0)\big\| = \|S_{y_0}(g(y)) - S_{y_0}(g(y_0))\| \le \frac{1}{2}\|g(y) - g(y_0)\|, \tag{1.12.9}
$$

thus

$$
\begin{aligned}
\forall y \in W, \qquad \big\|A_0^{-1}\big\| \cdot \|y - y_0\| &\ge \big\|A_0^{-1}(y - y_0)\big\| \\
&\ge \|g(y) - g(y_0)\| - \big\|g(y) - g(y_0) - A_0^{-1}(y - y_0)\big\| \\
&\ge \frac{1}{2}\|g(y) - g(y_0)\|,
\end{aligned} \tag{1.12.10}
$$

or equivalently[(7)],

$$
\forall y \in W, \qquad \|g(y) - g(y_0)\| \le 2\big\|A_0^{-1}\big\| \cdot \|y - y_0\|. \tag{1.12.11}
$$

Now observe that

$$
\forall y \in W, \qquad S_{y_0}(g(y)) - S_{y_0}(g(y_0)) = \Big(g(y) - A_0^{-1}f(g(y))\Big) - \Big(g(y_0) - A_0^{-1}f(g(y_0))\Big). \tag{1.12.12}
$$

Note $g(W) = V$, (1.12.12) leads us to consider the map

$$
\begin{aligned}
h : V &\to E \\
x &\mapsto x - A_0^{-1}f(x),
\end{aligned} \tag{1.12.13}
$$

which is clearly C^p with $\mathrm{d}h(x_0) = 0$. Now take any $\varepsilon > 0$. From the continuity of $\mathrm{d}h$, we may find a small open ball $V_\varepsilon := B(x_0, r_\varepsilon)$ inside V such that

$$
x \in V_\varepsilon \implies \|\mathrm{d}h(x)\| \le \varepsilon. \tag{1.12.14}
$$

By Corollary 1.4.3, we see that

$$
\forall x_\varepsilon, x_2 \in V_\varepsilon \subseteq V, \qquad \|h(x_\varepsilon) - h(x_2)\| \le \varepsilon\|x_\varepsilon - x_2\|. \tag{1.12.15}
$$

Set $W_\varepsilon := g^{-1}(V_\varepsilon)$, which is an open neighborhood of y_0 since g is continuous and $y_0 = g(x_0)$, by (1.12.11), (1.12.12), (1.12.15) and also (1.12.9), we have

$$
\begin{aligned}
\forall y \in W_\varepsilon, \qquad \big\|g(y) - g(y_0) - A_0^{-1}(y - y_0)\big\| &= \|S_{y_0}(g(y)) - S_{y_0}(g(y_0))\| \\
&= \|h(g(y)) - h(g(y_0))\| \\
&\le \varepsilon\|g(y) - g(y_0)\| \\
&\le 2\varepsilon\big\|A_0^{-1}\big\| \cdot \|y - y_0\|,
\end{aligned} \tag{1.12.16}
$$

which implies (4) since ε is arbitrary. □

Remark 1.12.2. — We now explain why we define S_y in the form of (1.12.2). The problem we wish to settle at this stage is that for any y near y_0, we want to find an x near x_0 that solves the equation $f(x_y) = y$. Imagine we don't have a precise value for such an x_y but we do have a good guess, say x, then the difference $y - f(x)$ is small, in which case, we know this difference should be roughly $\mathrm{d}f(x).(x_y - x)$. On the other hand, $\mathrm{d}f(x)$ should be close to $A_0 = \mathrm{d}f(x_0)$ since $\mathrm{d}f$ is continuous. We thus have

$$
y - f(x) \sim A_0(x_y - x) \tag{1.12.17}
$$

hence

$$
x_y \sim x + A_0^{-1}(y - f(x)) \tag{1.12.18}
$$

which means, by defining $S_y(x)$ to be the right side of (1.12.18), we should obtain a better guess $S_y(x)$ that is closer to x_y, and the precise value of x_y should make the roughly equal $\sim$ become strictly equal,

---

(7) Note that (1.12.11) provides another proof of the continuity of g at points $f(x)$ where $\mathrm{d}f(x)$ is invertible.

i.e. x_y should be the fixed point of S_y. The rest now easily amounts to the contraction mapping principle. The informed reader may find that this is actually a variant of Newton's tangent method for finding the root of a function, for which the main idea is that finding better approximation of a root using the given data and some approximate root, with the help of derivatives, since locally, the function should be roughly affine.

The following notion is of fundamental importance in differential topology.

Definition 1.12.3. — Let $U \subseteq E$ (resp. $V \subseteq F$) be open with E (resp. F) being a normed space. A map $f : U \to V$ is called a C^p **(**$p \in \mathbb{N}^+$**)-diffeomorphism** if f is bijective, of class C^p and the inverse of f is also of class C^p as a map from V to U. We say f is a **smooth diffeomorphism**, or simply, a **diffeomorphism**, if f is a C^p-diffeomorphism for all $p \in \mathbb{N}^+$. We say U and V are C^p**-diffeomorphic** (resp. **diffeomorphic**) if there is a C^p-diffeomorphism (resp. diffeomorphism) between them.

Remark 1.12.4. — In terms of the terminology of Definition 1.12.3, we see that the inverse mapping theorem provides a way to obtain a *local* (C^p-)diffeomorphism.

***Corollary 1.12.5* (Open mapping theorem).** — *Keep the setting as in Theorem 1.12.1, and set*

$$U^* := \Big\{ x \in U \;\Big|\; \mathrm{d}f(x) \in \mathrm{Iso}(E, F) \Big\}, \tag{1.12.19}$$

then both U^* *and* $f(U^*)$ *are open.*

Proof. — Since $\mathrm{d}f : U \to \mathcal{L}(E, F)$ is continuous and $\mathrm{Iso}(E, F)$ is open in $\mathcal{L}(E, F)$ (see Example 1.3.5), we have

$$U^* = (\mathrm{d}f)^{-1}\Big(\mathrm{Iso}(E, F)\Big) \tag{1.12.20}$$

is open. Now take any $y \in f(U^*)$. There is some $x \in U^*$ with $f(x) = y$. By the inverse mapping theorem applied to $f \restriction_{U^*}$, there exists an neighborhood V of x and an open neighborhood W of y with $V \subseteq U^*$ such that $W = f(V)$. In particular, $f(U^*) \supseteq f(V) = W \ni y$ and $f(U^*)$ is a neighborhood of each of its point, hence open. □

Definition 1.12.6. — Let X, Y be topological spaces[(8)], and $f : X \to Y$ a map. We say that f is **locally open** at $x_0 \in X$, if there exists a neighborhood U of x_0, such that the restriction $f \restriction_{U_0} : U \to Y$ is an open map. Equivalently, f is locally open at x_0 if $f(U_0)$ is open for any open neighborhood U_0 that are contained in a fixed neighborhood of x_0.

Using Definition 1.12.6, we immediately obtain the following corollary of the inverse mapping theorem.

Corollary 1.12.7. — *Keep the setting as in Theorem 1.12.1, then f is locally open at any point in U on which its derivative is invertible.* □

Remark 1.12.8. — For finite dimensional spaces, there is a much stronger result called the **invariance of domain** due to Brower. The result says that if $U \subseteq \mathbb{R}^n (n \in \mathbb{N}^+)$ is open and $f : U \to \mathbb{R}^n$ is continuous and injective, then f is an open map. Note that how this result does not need the differentiability of f unlike in Corollary 1.12.7. However, it will be seen in the exercise that our Corollary 1.12.7 is already strong enough to allow a proof of the fundamental theorem of algebra.

On the other hand, Brower's invariance of domain essentially depends on the finite dimensionality and fails immediately for infinite dimensional spaces. As a counter-example, consider any isometry on ℓ^2 that are not an isomorphism, e.g. $f \in \mathcal{L}(\ell^2)$ defined by

$$f : (a_n) \mapsto \left(\frac{1 + (-1)^n}{2} a_{\lfloor n/2 \rfloor} \right). \tag{1.12.21}$$

The problem in this case is that $\mathrm{d}f = f$ (since f in (1.12.21) is bounded linear) is merely injective but not surjective at each point (compare with Theorem 1.12.1).

With the help of the inverse mapping theorem, the following implicit mapping theorem can now be easily established.

(8) If you haven't seen the basic notions in general topologies, just take X and Y to be metric spaces.

Theorem 1.12.9 **(Implicit Mapping Theorem (IMT), aka. Implicit Function Theorem (IFT))**
Let E, F and G be Banach spaces, $\mathcal{O} \subseteq E \times F$ an open set and $f : \mathcal{O} \to G$ a C^p map for some fixed $p \in \mathbb{N}^+ \cup \{\infty\}$. Suppose that at some point $(x_0, y_0) \in \mathcal{O} \subseteq E \times F$, we have $\partial_2 f(x_0, y_0) \in \operatorname{Iso}(F, G)$. Set $z_0 = f(x_0, y_0)$. Then there exists an open neighborhood U_0 of x_0 in E, an open neighborhood V_0 of y_0 in F, such that

(1) *For each $x \in U_0$, there exists a unique point in V_0, denoted e.g. by $\varphi(x)$, such that $f(x, \varphi(x)) = z_0$.*
(2) *The map $\varphi : U_0 \to V_0$ as in (1) is of class C^p.*
(3) *We have $\partial_2 f(x, \varphi(x)) \in \operatorname{Iso}(F, G)$ for all $x \in U_0$ and*

$$\mathrm{d}\varphi(x) = -[\partial_2 f(x, \varphi(x))]^{-1} \partial_1 f(x, \varphi(x)). \tag{1.12.22}$$

Proof. — Consider the map $\Phi = (\Phi_1, \Phi_2) : \mathcal{O} \to E \times G$ defined by $\Phi_1 = \pi_E \restriction_{\mathcal{O}} : \mathcal{O} \to E$ (the restriction of the canonical projection $\pi_E : E \times F \to E$) and $\Phi_2 = f$. Set

$$A := \partial_2 f(x_0, y_0) = \partial_2 \Phi_2(x_0, y_0) \in \operatorname{Iso}(F, G) \subseteq \mathcal{L}(F, G) \tag{1.12.23}$$

and

$$B := \partial_1 f(x_0, y_0) = \partial_1 \Phi_2(x_0, y_0) \in \mathcal{L}(E, F). \tag{1.12.24}$$

Then

$$\forall (\xi, \eta) \in E \times F, \qquad \mathrm{d}\Phi(x_0, y_0)(\xi, \eta) = (\xi, B\xi + A\eta) \in E \times G.$$

It is easily checked that the linear map

$$\begin{aligned} E \times F &\to E \times G \\ (\xi, \eta) &\mapsto (\xi, A^{-1}(\eta - B\xi)) \end{aligned} \tag{1.12.25}$$

is bounded and is the inverse of $\mathrm{d}\Phi(x_0, y_0)$, i.e. $\mathrm{d}\Phi(x_0, y_0) \in \operatorname{Iso}(E \times F, E \times G)$. We may thus apply the inverse mapping theorem (Theorem 1.12.1) to Φ at (x_0, y_0), and obtain Φ as a local C^p-diffeomorphism from an open neighborhood of (x_0, y_0) onto an open neighborhood of $(x_0, z_0) = \Phi(x_0, y_0)$. By shrinking the former open neighborhood if necessary, we may find an open neighborhood U_0 (resp. V_0) of x_0 (resp. y_0), such that $\Phi \restriction_{U_0 \times V_0}$ is a C^p-diffeomorphism onto its image, which is open in $E \times G$. Hence

$$\forall (x, y) \in \mathcal{O}_0 := U_0 \times V_0, \qquad f(x, y) = z_0 \iff \Phi(x, y) = (x, z_0) \iff y = \Phi \restriction_{\mathcal{O}_0}^{-1} (x, z_0). \tag{1.12.26}$$

Thus (1) holds with $\varphi(x) = \Phi \restriction_{\mathcal{O}_0}^{-1} (x, z_0)$, which implies (2). Clearly,

$$\forall (x, y) \in \mathcal{O}, \qquad (\xi, \eta) \in E \times F \qquad \mathrm{d}\Phi(x, y).(\xi, \eta) = (\xi, \partial_1 f(x, y).\xi + \partial_2 f(x, y).\eta)\,. \tag{1.12.27}$$

A routine manipulation of linear algebra shows that $\mathrm{d}\Phi(x, y)$ is invertible if and only if $\partial_2 f(x, y)$ is so. This implies the first part of (3), and (1.12.22) follows by differentiating the equation $\Phi(x, \varphi(x)) = z_0$. $\square$

Remark 1.12.10. — In the exercise session, we will see that one can also independently establish the implicit mapping theorem first via a variant of the contraction mapping principle and a similar consideration as in Remark 1.12.2, then obtains the inverse mapping theorem from the implicit mapping theorem. Adopting such an approach, we may obtain a slightly stronger version of Theorem 1.12.9 which drops the completeness of the normed space E.

1.13. The Rank Theorem

The rank theorem traditionally concerns differential calculus on finite dimensional spaces, and refers to some results of the Jacobian matrix being of a constant rank. In keeping with the same spirit of doing differential calculus on general Banach spaces, we adopt an alternate, more general perspective which makes the essence of the proof more transparent.

As preparation, we first point out the correspondence between idempotents and direct product/sum decomposition. Given a normed space E. We say $P \in \mathcal{L}(E)$ is an **idempotent**, if $P^2 = P$. Idempotents are also known as **projections** in the literature. We denote by $\operatorname{Idem}(E)$ the set of idempotents in $\mathcal{L}(E)$.

Given a $P \in \mathrm{Idem}(E)$, we may decompose E into two parts: the kernel $\mathcal{N}(P)$ and the range $\mathcal{R}(P)$ of P. It is clear that $P_{\mathcal{N}(P)} = 0$ and that $P_{\mathcal{R}(P)}$ acts as identity due to idempotency. It follows that $\mathcal{N}(P) \cap \mathcal{R}(P) = 0$. On the other hand, for any $\xi \in E$, we have

$$\xi = (\mathrm{Id} - P)\xi + P\xi \tag{1.13.1}$$

with $(\mathrm{Id} - P)\xi \in \mathcal{N}(P)$ and $P\xi \in \mathcal{R}(P)$. Hence as vector spaces, we have the decomposition

$$E = \mathcal{N}(P) \oplus \mathcal{R}(P) \simeq \mathcal{N}(P) \times \mathcal{R}(P), \tag{1.13.2}$$

where the canonical isomorphism of vector spaces is given by $\xi + \eta \leftrightarrow (\xi, \eta)$. The vector space isomorphism $E \simeq \mathcal{N}(P) \times \mathcal{R}(P)$ is actually an isomorphism of normed spaces, since both $(\xi, \eta) \in \mathcal{N}(P) \times \mathcal{R}(P) \mapsto \xi + \eta \in E$ and its inverse $x \in E \mapsto (Px, x - Px) \in \mathcal{N}(P) \times \mathcal{R}(P)$ are continuous linear maps.

Being subspaces of E, we know that both $\mathcal{N}(P)$ and $\mathcal{R}(P)$ are normed. It is interesting to note that these are actually closed subspaces of E, hence are Banach spaces themselves provided E is a Banach space. Indeed, $\mathcal{N}(P) = P^{-1}(0)$ and the singleton $\{0\}$ is closed, which shows the closedness of $\mathcal{N}(P)$ by continuity of P. Similarly, one checks immediately, using $P^2 = P$, that $\mathcal{R}(P) = \mathcal{N}(\mathrm{Id} - P) = (\mathrm{Id} - P)^{-1}(0)$, hence is closed by continuity of $\mathrm{Id} - P$.

In summary, if we identify E with $\mathcal{N}(P) \times \mathcal{R}(P)$ via (1.13.2), then the idempotent P is nothing but the projection onto the first component. Conversely, if we have a decomposition $E = E_1 \times E_2$, and we identify E_1 (resp. E_2) with the subspace $E_1 \times \{0\}$ (resp. $\{0\} \times E_2$) of E via $x_1 \leftrightarrow (x_1, 0)$ (resp. $x_2 \leftrightarrow (x_2, 0)$), then the projection onto the first component takes the form $\pi_1 : (x_1, x_2) \in E \to (x_1, 0) \in E$, and we immediately have $\pi_1 \in \mathrm{Idem}(E)$ with $\mathcal{N}(\pi_1) = E_2$ and $\mathcal{R}(\pi_1) = E_1$.

With the above preparation ready, we may now motivate our seemingly involved Theorem 1.13.1. The idea is that we want to characterize maps that can be "straighten out" via a change of variable. Here by the vague of saying of "straighten out", we mean something that is very close to a bounded linear operator. On the other hand, the graph of a map f is always parameterized by its domain, via $x \in \mathrm{dom}(f) \leftrightarrow (x, f(x)) \in \mathrm{graph}(f)$, and this correspondence is a very naive form of C^p-diffeomorphism between $\mathrm{dom}(f)$ and $\mathrm{graph}(f)$ if f is C^p in a suitable sense, which is to be made more precise when we talk about sub-manifolds. The idea of content of Theorem 1.13.1 is a way to combine two types of simple maps—the bounded linear one, as well as the homeomorphism from the domain of a map onto its graph. More precisely we want to characterize maps that, after a change of variable on the domain, take the form

$$f : x \in U \subseteq E \mapsto (Ax, \varphi(Ax)) \in F_1 \times F_2 = F, \tag{1.13.3}$$

where U is an open set in the normed space E, and $F = F_1 \times F_2$ is a product of normed spaces F_1 and F_2 that accommodate the graph of the map φ.

Theorem 1.13.1. — *Let E, F be Banach spaces, $U \subseteq E$ open, $a \in U$, and $f \in C^p(U, F)$ with $p \in \mathbb{N}^+ \cup \{\infty\}$. Set $A = \mathrm{d}f(a) \in \mathcal{L}(E, F)$. Suppose that there exists a $P \in \mathrm{Idem}(F)$ such that $\mathcal{R}(P) = \mathcal{R}(A)$ and $\mathcal{N}(P) \cap \mathcal{R}\big(\mathrm{d}f(x)\big) = 0$ for any x in some small open neighborhood U_0 of a in U. If there exists $S \in \mathcal{L}\big(\mathcal{R}(A), E\big)$ with $AS = \mathrm{Id}_{\mathcal{R}(A)}$, then*

1. *There exists an open ball W_0 in E, a C^p-diffeomorphism $g : W_0 \to g(W_0)$ with $g(W_0)$ being open in E and $g(W_0) \subseteq U_0$, and a $\varphi \in C^p\big(A(W_0), \mathcal{N}(P)\big)$, such that*

$$\forall \xi \in W_0, \qquad f(g(\xi)) = A\xi + \varphi(A\xi), \tag{1.13.4}$$

   with $A(W_0)$ being an open set in the Banach space $\mathcal{R}(A) = \mathcal{R}(P)$.

2. *Identify F with $\mathcal{R}(P) \times \mathcal{N}(P)$ via the canonical isomorphism determined by P and consider the C^p-map*

$$\begin{aligned} h : A(W_0) \times \mathcal{N}(P) &\to A(W_0) \times \mathcal{N}(P) \\ (y_1, y_2) &\mapsto (y_1, y_2 - \varphi(y_1)), \end{aligned} \tag{1.13.5}$$

   then h is a C^p-diffeomorphism from the open subset $A(W_0) \times \mathcal{N}(P)$ onto itself such that

$$\forall \xi \in W_0, \qquad (h \circ f \circ g)(\xi) = A\xi. \tag{1.13.6}$$

Proof. — We first establish (1), which is the main part of the proof. Consider the map $\Phi \in C^p(U, F)$ defined by

$$\Phi(x) = x + SP\big(f(x) - Ax\big). \tag{1.13.7}$$

Here we use the same letter P to denote the restriction (on the range) $P : E \to \mathcal{R}(P)$ of P, so the composition SP makes sense. Note that by assumption, $AS = \mathrm{Id}_{\mathcal{R}(A)}$ and $\mathcal{R}(A) = \mathcal{R}(P)$, we have

$$\forall x \in U, \qquad A\Phi(x) = Ax + ASP\big(Ax - f(x)\big) = Ax + P\big(Ax - f(x)\big) = Pf(x). \tag{1.13.8}$$

It follows from (1.13.7) that

$$\mathrm{d}\Phi(a) = \mathrm{Id}_E + SP(\mathrm{d}f(a) - A) = \mathrm{Id}_E \in \mathrm{Aut}(E) := \mathrm{Iso}(E, E). \tag{1.13.9}$$

By the inverse mapping theorem, Φ restricts to a C^p-diffeomorphism from an open neighborhood of a inside U_0 onto some open ball W_0 in E, and we take g to be the inverse of this C^p-diffeomorphism. Clearly $g(W_0) \subseteq U_0$, and by (1.13.8), we have

$$\forall \xi \in W_0, \qquad A\xi = A\Phi\big(g(\xi)\big) = Pf\big(g(\xi)\big). \tag{1.13.10}$$

We thus define $\psi \in C^p(W_0, \mathcal{R}(P))$ (note again that $\mathcal{R}(A) = \mathcal{R}(P)$ by assumption) by

$$\psi(\xi) := f\big(g(\xi)\big) - A\xi. \tag{1.13.11}$$

Note that $\mathcal{R}(A) = \mathcal{R}(P)$ implies $PA = A$, and it follows from (1.13.10) that $P\psi = 0$. Hence it now suffices to show that $A(W_0)$ is open $\mathcal{R}(A) = \mathcal{R}(P)$, and that there exists $\varphi \in C^p(A(W_0), \mathcal{N}(P))$ with

$$\forall \xi \in W_0, \qquad \varphi(A\xi) = \psi(\xi). \tag{1.13.12}$$

Note that $A : E \to \mathcal{R}(A) = \mathcal{R}(P)$ is a linear, bounded and surjective operator between Banach spaces, the open mapping theorem[9] says that A is open. In particular, $A(W_0)$ is open in $\mathcal{R}(A) = \mathcal{R}(P)$. On the other hand, take any $\xi_1, \xi_2 \in W_0$ with $\psi(\xi_1) = \psi(\xi_2)$. For all $t \in]0, 1[$, we have $\xi_1 + t(\xi_2 - \xi_1) \in W_0$ since the open ball W_0 is convex. Since $A\xi_1 = A\xi_2$, we have

$$\psi(\xi_1) - \psi(\xi_2) = f\big(g(\xi_1)\big) - f\big(g(\xi_2)\big). \tag{1.13.13}$$

We define

$$\begin{aligned} \alpha : [0, 1] &\to W_0 \subseteq E \\ t &\mapsto \xi_1 + t(\xi_2 - \xi_1), \end{aligned} \tag{1.13.14}$$

and

$$\beta = f \circ g \circ \alpha. \tag{1.13.15}$$

Then for any $t \in]0, 1[$, we have

$$\beta'(t) = \mathrm{d}f\big(g(\alpha(t))\big)[\mathrm{d}g(\alpha(t)).(\xi_2 - \xi_1)] \in \mathcal{R}\Big(\mathrm{d}f\big(g(\alpha(t))\big)\Big). \tag{1.13.16}$$

By assumption, $\mathcal{N}(P) \cap \mathcal{R}\Big(\mathrm{d}f\big(g(\alpha(t))\big)\Big) = 0$, thus

$$\beta'(t) = 0 \iff P\beta'(t) = 0. \tag{1.13.17}$$

On the other hand, by (1.13.10), we have $P \circ f \circ g = A \restriction_{W_0}$, thus

$$\forall t \in]0, 1[, \qquad P\beta(t) = A\alpha(t) = A(\xi_1 + t(\xi_2 - \xi_1)) = A\xi_1 = A\xi_2. \tag{1.13.18}$$

The right side of (1.13.18) is independent of t, thus taking derivatives of (1.13.18), we obtain

$$P\beta'(t) = 0. \tag{1.13.19}$$

By (1.13.17) and (1.13.18), we obtain

$$\forall t \in]0, 1[, \qquad \beta'(t) = 0. \tag{1.13.20}$$

[9] This will be proved in later course on functional analysis. We also include a proof in the optional exercise.

Now by the mean value theorem, we see that

$$f\big(g(\xi_1)\big) = \beta(0) = \beta(1) = f\big(g(\xi_2)\big), \tag{1.13.21}$$

and we obtain from (1.13.13) that

$$\forall \xi_1, \xi_2 \in W_0 \cap \mathcal{N}(A), \qquad \psi(A\xi_1) = \psi(A\xi_2). \tag{1.13.22}$$

Therefore, the equation

$$\varphi(A\xi) = \psi(\xi) \tag{1.13.23}$$

gives a well-defined map φ from $A(W_0)$ to $\mathcal{R}(P)$. It remains to check that φ is of class C^p. For this, fix any $\xi_0 \in W_0$ and let $\eta_0 = A\xi_0$. For each $\xi \in W_0$ and $\eta = A\xi$, we have

$$A\big(\xi_0 + S(\eta - \eta_0)\big) = \eta_0 + (\eta - \eta_0) = \eta. \tag{1.13.24}$$

Since S is bounded, there exists $r_0 > 0$, such that $\eta \in B(\eta_0, r)$ implies that $\xi_0 + S(\eta - \eta_0) \in W_0$. This means

$$\forall \eta \in B(\eta_0, r_0), \qquad \varphi(\eta) = \varphi\big(A(\xi_0 + S(\eta - \eta_0))\big) = \psi(\xi_0 + S(\eta - \eta_0)), \tag{1.13.25}$$

which clearly implies that $\varphi \restriction_{B(\eta_0, r_0)}$ is of class C^p. Since $\eta_0 = A\xi_0$ is arbitrary in $A(W_0)$, this shows that $\varphi \in C^p\big(A(W_0), \mathcal{R}(P)\big)$, and the proof of (1) is complete.

For (2), note first that $A(W_0)$ is open in $\mathcal{R}(A) = \mathcal{R}(P)$ by the open mapping theorem as noted above. One now checks immediately that the C^p-map

$$\begin{aligned} A(W_0) \times \mathcal{N}(P) &\to A(W_0) \times \mathcal{N}(P) \\ (y_1, y_2) &\mapsto (y_1, y_2 + \varphi(y_1)) \end{aligned} \tag{1.13.26}$$

is the inverse of h, and h is indeed a C^p-diffeomorphism. Now (1.13.6) follows from applying h to (1.13.4) and the proof is complete. $\square$

In finite dimensional spaces, Theorem 1.13.1 can take a more concrete form that employs the Jacobian matrices, which is a strong form of the classical rank theorem.

***Theorem 1.13.2* (The Rank Theorem).** — *Let $m, n \in \mathbb{N}^+$, $p \in \mathbb{N}^+ \cup \{\infty\}$, $U \subseteq \mathbb{R}^n$ open, $a \in U$ and $f \in C^p(U, \mathbb{R}^m)$. Set $A = \mathrm{d}f(a) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ and $r = \mathrm{rank}(A)$. Suppose $P \in \mathrm{Idem}(\mathbb{R}^m)$ is a projection onto $\mathcal{R}(A)$. If there exists an open neighborhood U_0 of a inside U such that* $\mathrm{rank}\, Jf(x) = r$ *for all $x \in U_0$, then*

1. *There exists an open set W_0 in $\mathbb{R}^n$, a C^p-diffeomorphism g from W_0 onto some open neighborhood of a inside U_0, and a $\varphi \in C^p(A(W_0), \mathcal{N}(P))$ such that*

$$\forall \xi \in W_0, \qquad f(g(\xi)) = A\xi + \varphi(A\xi). \tag{1.13.27}$$

2. *Identify $\mathbb{R}^m$ with $\mathcal{R}(P) \times \mathcal{N}(P)$ via the canonical isomorphism determined by P and consider the C^p-map*

$$\begin{aligned} h : A(W_0) \times \mathcal{N}(P) &\to A(W_0) \times \mathcal{N}(P) \\ (y_1, y_2) &\mapsto (y_1, y_2 - \varphi(y_1)), \end{aligned} \tag{1.13.28}$$

   then h is a C^p-diffeomorphism from the open subset $A(W_0) \times \mathcal{N}(P)$ onto itself such that

$$\forall \xi \in W_0, \qquad (h \circ f \circ g)(\xi) = A\xi. \tag{1.13.29}$$

Proof. — We establish this theorem by apply Theorem 1.13.1, so it suffices to check that the hypothesis in Theorem 1.13.1 are satisfied under the assumption here.

By choosing a basis $(v_1, \ldots, v_r)$ for $\mathcal{R}(A)$ and extend it to basis $(v_1, \ldots, v_r, u_1, \ldots, u_{n-r})$ for $\mathbb{R}^m$, we may construct an $S \in \mathcal{L}(\mathcal{R}(A), \mathbb{R}^n)$ with $AS = \mathrm{Id}_{\mathcal{R}(A)}$. Indeed, choose $u_i \in A^{-1}(v_i)$ for each $i = 1, \ldots, r$, and it suffices to define S to be the unique linear map from $\mathcal{R}(A)$ to $\mathbb{R}^n$ that sends each v_i to u_i.

On the other hand, since $\mathrm{d}f$ is continuous, we know the Jacobian matrix $Jf(x)$ is a continuous function of x, so the product $M_P Jf(x) = (\alpha_{i,j}(x))$ of matrices is a continuous function of x, where M_P is the matrix of P under the canonical basis, and each $\alpha_{i,j}(x)$ is a continuous function of x. We know that $PA = A$, hence $M_P Jf(a)$ is of rank $\operatorname{rank} A = r$. There is thus an $r \times r$ minor, say the $(i_1, \dots, i_r) \times (j_1, \dots, j_r)$ minor, of $M_P Jf(a)$ whose determinant is nonzero, i.e.

$$\det\big(\alpha_{i_k, j_l}(a)\big)_{1 \le k,l \le r} \neq 0. \tag{1.13.30}$$

But $\det\big(\alpha_{i_k, j_l}(x)\big)_{1 \le k,l \le r}$ is a continuous function of x, thus there exists an open neighborhood U_1 of x inside U_0, such that

$$\forall x \in U_1, \qquad \det\big(\alpha_{i_k, j_l}(x)\big)_{1 \le k,l \le r} \neq 0, \tag{1.13.31}$$

which in turn implies

$$\forall x \in U_1, \qquad r = \operatorname{rank} P \geq \operatorname{rank}\big(M_P Jf(x)\big) \geq r, \tag{1.13.32}$$

forcing

$$\forall x \in U_1, \qquad \operatorname{rank}\big(P\mathrm{d}f(x)\big) = \operatorname{rank}\big(M_P Jf(x)\big) = r. \tag{1.13.33}$$

Since $\operatorname{rank} P = r$, this means that P maps $\mathcal{R}(\mathrm{d}f(x))$ onto $\mathcal{R}(P)$. But $\mathrm{d}f(x)$ is of rank r, this means $\mathcal{R}(\mathrm{d}f(x))$ and $\mathcal{R}(P)$ both have dimension r. We know from linear algebra that a surjective linear map between two finite dimensional vector spaces of the same dimension must be bijective, it follows that P maps $\mathcal{R}(\mathrm{d}f(x))$ bijectively onto $\mathcal{R}(P)$. In particular, $\mathcal{R}(\mathrm{d}f(x)) \cap \mathcal{N}(P) = 0$.

Now apply Theorem 1.13.1 and the proof is complete. □

Remark 1.13.3. — Theorem 1.13.2 implies immediately the local structures of immersions and submersions of smooth manifolds. We will see these results in the special case of open subsets of $\mathbb{R}^n$ in the exercise.

1.14. Submanifolds of $\mathbb{R}^n$ and Lagrange Multipliers

We have seen in some results in § 1.13 that transforms a general C^p-map satisfying certain conditions to the graph of C^p-maps from a "smaller" space, with the implicit understanding that the graph of C^p-maps should be certain "good" analytic/geometric object. We now formalize this implicit understanding by introducing submanifolds of finite dimensional normed spaces, i.e. normed spaces that are isomorphic to $\mathbb{R}^n$, where n is the dimension.

Before giving the formal definition, let's look at a simple example. Fix an $n \in \mathbb{N}^+$ and consider the $(n-1)$-unit sphere S^{n-1}, which is defined by

$$S^{n-1} := \left\{ x = (x_1, \dots, x_n) \in \mathbb{R}^n \;\middle|\; \|x\|_2 = \sum_{i=1}^n x_i^2 = 1 \right\}, \tag{1.14.1}$$

which is clearly a compact subset of $\mathbb{R}^n$. From its definition, the particularly nice geometric object S^{n-1} is clearly closely related to the smooth function $F : \mathbb{R}^n \to \mathbb{R}$ defined by $F(x_1, \dots, x_n) = \sum_{i=1}^n x_i^2$. In fact, $S^{n-1} = F^{-1}(1)$. An easy computation gives

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n, \qquad \operatorname{grad} F(x) = \operatorname{grad} F(x_1, \dots, x_n) = (2x_1, \dots, 2x_n) = 2x. \tag{1.14.2}$$

Hence $\operatorname{grad} F(x) \neq 0$ at any point $x \neq 0$. We will later formalize this property by saying that the restriction $F : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ is a submersion.

Now given any $a = (a_1, \dots, a_n) \in S^{n-1}$, we have $a_i \neq 0$ for some $i \in \{1, \dots, n\}$. It follows from the implicit function theorem that near a, we may write points of S^{n-1} as the graph of a smooth function. More specifically, we may consider the decomposition of $\mathbb{R}^n$ as $\mathbb{R}^{n-1} \times \mathbb{R}$ by putting the i-th coordinate to the end, namely we identify $\mathbb{R}^n$ with $\mathbb{R}^{n-1} \times \mathbb{R}$ via the isomorphism (of Banach spaces)

$$\begin{aligned} \Phi_i : \mathbb{R}^n &\xrightarrow{\sim} \mathbb{R}^{n-1} \times \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto \Big((x_1, \dots, \widehat{x_i}, \dots x_n), x_i \Big) \end{aligned} \tag{1.14.3}$$

where the hat $\widehat{\cdot}$ notation means that coordinate is omitted. We see that F is invariant under the identification Φ_i, i.e. $F \circ \Phi_i = F$. Since $a_i \neq 0$, we see that $\partial_i F(a) = 2a_i \neq 0$, and the implicit function theorem says that there exists an open neighborhood V_i of a_i in $\mathbb{R}$, an open neighborhood U_i of $(a_1, \dots, \widehat{a_i}, \dots, a_{i-1})$ in $\mathbb{R}^{n-1}$, such that there exists a unique map $\varphi : U_i \to V_i$ satisfying

$$\Phi_i(S^{n-1}) \cap (U_i \times V_i) = \operatorname{graph}(\varphi). \tag{1.14.4}$$

Moreover, this φ is smooth. The important property of S^{n-1} that we want to abstract from (1.14.4) is that locally near any point $a \in S^{n-1}$, we may describe S^{n-1} as a graph of a smooth map, modulo certain identification Φ_i. Here comes the formal definition.

Definition 1.14.1. — Let E be a finite dimensional normed space and $S \subseteq E$. We say S is an d-**dimensional** C^p-**submanifold** of E, where $d \in \mathbb{N}$ and $p \in \mathbb{N} \cup \{\infty\}$, if for each $a \in S$, we may describe S locally near a as the graph of a C^p-map in the sense that there exists a decomposition $E = E_1 \times E_2$ of E as a product of normed spaces with $\dim E_1 = d$, an open set U_i of E_i $(i = 1, 2)$ such that $a = (a_1, a_2) \in U_1 \times U_2$, and a $\varphi \in C^p(U_1, U_2)$, satisfying

$$S \cap (U_1 \times U_2) = \operatorname{graph}(\varphi) = \{(x_1, \varphi(x_1)) \mid x_1 \in U_1\}. \tag{1.14.5}$$

When $p = \infty$, instead of C^p-submanifold, we often speak of **smooth submanifold**, or simply a **submanifold**.

Remark 1.14.2. — As a subset of E, there is naturally a topological structure on S, namely the subspace topology induced from the topology of E. When $p = 0$ in Definition 1.14.1, the C^0-submanifold S is actually a d-dimensional topological manifold in E as you have seen in your topology course. The verification for this is just an easy exercise.

It is now clear from the consideration before Definition 1.14.1 that S^{n-1} is a smooth submanifold of $\mathbb{R}^n$. It is also clear that S^{n-1} is the inverse image of a single point, namely $\{1\}$, of the smooth map $F : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ whose gradient never vanishes. The latter observation is actually a typical feature of submanifolds, and can be used to characterize submanifolds as well as giving a lot of examples. We now introduce the relevant definition and result for this.

Definition 1.14.3. — Let E and F be normed spaces, U (resp. V) is open in E (resp. F). For $p \in \mathbb{N}^+ \cup \{\infty\}$, we say φ is a C^p-**submersion** from U to V, if $\varphi \in C^p(U, V)$ and $\mathrm{d}\varphi(a) \in \mathcal{L}(E, F)$ is surjective, i.e. $\mathcal{R}(\mathrm{d}\varphi(a)) = F$, for any $a \in U$.

Proposition 1.14.4. — *Let E be a finite dimensional normed space and $S \subseteq E$. Suppose $n = \dim E$ and*[10] *$p \in \mathbb{N}^+ \cup \{\infty\}$. Then the following are equivalent:*

1. *S is a d-dimensional C^p submanifold of E.*
2. *S is locally the level set of a submersion to an $(n-d)$-dimensional normed space, i.e. for any $a \in S$, there exists an open neighborhood U_a of a in E, an $(n-d)$-dimensional normed space G and a C^p-submersion $\psi_a \in C^p(U_a, F)$, such that*

$$S \cap U_a = \psi_a^{-1}(\psi_a(a)). \tag{1.14.6}$$

Proof. — (1) implies (2). Without loss of generality (why?), we may assume that $E = F \times G$, with F (resp. G) being a d-dimensional(resp. $(n-d)$-dimensional) normed space, and $S = \operatorname{graph}(\varphi)$ for some $\varphi \in C^p(U, V)$ with U (resp. V) being an open subset in F (resp. G). Now define $\psi \in C^p(U \times V, G)$ by setting $\psi(x, y) = y - \varphi(x)$. Since $\partial_y \psi(x, y) = \mathrm{Id}$ for all $(x, y) \in U \times V$, we see that ψ is a C^p-submersion. By definition, for any point

$$a = (a_x, a_y) \in S = \operatorname{graph}(\varphi) \subseteq F \times G = E,$$

we have $\psi(a) = 0$ and we indeed have

$$\operatorname{graph}(\varphi) = \psi^{-1}(0) = \psi^{-1}(\psi(a)). \tag{1.14.7}$$

Now just set $\psi_a = \psi$.

(2) implies (1). For each $a \in S$. By (2), we have obtained an open set $U_a \subseteq E$, a decomposition $E = F \times G$ and a C^p-submersion ψ_a. Let $(g_1, \dots g_{n-d})$ be a basis for the $(n-d)$-dimensional vector space G, since $\mathrm{d}\psi_a(a) \in \mathcal{L}(E, G)$ is surjective, we may take an $e_i \in E$ with $\mathrm{d}\psi_a(a).e_i = g_i$ for each

[10] Note that we no longer allow $p = 0$, as we want to use the implicit function theorem in the proof

$i = 1, \dots, n-d$. Let E_0 be the subspace of E spanned by $e_1, \dots, e_{n-d}$, and F a complementary subspace of E_0 in E, we then have $E = F \oplus E_0 \simeq F \times E_0$, where the last isomorphism is the canonical one, through which we identify E with $E_0 \times F$. Since U_a is an open neighborhood of $a \in E = F \times E_0$ (note that we have identified E with $F \times E_0$), we may find an open set U_1 in E_0 and an open set V_1 in F, such that $a \in U_1 \times V_1 \subseteq U_a$. Following the decomposition $E = F \times E_0$ and apply Proposition 1.6.7, and note that each $\eta \in E_0$ is identified with $(0, \eta) \in F \times E_0 = E$, we see that

$$\forall \eta \in E_0, \qquad \partial_2 f(a).\eta = \mathrm{d}f(a).\eta. \tag{1.14.8}$$

In particular, $\partial_2 \psi_a.e_i = \mathrm{d}\psi_a.e_i = g_i$ for each $i = 1, \dots, n-d$, so $\partial_2 \psi(a) \in \mathcal{L}(E_0, G)$ is surjective, hence invertible because $\dim E_0 = n - d = \dim G$. By the implicit function theorem applied to φ_a and a, we see, shrinking U_1 and V_1 if necessary, that there is a $\varphi \in C^p(V_1, U_1)$ such that for each $x \in V_1$, $\varphi(x)$ is the unique point in U_1 such that

$$\psi(x, \varphi(x)) = \psi(a). \tag{1.14.9}$$

In other words,

$$S \cap U_a = \psi^{-1}(\psi(a)) = \mathrm{graph}(\varphi). \tag{1.14.10}$$

Since $a \in S$ is arbitrary and

$$\dim F = \dim E - \dim E_0 = n - (n-d) = d, \tag{1.14.11}$$

this means precisely that S is a d-dimensional C^p-submanifold of E. □

Corollary 1.14.5. — *Let E, F be finite dimensional normed spaces, $U \subseteq E$ open, and $\varphi \in C^p(U, F)$ with $p \in \mathbb{N}^+ \cup \{\infty\}$. Fix any $y_0 \in F$, if $\mathrm{d}\varphi(x_0) \in \mathcal{L}(E, F)$ is surjective at any $x_0 \in \varphi^{-1}(y_0)$, then $\varphi^{-1}(y)$ is a $(n-m)$-dimensional C^p-submanifold of E.*

Proof. — Since $\mathrm{rank}(J\varphi(x_0)) = m$, i.e. $J\varphi(x_0)$ is of maximal rank, at each $x_0 \in \varphi^{-1}(y)$, there is a $m \times m$ submatrix of $J\varphi(x_0)$, say the one with the $i_1, \dots, i_m$ rows of $J\varphi(x_0)$, with nonzero determinant. By continuity, there exists an open neighborhood U_0 of x_0 in E, such that the same $m \times m$ submatrix formed by the $i_1, \dots, i_m$ rows of $J\varphi(x)$ still has a nonzero determinant for all $x \in U_{x_0}$. Now set $U = \cup_{x_0 \in \varphi^{-1}(y_0)} U_{x_0}$. It is clear from our construction that $\varphi|_U : U \to F$ is a submersion with $U \supseteq \varphi^{-1}(y_0)$. Hence

$$\varphi^{-1}(y_0) = (\varphi \restriction_U)^{-1}(y_0) \tag{1.14.12}$$

is a $(n-m)$-dimensional C^p-submanifold of E by Proposition 1.14.4. □

With these preparatory consideration on submanifolds, we shift our attention to the problem of finding extreme values of a function that are defined on high dimensional curved objects. In the simplest case where our curved object is an open set in a finite dimensional normed space, Proposition 1.10.7 says that the extreme values can only be attained at the so-called critical points as defined in the following definition.

Definition 1.14.6. — Let E be a normed spaces, $U \subseteq E$ open, $a \in U$ and $f : U \to E$ differentiable. We call the point a a **critical point** of f, if $\mathrm{d}f(a) = 0$.

Our discussion of sub-manifolds gives an elegant method for finding the extreme value of certain functions under certain constraints, still with the help of critical points. Before describing this method, let's first look at a simple example. Consider again the unit sphere

$$S^2 = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 \;\middle|\; \sum_{i=1}^{3} x_i^2 = 1 \right\}. \tag{1.14.13}$$

Suppose that we want to find the maximum of the function $s(x_1, x_2, x_3) = x_1 + x_2 + x_3$ on S^2 that represents taking the sum of each coordinate. The subtlety lies in the fact that even s is defined on the whole $\mathbb{R}^3$, we are merely interested in maximizing the restriction $s \restriction_{S^2}$, and S^2 is *not* open in $\mathbb{R}^3$. This means that we may no longer find local extreme points by solving $\mathrm{d}s(x) = 0$ for $x \in S^2$. In fact, it is clear that $\mathrm{grad}\, s(x) = (1, 1, 1)$ at all points $x \in \mathbb{R}^3$ so one can never find a solution to the equation

$\mathrm{d}s(x) = 0$. Nevertheless, since S^2 is compact, the continuous function $s \restriction$ must reach its maximum at some $P_0 \in S^2$, and this P_0 should be some sort critical point for our problem of finding $\max s \restriction_{S^2}$.

Intuitively, this failure of finding critical points naively is due to the fact that S^2 is a "low dimensional" object while our theory of differential calculus is based on maps that are defined on open sets (so that it makes sense to speak rigorously the idea of differentiability is a *local* property) which should be of the same dimension as the ambient space. Here, though we haven't learned a good theory of topological spaces, we still have the naive feeling that being low dimensional means there are less free variables. Thinking in this way, one natural attempt to circumvent this "lack of free variables" would of course be introducing more free variables "without destroying the critical points". Here's a positive result, due to Lagrange, that realizes this idea.

***Theorem 1.14.7* (Lagrange).** — *Let E, F be a finite dimensional normed spaces, $U \subseteq E$ open, $p \in \mathbb{N}^+ \cup \{\infty\}$, $v \in F$, and $\Phi \in C^p(U, F)$. Set $S := \Phi^{-1}(v) \subseteq U \subseteq E$, and consider an $f \in C^p(V, \mathbb{R})$ where V is open set in E containing S. Suppose the following hold:*

1. *$f \restriction_S$ achieves its maximum or minimum at some $a \in S$;*
2. *$\mathrm{d}\Phi(x)$ is surjective for each $x \in S$.*

Then there exists an $\omega \in F' = \mathcal{L}(F, \mathbb{R})$, such that $a \in U \cap V$ is a critical point for the function $f - \omega \circ \Phi \in C^p(U \cap V, \mathbb{R})$ that is defined on the open subset $U \cap V$ of E.

Proof. — By Corollary 1.14.5, S is a C^p-submanifold of E whose dimension is $\dim E - \dim F$. But we actually need to use the proof of this fact which we repeat here for completeness. By hypothesis, $\mathrm{d}\Phi(a)$ is surjective, which implies that we can find a linear subspace E_2 of E such that $\mathrm{d}\Phi(a) \restriction_{E_2} \in \mathrm{Iso}(E_2, F)$ and hence a decomposition $E = E_1 \times E_2$ of E into the product of two finite dimensional normed spaces such that $\partial_2\Phi(x) \in \mathrm{Iso}(E_2, F)$ for all x in an open neighborhood V of a contained in U. By the implicit function theorem, writing $a = (a_1, a_2)$ with respect to this decomposition, there exists an open neighborhood U_i of a_i in E_i $(i = 1, 2)$ with $U_1 \times U_2 \subseteq V$, and a $\varphi \in C^p(U_1, U_2)$, such that for each $x_1 \in U_1$, $\varphi(x_1)$ is the unique point in U_2 satisfying $\Phi(x_1, \varphi(x_1)) = v$. We then have $\varphi(a_1) = a_2$, and

$$S \cap (U_1 \cap U_2) = \mathrm{graph}(\varphi) = \big\{ \big(x_1, \varphi(x_1)\big) \bigm| x_1 \in U_1 \big\}. \tag{1.14.14}$$

Now define $g \in C^p(U_1, \mathbb{R})$ by $g(x_1) := f(x_1, \varphi(x_1))$. Since $(x_1, \varphi(x_1)) \in \mathrm{graph}(\varphi) \subseteq \varphi$ for all $x_1 \in U_1$, $(a_1, \varphi(a_1)) = (a_1, a_2) = a$, we see that g achieves its maximum or minimum at a_1. Hence

$$\mathrm{d}g(a_1) = 0. \tag{1.14.15}$$

i.e. a_1 is a critical point for g_1.

On one hand, by the chain rule, we have

$$\forall x_1 \in U_1, \qquad \mathrm{d}g(x_1) = \partial_1 f(x_1, \varphi(x_1)) + \partial_2 f(x_1, \varphi(x_1))\mathrm{d}\varphi(x_1). \tag{1.14.16}$$

On the other hand, since $S = \Phi^{-1}(c)$, we have

$$\forall x_1 \in U_1, \qquad \Phi(x_1, \varphi(x_1)) = v. \tag{1.14.17}$$

Differentiating (1.14.17) as maps of the variable x_1, we obtain

$$\forall x_1 \in U_1, \qquad \partial_1\Phi(x_1, \varphi(x_1)) + \partial_2\Phi(x_1, \varphi(x_1))\mathrm{d}\varphi(x_1) = 0 \tag{1.14.18}$$

As $(x_1, \varphi(x_1)) \in U_1 \times U_2 \subseteq V$, the operator $\partial_2\Phi(x_1, \varphi(x_1))$ is invertible, hence by (1.14.18), we have

$$\forall x_1 \in U_1, \qquad \mathrm{d}\varphi(x_1) = -[\partial_2\Phi(x_1, \varphi(x_1))]^{-1}\partial_1\Phi(x_1, \varphi(x_1)). \tag{1.14.19}$$

By (1.14.16) and (1.14.19), we have

$$\forall x_1 \in U_1, \qquad \mathrm{d}g(x_1) = \partial_1 f(x_1, \varphi(x_1)) - \partial_2 f(x_1, \varphi(x_1))[\partial_2\Phi(x_1, \varphi(x_1))]^{-1}\partial_1\Phi(x_1, \varphi(x_1)). \tag{1.14.20}$$

Now we set

$$\omega := \partial_2 f(a_1, \varphi(a_1))[\partial_2\Phi(a_1, \varphi(a_1))]^{-1} = \partial_2 f(a)[\partial_2\Phi(a)]^{-1} \in \mathcal{L}(F, \mathbb{R}) = F'. \tag{1.14.21}$$

By (1.14.15) and (1.14.20), we see that

$$\partial_1(f - \omega \circ \Phi)(a) = \partial_1 f(a) - \omega\partial_1\Phi(a) = 0. \tag{1.14.22}$$

And by the definition of ω, we have

$$\partial_2(f - \omega \circ \Phi)(a) = \partial_2 f(a) - \omega\partial_2\Phi(a) = 0. \tag{1.14.23}$$

It follows from Proposition 1.6.7 and (1.14.22) as well as (1.14.23) that

$$\mathrm{d}(f - \omega \circ \Phi)(a) = 0, \tag{1.14.24}$$

i.e. a is indeed a critical point for $f - \omega \circ G$. $\square$

We now explain how Theorem 1.14.7 can be applied to find the $a \in S$ that maximizes (or minimizes) $f \restriction_S$. By choosing vector basis and use the uniqueness of norms up to equivalence on finite dimensional vector spaces, we may simply assume $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$, and the functional $\omega \in F' = (\mathbb{R}^m)'$ is completely determined by the m numbers $c_1, \dots, c_m$, where $c_i = \omega(e_i)$, and $(e_1, \dots, e_m)$ is the standard basis for $\mathbb{R}^m$. Write $\Phi = (\Phi_1, \dots, \Phi_m)$ with each $\Phi_i \in C^p(U, \mathbb{R})$, i.e. $\Phi(x) = (\Phi_1(x), \dots, \Phi_m(x))$ for each $x \in U$, and set $L = f - \omega \circ \Phi \in C^p(U \cap V, \mathbb{R})$. We then have

$$\forall x = (x_1, \dots, x_n) \in U \cap V, \qquad L(x) = f(x) - \sum_{i=1}^{m} c_i \Phi_i(x). \tag{1.14.25}$$

Now Theorem 1.14.7 says that if $f \restriction_S$ achieves its maximum/minimum at some $a \in S$, then we may find a ω, or equivalently, m numbers c_1, ..., c_m, such that

$$\mathrm{d}L(a) = 0. \tag{1.14.26}$$

Writing (1.14.26) in terms of partial derivatives, we have

$$\forall i \in \{1, \dots, n\}, \qquad \frac{\partial f}{\partial x_i}(a) = \sum_{j=1}^{m} c_j \frac{\partial \Phi_j}{\partial x_i}(a). \tag{1.14.27}$$

(1.14.27) provides us with n equations for the $n + m$ unknowns $(a_1, \dots, a_n, c_1, \dots, c_m)$. We also have m equations

$$\forall j \in \{1, \dots, m\}, \qquad \Phi_j(a_1, \dots, a_n) = v_j \tag{1.14.28}$$

where $v = (v_1, \dots, v_j) \in \mathbb{R}^m = F$. Combining (1.14.27)(1.14.28), we obtain $n + m$ equations for the $n + m$ unknowns $(a_1, \dots, a_n, c_1, \cdots, c_m)$, which gives a necessary condition for $f \restriction_S$ achieves its maximum/minimum at a. Of course, this necessary condition needs not be sufficient, for which one usually has to check that $f(a)$ is indeed such an extreme value. Note that the m auxiliary constants $c_1, \dots, c_m$ appear as the scalar multiplication in front of the Φ_i's, it is for this reason that these constants are referred as **Lagrange multipliers**, and the above procedure for finding maximum/minimum of $f \restriction_S$ as the **method of Lagrange multipliers**.

We now revisit our previous example of finding the maximum value of $s \restriction_{S^2}$, where $s(x_1, x_2, x_3) = x_1 + x_2 + x_3$. It is already remarked that s does achieve its maximum at some point $a \in S^2$, our task is to find this a and the value $s(a)$. Note that we may now take $\Phi \in C^\infty(\mathbb{R}^3, \mathbb{R})$ defined by

$$\Phi(x_1, x_2, x_3) = \sum_{i=1}^{3} x_i^2, \tag{1.14.29}$$

so that $S^2 = \Phi^{-1}(1)$, and $\mathrm{d}\Phi(x) = \nabla\Phi(x) \neq 0$ for each $x \in S^2$, so that Theorem 1.14.7 and the method of Lagrange multipliers apply, and we may find a Lagrange multiplier $c \in \mathbb{R}$, such that $a = (a_1, a_2, a_3)$ satisfies the equations

$$\forall i = 1, 2, 3, \qquad 0 = \partial_i s(a) - c\partial_i \Phi(a) = 1 - 2ca_i, \tag{1.14.30}$$

and

$$1 = \Phi(a) = \sum_{i=1}^{3} a_i^2. \tag{1.14.31}$$

It follows from (1.14.30) that $c \neq 0$ and $a_1 = a_2 = a_3 = (2c)^{-1}$. Now substituting this into (1.14.31), we obtain $1 = 3(2c)^{-2}$, and $c = \pm(2\sqrt{3})^{-1}$. Hence either $a = (\sqrt{3})^{-1}(1, 1, 1)$ or $-(\sqrt{3})^{-1}(1, 1, 1)$. Clearly, only the former case, i.e. $a = 2\sqrt{3}(1,1,1)$, maximizes $s\restriction_{S^2}$. So we must have $a = (\sqrt{3})^{-1}(1,1,1)$ and the maximum value of $s\restriction_{S^2}$ is $s(a) = \sqrt{3}$. Note that $a = -(\sqrt{3})^{-1}(1,1,1)$ also solves the equations (1.14.30) and (1.14.31), but this value corresponds to the minimum of $s\restriction_{S^2}$, which is $-\sqrt{3}$. The proof is similar to the maximum value case.

Remark 1.14.8. — We emphasize in general, solving the equations (1.14.27) and (1.14.28) only yields a better selection of all the possible values of a on which our target function $f\restriction_S$ reaches its maximum/minimum. Which one corresponds to the actual maximum/minimum value still needs a verification, as illustrated in the above example. Sometimes, it could happen that even (1.14.27) and (1.14.28) do admit a solution a, but $f(a)$ is neither a local maximum nor a local minimum. For example, we may take $\Phi \in C^\infty(\mathbb{R}^2, \mathbb{R})$, $\Phi(x,y) = y - x^3$ and $f \in C^2(\mathbb{R}^2, \mathbb{R}), f(x,y) = y$. Consider $S = \Phi^{-1}(0) = \operatorname{graph}(\varphi)$ where $\varphi \in C^\infty(\mathbb{R}, \mathbb{R})$ is the cube function $\varphi(x) = x^3$. Now (1.14.27) and (1.14.28) become

$$\begin{aligned} 0 + 3cx^2 &= 0 \\ 1 - c &= 0 \\ y - x^3 &= 0. \end{aligned} \tag{1.14.32}$$

The solution for (1.14.32) does exist, and is given by $x = y = 0$, but $f\restriction_S$ does not have a local maximum or local minimum anywhere! This also shows that the hypothesis (1) in Theorem 1.14.7 can not be dropped.

We will see more applications of the method of Lagrange multipliers in the exercises.

1.15. (Not Covered in Course) The Morse Lemma

There is a general notion for critical points of smooth maps between smooth submanifolds. The method of Lagrange multipliers described in § 1.14 can be viewed as a way to find such critical points under the important special cases of smooth functions with values in $\mathbb{R}$. It is a natural and important question to investigate the local behavior of a smooth function near its critical points. The general form of this question is to consider smooth functions on smooth sub-manifolds. We will return to this point after introducing more basics about sub-manifolds in a later part of these notes. For now, we focus on the case of smooth functions defined on an open subset of $\mathbb{R}^n$. In the case of non-degenerate critical point, this question has a nice answer known as the Morse lemma, which opens the gate to a whole theory, the Morse theory , that studies the topology of manifolds via looking at the functions defined on them.

Definition 1.15.1. — Let $n \in \mathbb{N}^+$, $k \in \mathbb{N}_{\geq 2} \cup \{\infty\}$ and $U \subseteq \mathbb{R}^n$ open. A critical point $a \in U$ for $f \in C^k(U, \mathbb{R})$ is called **non-degenerate**, if the symmetric bilinear form $\mathrm{d}^2 f(a)$ on $\mathbb{R}^n$ is non-degenerate, i.e. if the Hessian matrix $Hf(a)$ is invertible.

Theorem 1.15.2 **(The Morse Lemma).** — *Let $U \subseteq \mathbb{R}^n$ $(n \in \mathbb{N}^+)$ be open and $f \in C^k(U, \mathbb{R})$ with $k \in \mathbb{N}_{\geq 3} \cup \{\infty\}$. Suppose $a \in U$ is a non-degenerate critical point for f. Then the following hold.*

(1) *There exists a C^{k-2}-diffeomorphism Ψ from an open neighborhood $V \subseteq \mathbb{R}^n$ of 0 onto an open set in $\mathbb{R}^n$ containing a, such that $\Psi(0) = a$, and*

$$\forall y \in V, \qquad (f \circ \Psi)(y) = f(a) + \langle Hf(a)y\,,\, y\rangle, \tag{1.15.1}$$

where $\langle \cdot\,,\,\cdot\rangle$ denotes the canonical inner product on $\mathbb{R}^n$.

(2) *There exists a C^{k-2}-diffeomorphism Θ from an open neighborhood $W \subseteq \mathbb{R}^n$ of 0 onto an open set in $\mathbb{R}^n$ containing a, such that $\Theta(0) = a$, and*

$$\forall w = (w_1, \ldots, w_n) \in W, \qquad (f \circ \Theta)(w) = f(a) + \sum_{i=1}^p w_i^2 - \sum_{j=1}^q w_{p+j}^2, \tag{1.15.2}$$

where (p,q) is the signature of the symmetric matrix $Hf(a)$, i.e. counting multiplicity, among the n-eigenvalues of $Hf(a)$, there are exactly p eigenvalues that are positive and q negative, so that $p + q = n$.

Proof. — By Taylor's formula up to order 2, and note that $\mathrm{d}f(a) = 0$ since a is a critical point for f, we have

$$f(x) = f(a) + \left(\int_0^1 (1-t)\mathrm{d}^2 f(a+t(x-a))\mathrm{d}t\right).(x-a)^{(2)} \tag{1.15.3}$$

for all x near a. Writing

$$Q(x) = \int_0^1 (1-t)Hf(a+t(x-a))\mathrm{d}t, \tag{1.15.4}$$

we may rewrite (1.15.3) in the form

$$f(x) = f(a) + \big\langle Q(x)(x-a)\,,\,x-a\,\big\rangle \tag{1.15.5}$$

for all x near a, and

$$Q(a) = Hf(a). \tag{1.15.6}$$

In particular, $Q(a)$ is invertible since the critical point a for f is non-degenerate. Moreover, since $\mathrm{d}^2 f$, hence equivalently Hf, is of class C^{k-2}, the matrix-valued function $Q(x)$ is also $(k-2)$-times continuously differentiable by repeated application of Proposition 1.9.3. Now consider the map

$$\begin{aligned}\Phi(B,x) :&= \left(I_n + \frac{1}{2}Q(a)^{-1}B\right)^t Q(a)\left(I_n + \frac{1}{2}Q(a)^{-1}B\right)\\ &= Q(a) + B + \frac{1}{4}BQ(a)^{-1}B - Q(x)\end{aligned} \tag{1.15.7}$$

where B varies in an open neighborhood of 0 in the finite dimensional linear subspace $\mathrm{Sym}_{n\times n}(\mathbb{R})$ of symmetric $n \times n$ matrices over $\mathbb{R}$, and x varies in an open neighborhood of a in $\mathbb{R}^n$. We have Φ is of class C^{k-2} since Q is so. Moreover,

$$\Phi(0,a) = 0 \tag{1.15.8}$$

and

$$\partial_B \Phi(B,x) = \mathrm{Id}\,. \tag{1.15.9}$$

Hence by the implicit function theorem, the equation

$$\Phi(B,x) = \Phi(0,a) = 0 \tag{1.15.10}$$

defines B as a C^{k-2} function of x near $(0,a)$, i.e. there exists an open neighborhood V_0 of a in $\mathbb{R}$, an open neighborhood $\mathcal{O}$ of 0 in $\mathrm{Sym}_{n\times n}(\mathbb{R})$, and a $\beta \in C^{k-2}(V_0, \mathcal{O})$, such that Φ is defined on $\mathcal{O} \times V_0$, $\beta(a) = 0$ and

$$\forall x \in V_0, \qquad \Phi(\beta(x), x) = 0. \tag{1.15.11}$$

Now set

$$\alpha(x) := \left(I_n + \frac{1}{2}Q(a)^{-1}\beta(x)\right)(x-a) \tag{1.15.12}$$

then $\alpha \in C^{k-2}(V_0, \mathbb{R}^n)$ with $\alpha(a) = 0$. By (1.15.7), (1.15.11) and (1.15.12), we have

$$\forall x \in V_0, \qquad \langle Q(a)\alpha(x)\,,\,\alpha(x)\,\rangle = \langle Q(x)(x-a)\,,\,x-a\,\rangle = f(x) - f(a). \tag{1.15.13}$$

For all small h, we have

$$\alpha(a+h) - \alpha(a) = \left(I_n + \frac{1}{2}Q(a)^{-1}(a+h)\right)h \tag{1.15.14}$$

so

$$\|\alpha(a+h) - \alpha(h) - h\| = \left\|\frac{1}{2}Q(a)^{-1}\beta(a+h)h\right\| \leq \frac{1}{2}\cdot\left\|Q(a)^{-1}\right\|\cdot\|\beta(a+h)\|. \tag{1.15.15}$$

Since $\beta(a) = 0$ and β is continuous, (1.15.15) implies that $\mathrm{d}\alpha(a) = \mathrm{Id}$. We may now apply the inverse mapping theorem to find a C^{k-2}-diffeomorphism Ψ from an open neighborhood V of $0 = \alpha(a)$ onto an open neighborhood of a contained in V_0 with $\alpha \circ \Psi = \mathrm{Id}$. Now setting $x = \Psi(y)$, $y \in V$, (1.15.13) becomes

$$\forall y \in V, \qquad \langle Hf(a)y\,,\,y\rangle = \langle Q(a)y\,,\,y\rangle = f\big(\Psi(y)\big) - f(a). \tag{1.15.16}$$

This establishes (1).

To establish (2), it suffices to note that one may find a fixed invertible matrix $A \in \mathrm{GL}_n(\mathbb{R})$ such that

$$A^t Hf(a)A = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}, \tag{1.15.17}$$

and it suffices to take $W := A^{-1}(V)$ and $\Theta(w) := \Psi(Aw)$ for all $w \in W$. □

Remark 1.15.3. — The local structure of the map f near a is reflected completely by (1.15.2). We have already seen two special cases, namely, when $p = 0$ (resp. $q = 0$), then f reaches its local maximum (resp. minimum) at a. In the case where both $p > 0$ and $q > 0$, then (1.15.2) says that $f(a)$ is not a local extreme of f near a, and the particular quadratic form on the right side of (1.15.2) completely describes the growth of f, after a change of coordinates provided by Θ. We call the non-degenerate critical point a a **saddle** point for f, where the terminology comes from the shape of $(x_1, x_2, x_1^2 - x_2^2)$ near $(0, 0, 0)$, corresponding to the case $n = 2$ and $p = q = 1$.

CHAPTER 2

RIEMANN INTEGRALS OF SEVERAL VARIABLES

> It is frequently claimed that Lebesgue integration is as easy to teach as Riemann integration. This is probably true, but I have yet to be convinced that it is as easy to learn. Under these circumstances, it is reasonable to introduce Riemann integration as an ad hoc tool to be replaced later by a more powerful theory, if required. If we only have to walk 50 metres, it makes no sense to buy a car.
>
> *A Companion To Analysis: a second first and first second course in analysis, page 197*
> T.W. Körner

After developing differential calculus in the setting of Banach spaces, it seems tempting to introduce the more flexible and powerful Lebesgue integration in this chapter on integral calculus. Nevertheless, unlike the differential calculus situation where we developed the theory in the setting of Banach spaces, to develop the theory of Lebesgue integrals here will certainly incur quite some technical challenges, which seems still premature for most students following this course. Our principle of prioritizing clarity and understanding dictates that we develop the more beginner-friendly Riemann integrals instead, which is sufficient to give a basic first introduction to analysis on smooth submanifolds in the last part of this course. This being said, we emphasize that the technical complication mentioned above is merely about which sets can be reasonably assigned a size and in what sense, and the theory of Jordan sets and Jordan contents that is intimately related to Riemann integrals already provides an easy to understand first answer. We therefore develop our theory of Riemann integrals with the novelty of allowing the integrand to take values in a general Banach space. As we will see by comparing with the treatment elsewhere, this slight more generality introduces no essential technical complication but brings what actually makes this theory work into the spotlight, just as our treatment of the differential calculus in the setting of normed/Banach spaces. Besides the advantages of clarity and a deeper understanding of the essence of the Riemann integral, we point out that the general Banach space valued Riemann integration (and of course its Lebesgue upgrade) is also widely used in modern mathematics, e.g. in holomorphic functional calculus on Banach algebras, in projection valued measure in spectral resolution of self/normal operators etc., and it is our opinion that the theory of Banach space valued Riemann integrals deserves a systematic treatment in these notes that it rarely receives in the literature.

2.1. Theories of Volumes: an Axiomatic Approach

A seemingly naive but profound question asks what is the size of geometric objects. In particular, since the ancient times, people want to be able to tell the length of a line, the area of a surface, the volume of a solid ball, et cetera. Our intuition tells us that for any dimension $n \in \mathbb{N}^+$, a bounded subset $A \subseteq \mathbb{R}^n$ of $\mathbb{R}^n$ should occupy some sort of n-dimensional volume $\nu(A)$.

For example, by analogy with the case of dimension 1, 2 and 3, we want to assign the number $\prod_{i=1}^n (b_i - a_i)$ as the n-dimensional volume of the n-dimensional box

$$\square_a^b := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid a_i \le x_i < b_i\}, \tag{2.1.1}$$

where $a = (a_i), b = (b_i) \in \mathbb{R}^n$, with $a_i \le b_i$ for each i. Note that if $a_{i_0} = b_{i_0}$ for some i_0, our notation for $\square_a^b$ as defined in (2.1.1) says that $\square_a^b = \emptyset$, but then again, it should have $0 = \prod_{i=1}^n (b_i - a_i)$ as its n-dimensional volume as expected.

Notation 2.1.1. — *For $a = (a_i), b = (b_i) \in \mathbb{R}^n$, we write $a \le b$ (resp. $a_i < b$) if $a_i \le b_i$ (resp. $a_i < b$) for each $i = 1, \dots, n$. For any $a, b \in \mathbb{R}^n$, sometimes we still use $\square_a^b$ to denote the set defined in* (2.1.1)*. We see that $\square_a^b \ne \emptyset$ if and only if $a < b$.*

Notation 2.1.2. — *Let A_i, $i \in I$ be a family of sets and A a set. By*

$$A = \bigsqcup_{i \in I} A_i, \tag{2.1.2}$$

we mean $A = \cup_{i \in I} A_i$ and $A_i \cap A_j = \emptyset$ whenever $i \ne j$.

Let $\mathcal{O}(\mathbb{R}^n)$ be a collection of bounded sets in $\mathbb{R}^n$ (to be thought of subsets that do have a well-defined n-dimensional volume), and $\nu : \mathcal{O}(\mathbb{R}^n) \to \mathbb{R}$ a map (to be thought of the volume map). From our intuition of volumes, we are led to pose two set of axioms. The first set of axioms concerns the properties of the collection $\mathcal{O}(\mathbb{R}^n)$ of bounded sets:

(O1) (Stability under finite union, intersection and difference) If A, B are in $\mathcal{O}(\mathbb{R}^n)$, then $A \cup B$, $A \cap B$ and $A \setminus B$ are all in $\mathcal{O}(\mathbb{R}^n)$.
(O2) (Stability under translation) If $A \in \mathcal{O}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$, then $A + x := \{a + x \mid a \in A\}$.
(O3) (The unit box axiom) We have $\square_{\mathbf{0}}^{\mathbf{1}} \in \mathcal{O}(\mathbb{R}^n)$, where $\mathbf{0} = (0, \dots, 0), \mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$.

The second set of axioms concerns the properties of volume map ν:

(ν1) (Non-negativity) We have $\nu(A) \ge 0$ for each $A \in \mathcal{O}(\mathbb{R}^n)$.
(ν2) (Finite additivity) If $A, B \in \mathcal{O}(\mathbb{R}^n)$ with $A \cap B = \emptyset$, then $\nu(A \cup B) = \nu(A) + \nu(B)$.
(ν3) (Translation invariance) If $A \in \mathcal{O}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$, then the translation $\nu(A + x) = \nu(A)$.

Definition 2.1.3. — A **theory of n-dimensional volumes**, or simply a **theory of volumes** if n is clear from context, is a pair $(\mathcal{O}(\mathbb{R}^n), \nu)$ where $\mathcal{O}(\mathbb{R}^n)$ is a collection of bounded sets in $\mathbb{R}^n$ satisfying (O1)-(O3), and $\nu : \mathcal{O}(\mathbb{R}^n) \to \mathbb{R}$ a map satisfying (ν1)-(ν3).

Before proceeding further, we list some remarks concerning these axioms.

Remark 2.1.4. — In any theory of volumes, we have $\emptyset \in \mathcal{O}(\mathbb{R}^n)$ and $\nu(\emptyset) = 0$. Indeed, $\emptyset = \square_{\mathbf{0}}^{1} \setminus \square_{\mathbf{0}}^{\mathbf{1}} \in \mathcal{O}(\mathbb{R}^n)$ by (O1) and (O3). Then by (ν2), we have

$$\nu(\emptyset) = \nu(\emptyset \cup \emptyset) = \nu(\emptyset) + \nu(\emptyset), \tag{2.1.3}$$

forcing $\nu(\emptyset) = 0$.

Remark 2.1.5. — Another consequence of the axioms is the so-called **monotonicity** of ν: if $A, B \in \mathcal{O}(\mathbb{R}^n)$ with $A \subseteq B$, then $\nu(A) \le \nu(B)$. Indeed, we have $B \setminus A \in \mathcal{O}(\mathbb{R}^n)$, hence from the decomposition $B = A \sqcup (B \setminus A)$, we have

$$\nu(B) = \nu(A) + \nu(B \setminus A) \ge \nu(A) \tag{2.1.4}$$

by (ν1) and (ν2).

Remark 2.1.6. — For any theory of volumes as above, suppose $A_i \in \mathcal{O}(\mathbb{R}^n)$, $i = 1, \dots, m$ and $A = \cup_{i=1}^m A_i$, we then have

$$\nu(A) \le \sum_{i=1}^n \nu(A_i), \tag{2.1.5}$$

and equality holds if all the A_i's are pairwise disjoint. Indeed, let $A_0 = \emptyset$ and set $B_i := A_i \setminus (A_1 \cup \dots \cup A_{i-1})$ for $i = 1, \dots, n$. Then we have $A_i = B_1 \cup \dots \cup B_i$ for each i, and all the B_i's are pairwise disjoint. It follows from the finite additivity that

$$\nu(A) = \sum_{i=1}^{n} \nu(B_i) \leq \sum_{i=1}^{n} \nu(A_i). \tag{2.1.6}$$

When A_i's are already pairwise disjoint, we have $B_i = A_i$ for each i, and we indeed have equality in (2.1.5).

Remark 2.1.7. — There is an obvious way of constructing theory of volumes: just take $\mathcal{O}(\mathbb{R}^n)$ to be the collection of all bounded subsets of $\mathbb{R}^n$ and $\nu \equiv 0$. Or proceed slightly further, we may take $\mathcal{O}(\mathbb{R}^n)$ to be any collection of bounded sets in $\mathbb{R}^n$ that satisfies (O1), (O2) and (O3), and $\nu \equiv 0$. Understandably, these theories are called **trivial**. As we will see presently, it requires some work, albeit not too much, to actually develop a nontrivial theory of volumes.

Remark 2.1.8. — There are more advanced theories of volumes, which satisfy stronger axioms, e.g. the Lebesgue measure theory[(1)], that lie beyond the scope of these notes. They may lead to a corresponding theory of integrals that is more powerful than the Riemann integrals to be developed later in this chapter. The interested students are encouraged to consult any book on Lebesgue integrals to his/her liking.

We will now derive some simple consequences of the axioms of a theory of volumes, as a warmup for later, slightly more involved results.

Definition 2.1.9. — Sets of the form $\square_a^b$ (possibly empty) are called **grid boxes**, or simply **boxes**, in $\mathbb{R}^n$.

Proposition 2.1.10. — *Let $\big(\mathcal{O}(\mathbb{R}^n), \nu\big)$ be a nontrivial theory of volumes, set $c := \nu\big(\square_{\mathbf{0}}^{\mathbf{1}}\big)$. Then the following hold.*

(1) *We must have $c > 0$.*
(2) *Any grid box is an element of $\mathcal{O}(\mathbb{R}^n)$.*
(3) *For any $a = (a_i), b = (b_i) \in \mathbb{R}^n$ with $a \leq b$, we have*

$$\nu\left(\square_a^b\right) = c \prod_{i=1}^{n} (b_i - a_i). \tag{2.1.7}$$

Proof. — (1). Suppose by contrapositive that $c = 0$, then we need to show that the theory $\big(\mathcal{O}(\mathbb{R}^n), \nu\big)$ must be trivial, i.e. $\nu \equiv 0$. Recall the Gauß function $[\cdot] : \mathbb{R} \to \mathbb{Z}$ that sends each real number r to its integral part $[r]$, i.e. $[r]$ is the largest integer not exceeding r, or equivalently, $[r]$ is the unique integer such that $[r_i] \leq r_i < [r_i] + 1$. For each vector $x = (x_i) \in \mathbb{R}^n$, we set $[x] := ([x_i]) \in \mathbb{Z}^n$. It is clear that $[x]$ is the unique $k \in \mathbb{Z}^n$ such that $x \in \square_{\mathbf{0}}^{\mathbf{1}} + k$. In particular, we have

$$\mathbb{R}^n = \bigsqcup_{k \in \mathbb{Z}^n} (\square_{\mathbf{0}}^{\mathbf{1}} + k). \tag{2.1.8}$$

Now take any $A \in \mathcal{O}(\mathbb{R}^n)$. Since A is bounded, there exists an $N \in \mathbb{N}^+$ such that $\|a\|_\infty \leq N$ for all $a = (a_i) \in A$. For any such $a \in A$, we see that $[a_i] \leq a_i < [a_i] + 1$, so $-N - 1 \leq a_i - 1 < [a_i] \leq a_i \leq N$ for each $i \in \{1, \ldots, n\}$. Now $a \in \square_0^1 + [a]$ together with (2.1.8) implies that

$$A \subseteq B := \bigsqcup_{\substack{k=(k_1,\ldots,k_n)\in\mathbb{Z}^n,\\ -N-1\leq k_i<N}} \left(\square_{\mathbf{0}}^{\mathbf{1}} + k\right). \tag{2.1.9}$$

In (2.1.9), the set B is written as a finite disjoint union of translation of the unit box $\square_{\mathbf{0}}^{\mathbf{1}}$, thus $B \in \mathcal{O}(\mathbb{R}^n)$ by (O2) and (O3); and the finite additivity of ν ($\nu 2$) together with its translation invariance ($\nu 3$) imply that

$$\nu(B) = \sum_{\substack{k=(k_1,\ldots,k_n)\in\mathbb{Z}^n,\\ -N-1\leq k_i<N}} \nu\left(\square_{\mathbf{0}}^{\mathbf{1}} + k\right) = \sum_{\substack{k=(k_1,\ldots,k_n)\in\mathbb{Z}^n,\\ -N-1\leq k_i<N}} \nu\left(\square_{\mathbf{0}}^{\mathbf{1}}\right) = (2N+1)^n c = 0. \tag{2.1.10}$$

[(1)] This will certainly be covered in a later course on real analysis.

By the non-negativity (ν1) and monotonicity (Remark 2.1.5), we have $0 \le \nu(A) \le \nu(B) = 0$, forcing $\nu(A) = 0$. Since $A \in \mathcal{O}(\mathbb{R}^n)$ is arbitrary in the above reasoning, we conclude that $\nu \equiv 0$ and the proof of (1) is complete.

(2). We first establish a claim which is a special case of what needs to be proved.

Claim. For all $a, b \in \mathbb{R}^n$ with $b - a < \mathbf{1}$, we have $\square_a^b \in \mathcal{O}(\mathbb{R}^n)$.

Before proving the claim, we first show how this claim imply (2). We wish to show that $\square_a^b \in \mathcal{O}(\mathbb{R}^n)$ for all $a = (a_i), b = (b_i) \in \mathbb{R}^n$. If $a < b$ does not hold, then $\square_a^b = \emptyset \in \mathcal{O}(\mathbb{R}^n)$ by Remark 2.1.4. So we suppose $a < b$. In this case, for each $i = 1, \dots, n$, we can choose $m_i \in \mathbb{N}^+$ large enough so that $l_i := (b_i - a_i)/m_i < 1$. Set

$$M(a,b) := \{k = (k_i) \in \mathbb{Z}^n \mid \forall i,\ 0 \le k_i < m_i\}. \tag{2.1.11}$$

Let $(e_i)_{i=1}^n$ be the standard basis for $\mathbb{R}^n$. For each $k = (k_i) \in M(a,b)$, define

$$\alpha(k) := \sum_{i=1}^n (a_i + k_i l_i) e_i \tag{2.1.12}$$

and

$$\beta(k) := \sum_{i=1}^n \big(a_i + (k_i + 1) l_i\big) e_i. \tag{2.1.13}$$

Then the fact that each $l_i < 1$ implies $\beta(k) - \alpha(k) < \mathbf{1}$ for any fixed k. But a routine check shows that

$$\square_a^b = \bigsqcup_{k \in M(a,b)} \square_{\alpha(k)}^{\beta(k)}. \tag{2.1.14}$$

Now each box on the right side of (2.1.14) is an element of $\mathcal{O}(\mathbb{R}^n)$ and $M(a,b)$ is clearly finite, hence $\square_a^b \in \mathcal{O}(\mathbb{R}^n)$ by (O1) as desired.

We now establish the claim. Since $\square_a^b = \square_{\mathbf{0}}^{b-a} + a$, by stability of translations of $\mathcal{O}(\mathbb{R}^n)$, it suffices to show that $\square_{\mathbf{0}}^c \in \mathcal{O}(\mathbb{R}^n)$ whenever $c \in \mathbb{R}^n$ with $c < \mathbf{1}$. By translation invariance again and (O3) $\square_{\mathbf{0}}^{\mathbf{1}} \in \mathcal{O}(\mathbb{R}^n)$, we have

$$\square_{c-\mathbf{1}}^c = \square_{\mathbf{0}}^{\mathbf{1}} + (c - \mathbf{1}) \in \mathcal{O}(\mathbb{R}^n). \tag{2.1.15}$$

Hence by (O1), we indeed have

$$\square_{\mathbf{0}}^c = \square_{\mathbf{0}}^{\mathbf{1}} \cap \square_{c-\mathbf{1}}^c \in \mathcal{O}(\mathbb{R}^n). \tag{2.1.16}$$

(3). If $a < b$ does not hold, i.e. if $a_i = b_i$ for some $i \in \{1, \dots, n\}$, then $\square_a^b = \emptyset$ and both sides of (2.1.7) are 0 Remark 2.1.4. Hence we assume $a < b$ without loss of generality. Since $\square_a^b = \square_{\mathbf{0}}^{b-a} + a$, by (O2) and ($\nu$3), it suffices to prove (3) for the special case where $a = \mathbf{0}$. It thus suffices now to show that

$$\forall d = (d_i) \in \mathbb{R}_{>0}^n \subseteq \mathbb{R}^n, \quad \nu\left(\square_{\mathbf{0}}^d\right) = c \prod_{i=1}^n d_i. \tag{2.1.17}$$

We first show (2.1.17) when $d \in \mathbb{Q}_{>0}^n$. By choosing a common denominator for each d_i, we may suppose that there exists an $N, m_1, \dots, m_n \in \mathbb{N}^+$ such that $d_i = \frac{m_i}{N}$ for each i. Set $m :== (m_1, \dots, m_n)$. Moreover, define

$$\gamma(N) := \frac{1}{N}\mathbf{1} = \left(\frac{1}{N}, \dots, \frac{1}{N}\right) \in \mathbb{R}^n, \tag{2.1.18}$$

$$K(m) := \big\{k = (k_i) \in \mathbb{N}^n \bigm| k < m\big\}, \tag{2.1.19}$$

and

$$\forall k \in K(m), \qquad \delta(k) = \sum_{i=1}^n \frac{k_i}{N} e_i \in \mathbb{R}^n. \tag{2.1.20}$$

It is clear from $d_i = \frac{m_i}{N}$ for each i that

$$\square_{\mathbf{0}}^d = \bigsqcup_{k \in K(m)} \square_{\delta(k)}^{\delta_k + \gamma(N)}. \tag{2.1.21}$$

Now let $c_N := \nu\left(\square_{\mathbf{0}}^{\gamma(N)}\right)$. By (O2) and ($\nu$3), (2.1.21) implies that

$$\nu\left(\square_0^d\right) = m_1 m_2 \cdots m_n c_N, \tag{2.1.22}$$

since $|K(m)| = m_1 m_2 \cdots m_n$. On the other hand, setting $m_1 = \cdots = m_n = N$ in (2.1.22) shows that

$$c = \nu\left(\square_{\mathbf{0}}^{\mathbf{1}}\right) = N^n c_N. \tag{2.1.23}$$

By (1), $c > 0$, thus (2.1.22) and (2.1.23) together yield

$$\nu\left(\square_{\mathbf{0}}^d\right) = c\prod_{i=1}^n \frac{m_i}{N} = c\prod_{i=1}^n d_i. \tag{2.1.24}$$

Hence (2.1.17) holds for all $d \in \mathbb{Q}^n_{>0}$.

Now consider arbitrary $d = (d_i) \in \mathbb{R}^n_{>0}$. For any $0 < p_i < d_i < q_i < 0$ with $p_i, q_i \in \mathbb{Q}$ for each i, set $p := (p_i)$ and $q := (q_i)$, we have

$$\square_{\mathbf{0}}^p \subseteq \square_{\mathbf{0}}^d \subseteq \square_{\mathbf{0}}^q, \tag{2.1.25}$$

thus by the rational case and the monotonicity of ν, we have

$$c\prod_{i=1}^n p_i = \nu\left(\square_{\mathbf{0}}^p\right) \le \nu\left(\square_{\mathbf{0}}^d\right) \le \nu\left(\square_{\mathbf{0}}^q\right) = c\prod_{i=1}^n q_i. \tag{2.1.26}$$

Since p_i and q_i can be made arbitrarily closed to d_i, (2.1.26) forces (2.1.17) and the proof is complete. □

Remark 2.1.11. — Note that Proposition 2.1.10 does not claim at all that there indeed exists a nontrivial theory of volumes, it merely deduces some the properties this theory must possess.

We end this section with the introduction of the notion of paved sets and some of their basic properties which are natural after Proposition 2.1.10.

Definition 2.1.12. — A is called **paved set** in $\mathbb{R}^n$ is a finite union of boxes in $\mathbb{R}^n$. We use $\mathcal{P}(\mathbb{R}^n)$ to denote the collection of all paved sets in $\mathbb{R}^n$.

Corollary 2.1.13. — *For any theory $\left(\mathcal{O}(\mathbb{R}^n, \nu)\right)$ of volumes, the following hold.*

(1) *We have $\mathcal{P}(\mathbb{R}^n) \subseteq \mathcal{O}(\mathbb{R}^n)$.*
(2) *If B and B' are boxes, then $B \cap B'$ is a box, and $B \setminus B'$ is a finite disjoint union of boxes.*
(3) *Each $A \in \mathcal{P}(\mathbb{R}^n)$ is a finite* disjoint *union of boxes.*
(4) *The collection $\mathcal{P}(\mathbb{R}^n)$ contains the unit cube $\square_{\mathbf{0}}^{\mathbf{1}}$ and is stable under taking finite union, finite intersection, set difference and translations. In particular, $\left(\mathcal{P}(\mathbb{R}^n), \nu\restriction_{\mathcal{P}(\mathbb{R}^n)}\right)$ is a theory of volumes.*
(5) *If $A = \sqcup_{j=1}^m B_j$ where each B_j is a box, then*

$$\nu(A) = \sum_{j=1}^m \nu(B_j). \tag{2.1.27}$$

In particular, the theory $\left(\mathcal{P}(\mathbb{R}^n), \nu\restriction_{\mathcal{P}(\mathbb{R}^n)}\right)$ of volumes is completely determined by the number $c = \nu(\square_{\mathbf{0}}^{\mathbf{1}})$.

Proof. — (1) is clear since $\mathcal{O}(\mathbb{R}^n)$ contains all boxes and is stable under taking finite unions.

(2). Let $B = \square_a^b$ and $B' = \square_c^d$, with $a, b, c, d \in \mathbb{R}^n$. To facilitate discussion, for each $i = 1, \ldots, n$, let $I_i := [a_i, b_i[$ (could be empty) and $J_i := [c_i, d_i[$, so that $\square_a^b = I_1 \times \cdots \times I_n$ and $\square_c^d = J_1 \times \cdots \times J_n$. It is clear that

$$B \cap B' = \square_a^b \cap \square_b^d = (I_1 \cap J_i) \times \cdots \times (I_n \cap J_n), \tag{2.1.28}$$

which is still a box. Moreover, A routine check shows that we have

$$B \setminus B' = \square_a^b \setminus \square_b^d = \bigsqcup_{k=1}^n (I_1 \setminus J_1) \times \cdots \times (I_k \setminus J_k) \times I_{k+1} \times \cdots \times I_n, \tag{2.1.29}$$

which implies $B \setminus B'$ is a disjoint union of n boxes since each $I_k \setminus J_k$ is either of the form $[r, s[$, or of the form $[r, s[\sqcup [u, v[$.

(3). For each $m \in \mathbb{N}^+$, we propose

Statement $P(m)$: if $A \in \mathcal{P}(\mathbb{R}^n)$ can be written as the union of at most m boxes, then A is a finite disjoint union of boxes.

It is clear that $P(1)$ holds, and (3) will be proved if $P(m)$ holds for all $m \in \mathbb{N}^+$. We therefore establish by induction and suppose that $P(n)$ holds for some $n \in \mathbb{N}^+$ and it suffices to show that $P(n+1)$ holds. For this, it suffices to show that if $A = \cup_{j=1}^{n+1} B_j$ with each B_j being a box, then A is a finite disjoint union of boxes. Set $A_2 := \cup_{j=2}^{n+1} B_j$, which is union of n boxes, hence by $P(n)$, there exists a finite sequence of boxes $B_1', \dots, B_l'$, such that

$$A_2 = \bigsqcup_{j=1}^{l} B_j'. \tag{2.1.30}$$

Now

$$A = B_1 \cup A_2 = B_1 \sqcup (A_2 \setminus B_1) = B_1 \sqcup \left(\bigsqcup_{j=1}^{l} (B_j' \setminus B_1) \right). \tag{2.1.31}$$

In the finite disjoint union on the right side of (2.1.31), each $B_j' \setminus B_1$ is a finite disjoint union of boxes by (2), it follows from (2.1.31) again that A itself is a finite disjoint of boxes.

(4). By definition $\square_{\mathbf{0}}^{\mathbf{1}} \in \mathcal{P}(\mathbb{R}^n)$ and $\mathcal{P}(\mathbb{R}^n)$ is stable under finite unions. By (2) and (3), it is easy to see that $\mathcal{P}(\mathbb{R}^n)$ is stable under taking finite intersections and translates. It remains to show that $P_1 \setminus P_2 \in \mathcal{P}(\mathbb{R}^n)$ if $P_1, P_2 \in \mathcal{P}(\mathbb{R}^n)$. By (3), we may write

$$P_1 = \bigsqcup_{j=1}^{m} C_j \tag{2.1.32}$$

and

$$P_2 = \bigsqcup_{k=1}^{l} D_k \tag{2.1.33}$$

where each C_j(resp. D_k) is a box. Then, by (2), every $C_j \setminus D_k \in \mathcal{P}(\mathbb{R}^n)$, and the stability of finite unions and finite intersections of $\mathcal{P}(\mathbb{R}^n)$ now implies that we indeed have

$$P_1 \setminus P_2 = \bigsqcup_{j=1}^{m} (C_j \setminus P_2) = \bigsqcup_{j=1}^{m} \left(\bigcap_{k=1}^{l} (C_j \setminus D_k) \right) \in \mathcal{P}(\mathbb{R}^n). \tag{2.1.34}$$

(5). This follows directly from (3), (4) and Definition 2.1.3, as well as (2.1.7) in Proposition 2.1.10. $\square$

Remark 2.1.11 now raises the following important fundamental question.

Question. Fix any $c > 0$, we define ν via (2.1.27) and (2.1.17) as well as the translation invariance of ν, do we get a well-defined ν, i.e. for each $A \in \mathcal{P}(\mathbb{R}^n)$, is $\nu(A)$ in (2.1.27) independent of the choice of the finite disjoint union decomposition?

One can resolve the above question by a careful analysis of the finite disjoint decomposition of A, e.g. take two such decompositions, find a common refinement, and show carefully that the two original decomposition can both be obtained from this common refinement by merging certain of its boxes. This idea is simple, yet its rigorous formulation can be quite tedious. We therefore switched to an approach that simultaneously resolved this issue and paves the way to some important theoretical development. It is clear that with different choices of c, these are the minimal models of theory of volumes, and to resolve the above question, it is enough to show the existence of any theory of volumes for each $c > 0$.

2.2. Volumes of Paved Sets via Successive Riemann Integrals

Recall the notion of characteristic functions. Let A be a subset of X, the characteristic function χ_A of A defined on X is the map sending $\chi_A(x)$ to 1 if $x \in A$ and 0 otherwise. Clearly the characteristic function χ_A and A are in bijective correspondence since $A = \chi_A^{-1}(1)$. Consideration of characteristic functions gives a way to transform problems about sets (e.g. how to construct a theory of volumes) into problems about functions (e.g. how to build some sort of theory of integrals).

In view of the question raised at the end of § 2.1, we will now exploit the one-dimensional Riemann integrals we are already familiar with in our first year's study, except that we will use it repeated since we want to integrate out all of the n-variables in $\mathbb{R}^n$. If successful, this will produce some model of a theory of volumes by considering the characteristic functions of sets. If one tries this, immediately, one encounters the problem of integrability, so we simply circumvent this by using the upper and lower Riemann integrals instead, and demand them to be equal. Another small problem is that we need to fix an order of the n-variables with respect to which to take the one-dimensional upper/lower Riemann integrals.

We now fix an $L > 0$ and consider the region

$$\overline{B}_\infty(L) := \left\{x \in \mathbb{R}^n \;\middle|\; \|x\|_\infty \le L\right\} = [-L, L]^n, \tag{2.2.1}$$

which is a compact cube in $\mathbb{R}^n$. Take any *bounded* function $f : \overline{B}_\infty(L) \to \mathbb{R}$. We want to obtain a real number from f by repeatedly taking the lower (upper) Riemann integrals, for which we need to fix an order of the variables. We model this order by an element σ in the permutation group S_n, i.e. the group of all bijection from $\{1, \ldots, n\}$ onto itself. We set

$$\underline{\mathcal{I}}_\sigma(f) := \underline{\int}_{-L}^{L}\left(\underline{\int}_{-L}^{L}\left(\underline{\int}_{-L}^{L}\left(\cdots\underline{\int}_{-L}^{L} f(x_1, \ldots, x_n)\mathrm{d}x_{\sigma(n)}\cdots\right)\mathrm{d}x_{\sigma(3)}\right)\mathrm{d}x_{\sigma(2)}\right)\mathrm{d}x_{\sigma(1)}, \tag{2.2.2}$$

i.e. $\underline{\mathcal{I}}(f)$ is the real number we obtained from f by successively taking the lower Riemann integral on the interval $[-L, L]$, first with respect to x_n, then with respect to x_{n-1}, …, until lastly with respect to x_1. The reason we choose the lower Riemann integral in (2.2.2) is that it always exists, no matter what f is as long as it is bounded. Similar, we may define $\overline{\mathcal{I}_\sigma}(f)$ by replacing the lower Riemann integrals with the upper Riemann integrals in (2.2.2), i.e.

$$\overline{\mathcal{I}}_\sigma(f) := \overline{\int}_{-L}^{L}\left(\overline{\int}_{-L}^{L}\left(\overline{\int}_{-L}^{L}\left(\cdots\overline{\int}_{-L}^{L} f(x_1, \ldots, x_n)\mathrm{d}x_{\sigma(n)}\cdots\right)\mathrm{d}x_{\sigma(3)}\right)\mathrm{d}x_{\sigma(2)}\right)\mathrm{d}x_{\sigma(1)}, \tag{2.2.3}$$

Definition 2.2.1. — Fix $\sigma \in S_n$ and L as above. We say that a bounded real-valued function f defined on $\overline{B}_\infty(L)$ is **successively Riemann integral in the order** σ, if $\underline{\mathcal{I}}_\sigma(f) = \overline{\mathcal{I}}_\sigma(f)$. In this case, we set $\mathcal{I}_\sigma(f) := \underline{\mathcal{I}}_\sigma(f) = \overline{\mathcal{I}}_\sigma(f)$, and call it the **value of successive Riemann integrals of** f **in the order** σ. We will use the symbol $\mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right)$ to denote the space of all such f.

Before proceeding to Proposition 2.2.3, we recall the following basic results.

Lemma 2.2.2. — *Let $[a, b]$ be a compact interval. If $f, g : [a, b] \to \mathbb{R}$ are two arbitrary bounded maps, then*

$$\underline{\int}_a^b (f + g) \ge \underline{\int}_a^b f + \underline{\int}_a^b g, \tag{2.2.4}$$

and

$$\overline{\int}_a^b (f + g) \le \overline{\int}_a^b f + \overline{\int}_a^b g. \tag{2.2.5}$$

Proof. — Note the fact that for each closed subinterval J of I, we have

$$\begin{aligned} \forall x \in J, \qquad & \sup_{x\in J}\big(f(x) + g(x)\big) \le \sup_{x\in J} f(x) + \sup_{x\in J} g(x) \\ \text{and} \qquad & \inf_{x\in J}\big(f(x) + g(x)\big) \ge \inf_{x\in J} f(x) + \inf_{x\in J} g(x). \end{aligned} \tag{2.2.6}$$

Now the lemma is clear from the definition of the upper (resp. lower) Riemann integrals as the greatest lower bound (resp. lowest upper bound) of all Darboux upper (resp. lower) sums. □

Proposition 2.2.3. — *Let $\sigma \in S_n$ and f be as above. Then the following hold.*

(1) *The space* $\mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right)$ *is a linear subspace of the space of all bounded real-valued functions on* $\overline{B}_\infty(L)$.

(2) *The map* $\mathcal{I}_\sigma : \mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right) \to \mathbb{R}$, $f \mapsto \mathcal{I}_\sigma(f)$ *is linear.*

Proof. — It is clear the $\mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right)$ is stable under taking scalar multiplications and that $\mathcal{I}_\sigma$ commutes with scalar multiplications. It remains to show that if $f, g \in \mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right)$, then $f + g \in \mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right)$ and $\mathcal{I}_\sigma(f+g) = \mathcal{I}_\sigma(f) + \mathcal{I}_\sigma(g)$.

Repeated use of Lemma 2.2.2 imply that

$$\overline{\mathcal{I}}_\sigma(f+g) \le \overline{\mathcal{I}}_\sigma(f) + \overline{\mathcal{I}}_\sigma(g) = \underline{\mathcal{I}}_\sigma(f) + \underline{\mathcal{I}}_\sigma(g) \le \underline{\mathcal{I}}_\sigma(f+g). \tag{2.2.7}$$

This forces all inequalities in (2.2.7) is actually equal, and $f + g \in \mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right)$ with $\mathcal{I}_\sigma(f+g) = \mathcal{I}_\sigma(f) + \mathcal{I}_\sigma(g)$, just as desired. $\square$

Remark 2.2.4. — We can already build our first theory of volumes to settle the question at the end of § 2.1. First, if $B = \square_a^b$ is any box in $\mathbb{R}^n$, by choosing L large enough, we may suppose $B \subseteq \overline{B}_\infty(L)$, and an easy computation shows that

$$\underline{\mathcal{I}}_\sigma(\chi_B) = \overline{\mathcal{I}}_\sigma(\chi_B) = \prod_{i=1}^n (b_i - a_i) \tag{2.2.8}$$

so $\chi_B \in \mathcal{SR}_\sigma\left(\overline{B}_\infty(L)\right)$ with $\mathcal{I}_\sigma(\chi_B) = \prod_{i=1}^n (b_i - a_i)$, which is independent of L and of $\sigma \in S_n$. Now if $A \in \mathcal{P}(\mathbb{R}^n)$ with a finite disjoint decomposition $A = \sqcup_{k=1}^m B_k$ into boxes, we may choose $L > 0$ large enough so that the bounded set A, and all of its subset including every B_k, are contained in $\overline{B}_\infty(L)$. The decomposition of A and Proposition 2.2.3 imply that

$$\chi_A = \sum_{k=1}^m \chi_{B_k} \in \mathcal{SR}_\sigma\left(\overline{B_\infty(L)}\right), \tag{2.2.9}$$

and

$$\mathcal{I}_\sigma(\chi_A) = \sum_{k=1}^m \mathcal{I}_\sigma(\chi_{B_k}). \tag{2.2.10}$$

The left side of (2.2.10) is independent of our choice of $L > 0$ and $\sigma \in S_n$ since the right side is so. We denote this common value (when $L > 0$ is large and $\sigma \in S_n$) by $\mathcal{I}(\chi_A)$. Clearly $\mathcal{I}(\chi_{B+x}) = \mathcal{I}(\chi_B)$ for all $x \in \mathbb{R}^n$ by (2.2.8), it follows from (2.2.10) as well as (apply the finite additivity of ν again to the decomposition $A + x = \sqcup_{k=1}^m (B_k + x)$)

$$\mathcal{I}(\chi_{A+x}) = \sum_{k=1}^m \mathcal{I}(\chi_{B_k+x}) \tag{2.2.11}$$

that

$$\forall A \in \mathcal{P}(\mathbb{R}^n), \forall x \in \mathbb{R}^n, \qquad \mathcal{I}(\chi_{A+x}) = \mathcal{I}(\chi_A). \tag{2.2.12}$$

For any fixed $c > 0$. It is now trivial to check that the pair $\left(\mathcal{P}(\mathbb{R}^n), \nu_c\right)$, where $\nu_c : \mathcal{P}(\mathbb{R}^n) \to \mathbb{R}$ is the map defined by $\nu_c(A) := c\mathcal{I}(\chi_A)$, is indeed a n-dimensional theory of volumes, and the question at the end of § 2.1 is settled.

2.3. More on Successive Riemann Integrals

We still fix an $n \in \mathbb{N}^+$ in this section. Remark 2.2.4 settles the question at the end of § 2.1 by constructing a class of theories of volumes $\left(\mathcal{P}(\mathbb{R}^n), \nu_c\right)$, one for each constant c. Our discussion actually showed the following result.

Proposition 2.3.1. — *For each $c \in [0,+\infty[$, the pair $\big(\mathcal{P}(\mathbb{R}^n), \nu_c\big)$, with $\nu_c : \mathcal{P}(\mathbb{R}^n) \to \mathbb{R}$ defined by $\nu_c(A) := c\mathcal{I}(\chi_A)$ for all $A \in \mathcal{P}(\mathbb{R}^n)$, is a theory of volumes. Moreover, this theory is minimal in the sense that for any theory of volumes $\big(\mathcal{O}(\mathbb{R}^n), \nu\big)$, we must have $\mathcal{P}(\mathbb{R}^n) \subseteq \mathcal{O}(\mathbb{R}^n)$, and $\nu \restriction_{\mathcal{P}(\mathbb{R}^n)} = \nu_c$, where $c = \nu\left(\square_{\mathbf{0}}^{\mathbf{1}}\right)$.* □

The minimal theories $\big(\mathcal{P}(\mathbb{R}^n), \nu_c\big)$, $c \in [0,+\infty[$, while being of fundamental importance for the development of a good theory of volumes, has an obvious drawback: it can only assign a volume to paved sets for which all of the sides of the composing boxes must be parallel to the coordinate axis. We wish to obtain a theory that can answer basic questions such as what is the volume of the unit ball

$$B_n := \big\{x \in \mathbb{R}^n \;\big|\; \|x\| \le 1\big\} \subseteq \mathbb{R}^n. \tag{2.3.1}$$

This motivates us to look for theories of volumes that can assign a volume to simple geometric objects such as B_n. Consideration in Remark 2.2.4 already indicates us that one might obtain such a theory by considering successive Riemann integrals. In fact, for each fixed $\sigma \in S_n$, we may obtain a theory of successive Riemann integration on $\mathbb{R}^n$ with respect to the order σ, and consequently, by passing to characteristic functions, a theory of volumes. We now elaborate on how this can be done.

Definition 2.3.2. — Let $f : X \to A$ be a map from a topological space X to an abelian group A, the **support** of f, denoted by $\operatorname{supp}(f)$, is defined by

$$\operatorname{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}}, \tag{2.3.2}$$

which is clearly the smallest closed set in X outside which f vanishes.

Proposition 2.3.3. — *Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function with compact support. For all $L_1, L_2 > 0$ with* $\operatorname{supp}(f) \subseteq [-L_1, L_1] \cap [-L_2, L_2]$, *we have*

$$\overline{\int}_{-L_1}^{L_1} f = \overline{\int}_{-L_2}^{L_2} f, \tag{2.3.3}$$

and

$$\underline{\int}_{-L_1}^{L_1} f = \underline{\int}_{-L_2}^{L_2} f. \tag{2.3.4}$$

Proof. — By an obvious manipulation of Darboux upper/lower sums, we see that once chosen an $L > \max(L_1, L_2)$, we have

$$\overline{\int}_{-L}^{L} f = \overline{\int}_{-L}^{-L_1} f + \overline{\int}_{-L_1}^{L_1} f + \overline{\int}_{L_1}^{L} f = 0 + \overline{\int}_{-L_1}^{L_1} f + 0 = \overline{\int}_{-L_1}^{L_1} f, \tag{2.3.5}$$

and similarly

$$\overline{\int}_{-L_2}^{L_2} f = \overline{\int}_{-L}^{L} f = \overline{\int}_{-L_1}^{L_1} f. \tag{2.3.6}$$

The case for lower Riemann integrals is similar. □

Definition 2.3.4. — For a bounded compactly supported $f : \mathbb{R} \to \mathbb{R}$, we define the **upper Riemann integral** of f, denote by $\overline{\mathcal{I}}(f)$, $\overline{\int}_{\mathbb{R}} f$, $\overline{\int} f$, ..., to be $\overline{\int}_{-L}^{L} f$ whenever $\operatorname{supp}(f) \subseteq [-L, L]$. Because of Proposition 2.3.3, this is independent of the choice of L and hence is well-defined. Similarly, we define the **lower Riemann integral** of f and denote it by $\underline{\mathcal{I}}(f)$, $\underline{\int}_{\mathbb{R}} f$, $\underline{\int} f$, etc. We say $f : \mathbb{R} \to \mathbb{R}$ is **Riemann integrable** if f is bounded with compact support and $\overline{\int} f = \underline{\int} f$, in which case, call the common number $\overline{\int} f = \underline{\int}$ **the Riemann integral** of f and denote it by $\mathcal{I}(f)$, $\int f$, $\int_{\mathbb{R}} f$ etc. We denote by $\mathcal{R}(\mathbb{R})$ the space of all bounded compactly supported real-valued Riemann integrable functions on $\mathbb{R}$.

An important property we wish to establish for our theory of (successive) integrals is the so-called translation-invariance, for which we introduce the following definition.

Definition 2.3.5. — Fix an $n \in \mathbb{N}$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a map. The **translation** of f by $x \in \mathbb{R}^n$, denoted by $\tau_x f : \mathbb{R}^n \to \mathbb{R}$, is the map defined by $(\tau_x f)(t) = f(t - x)$.

The following properties concerning translation of functions can be established by routine verifications.

Proposition 2.3.6. — *The following hold.*

(1) *(Group action property) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a map ($n \in \mathbb{N}$). We have $\tau_0 f = f$, and $\tau_{x+y} = \tau_x(\tau_y f) = \tau_y(\tau_x f)$ for all $x, y \in \mathbb{R}^n$.*

(2) *(Translation of the graph) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a map ($n \in \mathbb{N}$) and fix an $a \in \mathbb{R}^n$, we have*

$$\operatorname{graph}(\tau_a f) = \operatorname{graph}(f) + (a, 0) \subseteq \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}. \tag{2.3.7}$$

In particular,

$$\operatorname{supp}(\tau_a f) = \operatorname{supp}(f) + a \subseteq \mathbb{R}^n. \tag{2.3.8}$$

(3) *(Linearity of the translation operator) Let f, g be maps from $\mathbb{R}^n$ to $\mathbb{R}$ and $x \in \mathbb{R}^n$, $\lambda, \mu \in \mathbb{R}$, then*

$$\tau_x(\lambda f + \mu g) = \lambda \tau_x f + \mu \tau_x g. \tag{2.3.9}$$

Proof. — Exercise. □

Lemma 2.3.7. — *If $f : \mathbb{R} \to \mathbb{R}$ is bounded with compact support, then for any $x \in \mathbb{R}$, we have*

$$\underline{\int} \tau_x f = \underline{\int} f \tag{2.3.10}$$

and

$$\overline{\int} \tau_x f = \overline{\int} f. \tag{2.3.11}$$

Proof. — The idea for these facts is simple: the corresponding Darboux sums are translation-invariant. We now elaborate this idea into a rigorous proof, with a small trick to make the proof slightly shorter.

Let $[a, b]$ be a compact interval in $\mathbb{R}$ that contains $\operatorname{supp}(f)$. Then

$$[a + x, b + x] = [a, b] + x \supseteq \operatorname{supp}(f) + x = \operatorname{supp}(\tau_x f). \tag{2.3.12}$$

Choose an arbitrary $\varepsilon > 0$. There exists a partition $\Delta = \{t_0 = a < \cdots < t_m = b\}$ of $[a, b]$, such that the corresponding Darboux lower sum satisfies

$$\underline{\int} f = \underline{\int_a^b} f \geq \underline{S}(f, \Delta) := \sum_{i=1}^m \Big(\inf f([t_{i-1}, t_i])\Big)(t_i - t_{i-1}) \geq -\varepsilon + \underline{\int_a^b} f = -\varepsilon + \underline{\int} f. \tag{2.3.13}$$

Now consider the partition $\Delta + x = \{s_0 = a + x < \cdots s_m = b + x\}$ of $[a+x, b+x]$, where each $s_i = t_i + x$. Since

$$\begin{aligned} \forall i = 1, \ldots, m, \qquad \inf(\tau_x f)([s_{i-1}, s_i]) &= \inf_{r \in [s_{i-1}, s_i]} (\tau_x f)(r) \\ &= \inf_{r \in [s_{i-1}, s_i]} f(r - x) = \inf_{t \in [t_{i-1}, t_i]} f(t) = \inf f([t_{i-1}, t_i]), \end{aligned} \tag{2.3.14}$$

and clearly that $s_i - s_{i-1} = t_i - t_{i-1}$ for each i, we have

$$\begin{aligned} \underline{\int} \tau_x f &= \underline{\int_{a+x}^{b+x}} f \\ &\geq \underline{S}(\tau_x f, \Delta + x) := \sum_{i=1}^m \left(\inf(\tau_x f)([s_{i-1}, s_i])\right)(s_i - s_{i-1}) \\ &= \underline{S}(f, \Delta) \geq -\varepsilon + \underline{\int} f. \end{aligned} \tag{2.3.15}$$

Since $\varepsilon > 0$ is arbitrary, from (2.3.15), we conclude that

$$\underline{\int} \tau_x f \geq \underline{\int} f. \tag{2.3.16}$$

In particular, apply (2.3.16) to $\tau_x f$ instead of f with x replaced by $-x$, we get

$$\underline{\int} f = \underline{\int} \tau_{-x}(\tau_x f) \geq \underline{\int} \tau_x f. \tag{2.3.17}$$

Now (2.3.10) follows from (2.3.16) and (2.3.17).

For the upper Riemann integral, we have

$$\overline{\int}\tau_x f = -\underline{\int}(-\tau_x f) = -\underline{\int}(\tau_x(-f)) = -\underline{\int}(-f) = \overline{\int} f, \tag{2.3.18}$$

which completes the proof. □

We track here an isolated curious result, which the hurried reader can safely omit. For the relevance of this result, see Remark 2.3.16.

Proposition 2.3.8. — *If $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are bounded with compact support, then*

(1) *For each compact interval $I = [a, b]$ in $\mathbb{R}$, we have*

$$\begin{aligned} \forall r, s \in I, \qquad & f(r)g(r) - f(s)g(s) = f(r)[g(r) - g(s)] + [f(r) - f(s)]g(s) \\ & \leq |f(r)| \cdot |g(r) - g(s)| + |f(r) - f(s)| \cdot |g(s)| \\ & \leq M_f\Big(\sup g(I) - \inf g(I)\Big) + M_g\Big(\sup f(I) - \inf f(I)\Big), \end{aligned} \tag{2.3.19}$$

where $M_f > 0$ (resp. $M_g > 0$) is a bound for the range of f (resp. g) in the sense that $|f(x)| \leq M_f$ (resp. $|g(x)| \leq M_g$) for all $x \in \mathbb{R}$. In particular, we have

$$\sup(fg)(I) - \inf(fg)(I) \leq M_f\Big(\sup g(I) - \inf g(I)\Big) + M_g\Big(\sup f(I) - \inf f(I)\Big). \tag{2.3.20}$$

(2) *We have*

$$\overline{\int} fg - \underline{\int} fg \leq M_g\left(\overline{\int} f - \underline{\int} f\right) + M_f\left(\overline{\int} g - \underline{\int} g\right). \tag{2.3.21}$$

In particular, if $f, g \in \mathcal{R}(\mathbb{R})$, then $fg \in \mathcal{R}(\mathbb{R})$.

Proof. — We give a rather detailed proof sketch and leave as an exercise to fill in more details.

(1). In (2.3.19), the first equality follows from a routine check, the first inequality from the triangle inequality and $x \leq |x|$ for all $x \in \mathbb{R}$, and the last inequality from the first. (2.3.20) follows directly from (2.3.19) and the definition of $\sup(fg)(I)$ and $\inf(fg)(I)$.

(2). For any partition $\Delta = \{t_0 = a < \cdots < t_m = b\}$, where $[a, b]$ is a fixed compact interval containing $\mathrm{supp}(f) \cup \mathrm{supp}(g)$, we have, by (2.3.20) applied to each of the subinterval $[t_{i-1}, t_i]$, that

$$\overline{S}(fg, \Delta) - \underline{S}(fg, \Delta) \leq M_f\big(\overline{S}(g, \Delta) - \underline{S}(g, \Delta)\big) + M_g\big(\overline{S}(f, \Delta) - \underline{S}(f, \Delta)\big). \tag{2.3.22}$$

Taking limits along the refinement of partitions in (2.3.22) establishes (2.3.21).

If $f, g \in \mathcal{R}(R)$, we then obtain

$$0 \leq \overline{\int} fg - \underline{\int} fg \leq 0, \tag{2.3.23}$$

which forces all inequalities in (2.3.23) to be equality, and $fg \in \mathcal{R}(\mathbb{R})$. □

Proposition 2.3.9. — *Consider the map*

$$\begin{aligned} \int : \mathcal{R}(\mathbb{R}) &\to \mathbb{R} \\ f &\mapsto \int f. \end{aligned} \tag{2.3.24}$$

The following hold.

(1) *(Linear structure of $\mathcal{R}(\mathbb{R})$) The space $\mathcal{R}(\mathbb{R})$ is a linear subspace of the vector space of all maps from $\mathbb{R}^n$ to $\mathbb{R}$.*
(2) *(Multiplications structure of $\mathcal{R}(\mathbb{R})$) The space $\mathcal{R}(\mathbb{R})$ is stable under taking products of functions.*
(3) *(Linearity of the integral) The integral map $\int : \mathcal{R}(\mathbb{R}) \to \mathbb{R}$ is linear.*
(4) *(Positivity of the integral) If $f \in \mathcal{R}(\mathbb{R})$ is non-negative, then $\int f \geq 0$.*
(5) *(Translation invariance of the integral) For all $f \in \mathcal{R}(\mathbb{R})$ and $x \in \mathbb{R}$, we have $\tau_x f \in \mathcal{R}(\mathbb{R})$ and $\int \tau_x f = \int f$.*

Proof. — We first simultaneously establish (1) and (3). It is obvious that if $c \in \mathbb{R}$ and $f \in \mathcal{R}(\mathbb{R})$, then $cf \in \mathcal{R}(\mathbb{R})$ and $\int cf = c\int f$. It remains to show that $f+g \in \mathcal{R}(\mathbb{R})$ and $\int(f+g) = \int f + \int g$ under the assumption of $f, g \in \mathcal{R}(\mathbb{R})$. In this case, we have, by Lemma 2.2.2, that

$$\overline{\int}(f+g) \leq \overline{\int} f + \overline{\int} g = \underline{\int} f + \underline{\int} g \leq \underline{\int}(f+g), \tag{2.3.25}$$

forcing all inequalities in (2.3.25) to be equalities. Hence $f+g \in \mathcal{R}(\mathbb{R})$ and $\int(f+g) = \int f + \int g$ as desired.

(2) follows from Proposition 2.3.8, (4) is clear from the definition, and (5) follows from Lemma 2.3.7. $\square$

We are now well prepared to develop the theory of successive Riemann integrals. To start, we fix a dimension $n \in \mathbb{N}^+$ and a permutation $\sigma \in S_n$, which keeps track of the order with respect to which we perform the successive one-dimensional Riemann integrals. As in § 2.2, we give the following definition.

Notation 2.3.10. — *To simplify notation, for each $i \in \{1, \ldots, n\}$, we use the symbol $\overline{\mathcal{I}}_i$ (resp. $\underline{\mathcal{I}}_i$) to denote the upper (resp. lower) Riemann integral with respect to the i-th variable. For example, if $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^3 \to \mathbb{R}$ are bounded compactly supported functions, then*

$$(\overline{\mathcal{I}}_2 f)(x_1) = \overline{\int} f(x_1, x_2)\mathrm{d}x_2$$

is a bounded compactly supported function from $\mathbb{R} \to \mathbb{R}$, and

$$(\underline{\mathcal{I}}_2 g)(x_1, x_3) = \underline{\int} g(x_1, x_2, x_3)\mathrm{d}x_2 \tag{2.3.26}$$

is a bounded compactly supported function from $\mathbb{R}^2 \to \mathbb{R}$ (note how we label its variables as (x_1, x_3) instead of (x_1, x_2), to indicate that the second variable x_2 has been integrated out).

Definition 2.3.11. — Let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded compactly supported function. The **lower** (resp. **upper**) **successive Riemann integral** of f **with respect to the order** σ is the number $\underline{\mathcal{I}}_\sigma(f)$ (resp. $\overline{\mathcal{I}}_\sigma(f)$) defined respectively by

$$\underline{\mathcal{I}}_\sigma(f) := \underline{\mathcal{I}}_{\sigma(1)}\Big(\underline{\mathcal{I}}_{\sigma(2)}\big(\cdots(\underline{\mathcal{I}}_{\sigma(n)} f)\cdots\big)\Big) = (\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n)})(f), \tag{2.3.27}$$

and

$$\overline{\mathcal{I}}_\sigma(f) := \overline{\mathcal{I}}_{\sigma(1)}\Big(\overline{\mathcal{I}}_{\sigma(2)}\big(\cdots(\overline{\mathcal{I}}_{\sigma(n)} f)\cdots\big)\Big) = (\overline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \overline{\mathcal{I}}_{\sigma(n)})(f). \tag{2.3.28}$$

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is **successively Riemann integrable with respect to the order** σ, if f is the upper successive integral $\overline{\mathcal{I}}_\sigma(f)$ coincide with the lower one $\underline{\mathcal{I}}_\sigma(f)$, and we denote this common value by $\mathcal{I}_\sigma(f)$, and call it the **successive Riemann integral** of f **with respect to the order** σ. We denote by $\mathcal{RS}_\sigma(\mathbb{R}^n)$ the space of all successively Riemann integrable functions with respect to the order σ.

Proposition 2.3.12. — *The following holds.*

(1) *(Linearity) The space $\mathcal{RS}_\sigma(\mathbb{R}^n)$ is a linear subspace of the space of all functions from $\mathbb{R}^n$ to $\mathbb{R}$, and the successive integration map $\mathcal{I}_\sigma : \mathcal{RS}_\sigma(\mathbb{R}^n) \to \mathbb{R}$ is linear.*
(2) *(Positivity) For each $f \in \mathcal{RS}_\sigma(\mathbb{R}^n)$ with $f \geq 0$, i.e. $f(x) \geq 0$ for all $x \in \mathbb{R}^n$, we have $\mathcal{I}_\sigma(f) \geq 0$.*
(3) *(Translation invariance) For any $f : \mathbb{R}^n \to \mathbb{R}$ that is bounded and compactly supported and any $a \in \mathbb{R}^n$, we have*

$$\underline{\mathcal{I}}_\sigma(\tau_a f) = \underline{\mathcal{I}}_\sigma(f) \tag{2.3.29}$$

and

$$\overline{\mathcal{I}}_\sigma(\tau_a f) = \overline{\mathcal{I}}_\sigma(f). \tag{2.3.30}$$

In particular, For each $f \in \mathcal{RS}_\sigma(\mathbb{R}^n)$ and each $a \in \mathbb{R}^n$, we have

$$\mathcal{I}_\sigma(\tau_a f) = \mathcal{I}_\sigma(f). \tag{2.3.31}$$

Proof. — (1). It is clear that $\mathcal{RS}_\sigma(\mathbb{R}^n)$ is stable under scalar multiplications, and scalar multiplications commute with $\mathcal{I}_\sigma$. It remains to establish the additivity. For $f, g \in \mathcal{RS}_\sigma(\mathbb{R}^n)$, a repeated use of Lemma 2.2.2 shows that

$$\overline{\mathcal{I}}_\sigma(f+g) \leq \overline{\mathcal{I}}_\sigma(f) + \overline{\mathcal{I}}_\sigma(g) = \underline{\mathcal{I}}_\sigma(f) + \underline{\mathcal{I}}_\sigma(g) \leq \underline{\mathcal{I}}_\sigma(f+g), \tag{2.3.32}$$

which forces all inequalities in (2.3.32) to be equalities, and we indeed have $f+g \in \mathcal{RS}_\sigma(\mathbb{R}^n)$ and $\mathcal{I}_\sigma(f+g) = \mathcal{I}_\sigma(f) + \mathcal{I}_\sigma(g)$.

(2) is clear from definition.

(3). We first treat the lower successive Riemann integral case.

The main difficulty here is about notation rather than mathematical complication. We thus write out carefully what we mean by the successive Riemann integral of f with respect to σ and the upper/lower variants. By definition and Lemma 2.3.7, we have

$$\begin{aligned}\underline{\mathcal{I}}_\sigma(\tau_a f) &= \underline{\int} \cdots \left(\underline{\int} f(x_1 + a_1, \ldots, x_{\sigma(n)} + a_{\sigma(n)}, \ldots, x_n + a_n) \mathrm{d}x_{\sigma(n)} \right) \cdots \mathrm{d}x_{\sigma(1)} \\ &= \underline{\int} \cdots \left(\underline{\int} f(x_1 + a_1, \ldots, x_{\sigma(n)}, \ldots, x_n + a_n) \mathrm{d}x_{\sigma(n)} \right) \cdots \mathrm{d}x_{\sigma(1)}.\end{aligned} \tag{2.3.33}$$

That is to say, by taking first the lower Riemann integral with respect to the variable $x_{\sigma(n)}$, we may replace $x_{\sigma_n} + a_{\sigma(n)}$ with $x_{\sigma(n)}$ in the first line of (2.3.33) and thus eliminate the effect of the translation on the $\sigma(n)$-th variable. We now perform an induction on n, i.e. the number of variables. The case $n = 1$ is settled by Lemma 2.3.7 already. We make the induction hypothesis, we assume that the corresponding result holds for bounded compactly supported functions of $n-1$ variables, and it suffices to establish (2.3.29) under the induction hypothesis.

We set $x_{\widehat{\sigma(n)}}$ to be the element $(x_1, \ldots, \widehat{x_{\sigma(n)}}, \ldots, x_n) \in \mathbb{R}^{n-1}$, where the hat notation $\widehat{x_{\sigma(n)}}$ means the number $x_{\sigma(n)}$ is removed from $(x_1, \ldots, x_n)$. The notation $a_{\widehat{\sigma(n)}} \in \mathbb{R}^{n-1}$ has a similar meaning. We set

$$g\left(x_{\widehat{\sigma(n)}}\right) = \underline{\int} f(x_1, \ldots, x_n) \mathrm{d}x_{\sigma(n)}, \qquad x_{\widehat{\sigma(n)}} \in \mathbb{R}^{n-1}. \tag{2.3.34}$$

Then $g : \mathbb{R}^{n-1} \to \mathbb{R}$ is clearly bounded and compactly supported because f is so. By definition, we have

$$\underline{\mathcal{I}}_\sigma(f) = \big(\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n-1)}\big)(g). \tag{2.3.35}$$

By the induction hypothesis, we have

$$\big(\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n-1)}\big)(g) = \big(\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n-1)}\big)(\tau_{a_{\widehat{\sigma(n)}}} g). \tag{2.3.36}$$

However, by definition, we have

$$\begin{aligned}&\big(\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n-1)}\big)(\tau_{a_{\widehat{\sigma(n)}}} g) \\ &= \underline{\int} \cdots \left(\underline{\int} f(x_1 + a_1, \ldots, x_{\sigma(n)}, \ldots, x_n + a_n) \mathrm{d}x_{\sigma(n)} \right) \cdots \mathrm{d}x_{\sigma(1)}.\end{aligned} \tag{2.3.37}$$

From the above, we obtain

$$\begin{aligned}\underline{\mathcal{I}}_\sigma(\tau_a f) &= \underline{\int} \cdots \left(\underline{\int} f(x_1 + a_1, \ldots, x_{\sigma(n)}, \ldots, x_n + a_n) \mathrm{d}x_{\sigma(n)} \right) \cdots \mathrm{d}x_{\sigma(1)} && \text{(by (2.3.33))} \\ &= \big(\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n-1)}\big)(\tau_{a_{\widehat{\sigma(n)}}} g) && \text{(by (2.3.37))} \\ &= \big(\underline{\mathcal{I}}_{\sigma(1)} \circ \cdots \circ \underline{\mathcal{I}}_{\sigma(n-1)}\big)(g) && \text{(by (2.3.36))} \\ &= \underline{\mathcal{I}}_\sigma(f) && \text{(by definition of } \underline{\mathcal{I}}_\sigma(f)\text{)}.\end{aligned}$$

This proves (2.3.29).

Now (2.3.30) follows from (2.3.29) by noting that

$$\overline{\mathcal{I}}_\sigma(\tau_a f) = -\underline{\mathcal{I}}_\sigma(\tau_a(-f)) = -\underline{\mathcal{I}}_\sigma(-f) = \overline{\mathcal{I}}_\sigma(f). \tag{2.3.38}$$

□

To construct a theory of volumes via Proposition 2.3.12, we need the following preparation on characteristic functions. Recall that we use χ_A to denote the characteristic function of the set A.

Proposition 2.3.13. — *Let X be a set. Consider all characteristic functions of subsets of X as maps from X to $\mathbb{R}$ (in fact, one can replace $\mathbb{R}$ with any non-zero ring). The following hold.*

(1) *If $A \subseteq X$, then $A = \chi_A^{-1}(1)$ and*

$$\chi_{A^c} = \chi_{X \setminus A} = 1 - \chi_A. \tag{2.3.39}$$

(2) *If $A, B \subseteq X$, then*

$$A \cap B = \emptyset \iff \chi_{A \cup B} = \chi_A + \chi_B. \tag{2.3.40}$$

(3) *If $A, B \subseteq X$, then*

$$\chi_{AB} = \chi_A \chi_B \tag{2.3.41}$$

$$\chi_{A \setminus B} = \chi_A(1 - \chi_B) = \chi_A - \chi_A \chi_B \tag{2.3.42}$$

and

$$\chi_{A \cup B} = \chi_A + \chi_{B \setminus A} = \chi_A + \chi_B - \chi_A \chi_B. \tag{2.3.43}$$

(4) *If in addition, X is a vector space, and $A \subseteq X$, then*

$$\chi_{A+x} = \tau_x \chi_A. \tag{2.3.44}$$

(5) *If in addition, X is a topological space, then*

$$\operatorname{supp} \chi_A = \overline{A}. \tag{2.3.45}$$

Proof. — All of these are just routine check which we leave to the reader. □

Definition 2.3.14. — Fix an $n \in \mathbb{N}^+$ and a $\sigma \in S_n$. A subset $A \subseteq X$ is called **successively contented with respect to** (the order) σ, if $\chi_A \in \mathcal{RS}_\sigma(\mathbb{R}^n)$. In this case, the number $\mathcal{I}_\sigma(\chi_A)$, which is denoted by $\nu_\sigma(A)$, is called the **successive content**, or **successive volume**, of A, with respect to σ. The collection of successively contented sets is denoted by $\mathcal{O}_\sigma(\mathbb{R}^n)$ and $\nu_\sigma : \mathcal{O}_\sigma(\mathbb{R}^n) \to \mathbb{R}$ denotes the obvious map $A \mapsto \sigma_\sigma(A)$.

Corollary 2.3.15. — *Fix an $n \in \mathbb{N}^+$ and a $\sigma \in S_n$. Let $\mathcal{O}(\mathbb{R}^n) \subseteq \mathcal{O}_\sigma(\mathbb{R}^n)$ be a sub-collection of successively contented sets with respect to σ. If $\mathcal{O}(\mathbb{R}^n)$ satisfies the axioms (O1), (O2) and (O3), then the pair $\big(\mathcal{O}(\mathbb{R}^n), \nu_\sigma \restriction_{\mathcal{O}(\mathbb{R}^n)}\big)$ is a theory of volumes.*

Proof. — Let $\nu := \nu_\sigma \restriction_{\mathcal{O}(\mathbb{R}^n)}$. It suffices to check that $\big(\mathcal{O}(\mathbb{R}^n), \nu\big)$ satisfies the axioms (ν1), (ν2) and (ν3). The positivity of successive integrals (Proposition 2.3.12 (2)) implies that $\nu(A) = \mathcal{I}_\sigma(\chi_A) \geq 0$ for any $A \in \mathcal{O}(\mathbb{R}^n) \subseteq \mathcal{O}_\sigma(\mathbb{R}^n)$, so ($\nu$1) is verified. If $A, B \in \mathcal{O}(\mathbb{R}^n)$ with $A \cap B = \emptyset$, then $\chi_{A \cup B} = \chi_A + \chi_B$ by Proposition 2.3.13 (3), hence by Proposition 2.3.12 (1), we have $\chi_{A \cup B} \in \mathcal{RS}_\sigma(\mathbb{R}^n)$ with

$$\nu(A \cup B) = \mathcal{I}_\sigma(\chi_{A \cup B}) = \mathcal{I}_\sigma(\chi_A + \chi_B) = \mathcal{I}_\sigma(\chi_A) + \mathcal{I}_\sigma(\chi_B) = \nu(A) + \nu(B).$$

This establishes (ν2). For (ν3), it suffices to apply Proposition 2.3.12 (4) and Proposition 2.3.12 (3). □

Remark 2.3.16. — There is an obvious example of Corollary 2.3.15: just take taking $\mathcal{O}(\mathbb{R}^n)$ to be $\mathcal{P}(\mathbb{R}^n)$. However, the real interest of this corollary lies in studying "larger theory" of volumes. This leads to the natural question of whether $\mathcal{O}_\sigma(\mathbb{R}^n)$ itself satisfies the axioms (O1), (O2) and (O3). It is easy to see from Proposition 2.3.12 and Proposition 2.3.13 that the axioms (O2) and (O3) are indeed satisfied by taking $\mathcal{O}(\mathbb{R}^n) := \mathcal{O}_\sigma(\mathbb{R}^n)$, and that (O1) is satisfied in this case if and only if $\mathcal{O}_\sigma(\mathbb{R}^n)$ is stable under taking finite intersections. As of this writing, the author of these notes does not know whether the latter always holds or not. More generally, the author does not know whether $\mathcal{RS}_\sigma(\mathbb{R}^n)$ is stable under taking pointwise multiplication, unless $n = 1$, in which case we do have an affirmative answer (Proposition 2.3.8).

Remark 2.3.16 naturally leads us to consider the class of bounded, compactly supported real-valued functions on $\mathbb{R}^n$ that are successively Riemann integrable with respect to any $\sigma \in S_n$ *and* that the resulting integrals share a common value that is independent of the choice of σ.

Definition 2.3.17. — A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **successively Riemann integrable**, if $f \in \cap_{\sigma\in S_n}\mathcal{SR}_\sigma(\mathbb{R}^n)$ and $\mathcal{I}_\sigma(f)$ is independent of $\sigma \in S_n$. In this case, we write $\mathcal{I}(f)$, or $\int f$, or $\int_{\mathbb{R}^n} f(t)\mathrm{d}t$, for the common value of $\mathcal{I}_\sigma(f)$, $\sigma \in S_n$, and call it the **successive Riemann integral** of f. We use $\mathcal{SR}_s(\mathbb{R}^n)$ to denote the space of all successively Riemann integrable real-valued functions on $\mathbb{R}^n$, and we define $\mathcal{O}_s(\mathbb{R}^n)$ by

$$\mathcal{O}_s(\mathbb{R}^n) := \{A \subseteq \mathbb{R}^n \mid \chi_A \in \mathcal{RS}_s(\mathbb{R}^n)\}. \tag{2.3.46}$$

Elements of $\mathcal{O}_s(\mathbb{R}^n)$ will be called **successively contented sets**. We also define

$$\begin{aligned} \nu_s : \mathcal{O}_s(\mathbb{R}^n) &\to \mathbb{R} \\ A &\mapsto \mathcal{I}(\chi_A), \end{aligned} \tag{2.3.47}$$

and call $\nu_s(A)$ the **successive content**, or **successive volume**, of A for any $A \in \mathcal{O}_s(\mathbb{R}^n)$.

Remark 2.3.18. — We have a terminology for the property that the number $\mathcal{I}_\sigma(f)$ is independent of the order $\sigma \in S_n$ with respect to which we take successive integrals: this is called the Fubini-type property after the Italian mathematician Guido Fubini. We will encounter an important theorem in our integration theory that bares his name, which states that certain functions, called successively contented Riemann integrable functions, enjoy this Fubini-type property.

Remark 2.3.19. — Based on the materials in this section, one can already develop most, if not all, of the Riemann integration theory in $\mathbb{R}^n$, using only successive (upper/lower) Riemann integrals from the familiar one dimensional case. The only missing key property that we haven't established is the change of variable formula. But the latter can also be established without much pain if we are willing to do some work related to the inverse mapping theorem. This is the approach adopted e.g. in Walter Rudin's famous book *Principles of Mathematical Analysis* (see also one of the optional problem sets for a slightly more general case than considered by Rudin). One principle reason that we do not pursue this tempting idea further is that we wish to speak of more general symmetries than the translation-invariant ones for a "good" theory of volumes, another reason is that we wish to describe Banach space valued integrals, in which case the approach in this section no longer works, since there is no corresponding notion of upper/lower Riemann integrals because of the lack of a nice order structure on a general Banach space. This being said, let us mention that from the perspective of developing a good theory of volumes that can measure larger classes of sets than $\mathcal{P}(\mathbb{R}^n)$, the process of taking successive integrals almost succeeds, but still not quite yet (see Remark 2.3.16). This is a third good reason why we will introduce the theory of Jordan sets and contents, which do provide such a good theory of volumes.

2.4. Jordan Sets and Contents

Fix an $n \in \mathbb{N}^+$ through this section. Take any $c > 0$ and consider our unique theory of volumes $\big(\mathcal{P}(\mathbb{R}^n), \nu_c\big)$ on the collection $\mathcal{P}(\mathbb{R}^n)$ of paved sets with $\nu_c(\square_{\mathbf{0}}^{\mathbf{1}}) = c$. When $c = 1$, we simply write ν_c as ν, so $\nu_c = c\nu$ by uniqueness of ν_c. Since the very beginning of calculus, approximation is the principle idea, which naturally leads us to formulate the following definition of subsets of $\mathbb{R}^n$ that can be approximate by paved sets from the outside and inside to an arbitrary precision.

Definition 2.4.1. — A subset $A \subseteq \mathbb{R}^n$ is called a **Jordan set**, if for each $\varepsilon > 0$, there exists $P, Q \in \mathcal{P}(\mathbb{R}^n)$, such that $P \subseteq A \subseteq Q$, and

$$\nu(Q \setminus P) = \nu(Q) - \nu(P) < \varepsilon. \tag{2.4.1}$$

We use $\mathcal{J}(\mathbb{R}^n)$ to denote[(2)] the collection of all Jordan sets in $\mathbb{R}^n$.

Lemma 2.4.2. — *If $A \in \mathcal{J}(\mathbb{R}^n)$, then there exists a unique number $r \in \mathbb{R}$ such that for any $P \in \mathcal{P}(\mathbb{R}^n)$, we have $A \subseteq P$ implies that $r \le \nu(P)$ and $P \subseteq A$ implies $\nu(P) \le r$. Moreover,*

$$r = \sup\{\nu(P) \mid P \in \mathcal{P}(\mathbb{R}^n) \ \text{ and } \ P \subseteq A\} = \inf\{\nu(Q) \mid Q \in \mathcal{P}(\mathbb{R}^n) \ \text{ and } \ A \subseteq Q\}. \tag{2.4.2}$$

---

(2) We warn the readers that some authors, e.g. in Dustermaat & Kolk, *Multidimensional Real Analysis, volume 2*, use the same symbol $\mathcal{J}(\mathbb{R}^n)$ to denote the collection of *compact* Jordan sets.

Proof. — By definition of the supremum, $P \in \mathcal{P}(\mathbb{R}^n)$ and $P \subseteq A$ implies $\nu(P) \leq r$ if and only if

$$r \geq \sup\{\nu(P) \mid P \in \mathcal{P}(\mathbb{R}^n) \text{ and } P \subseteq A\}. \tag{2.4.3}$$

Similarly, by definition of the infimum, $Q \in \mathcal{P}(\mathbb{R}^n)$ and $A \subseteq Q$ implies $r \leq \nu(Q)$ if and only if

$$r \leq \inf\{\nu(Q) \mid Q \in \mathcal{P}(\mathbb{R}^n) \text{ and } A \subseteq Q\}. \tag{2.4.4}$$

By definition 2.4.1, we have

$$\sup\{\nu(P) \mid P \in \mathcal{P}(\mathbb{R}^n) \text{ and } P \subseteq A\} = \inf\{\nu(Q) \mid Q \in \mathcal{P}(\mathbb{R}^n) \text{ and } A \subseteq Q\}, \tag{2.4.5}$$

which is easy to see is the unique desired value of r. □

Clearly, if A itself is a paved set, then the unique number r in Lemma 2.4.2 is exactly $\nu(A)$. This means that the notation introduced in Definition 2.4.3 will cause no ambiguity.

Definition 2.4.3. — For any Jordan set $A \in \mathcal{J}(\mathbb{R}^n)$, the unique number $r \in \mathbb{R}$ in Lemma 2.4.2 is called the **Jordan content**, or the **volume** of A, and is denoted by $\nu(A)$, i.e.

$$\nu(A) = \sup\{\nu(P) \mid P \in \mathcal{P}(\mathbb{R}^n) \text{ and } P \subseteq A\} = \inf\{\nu(Q) \mid Q \in \mathcal{P}(\mathbb{R}^n) \text{ and } A \subseteq Q\}. \tag{2.4.6}$$

Notation 2.4.4. — *From now on, unless stated otherwise, we will no longer use ν to denote the volume function of an arbitrary theory of volumes, but reserve it for the volumes of Jordan sets.*

Example 2.4.5. — (Exercise) Take any box $\square_a^b$ in $\mathbb{R}^n$, the box itself, the interior of this box and the closure of this box are all Jordan sets. More generally, for $P \in \mathcal{P}(\mathbb{R}^n)$, we have $P, \mathrm{Int}(P), \overline{P} \in \mathcal{J}(\mathbb{R}^n)$.

More generally, we may also relate Jordan set to successively contented sets as follows.

Proposition 2.4.6. — *For any $A \in \mathcal{J}(\mathbb{R}^n)$, we have $A \in \mathcal{RS}_s(\mathbb{R}^n)$ and $\nu(A) = \nu_s(A)$.*

Proof. — We need to show that for any $\sigma \in S^n$, we have $\chi_A \in \mathcal{RS}_\sigma(\mathbb{R}^n)$, and $\mathcal{I}_\sigma(\chi_A) = \nu(A)$. Indeed, for any $\varepsilon > 0$, choose $P, Q \in \mathcal{P}(\mathbb{R}^n)$ so that $P \subseteq A \subseteq Q$, then $\chi_P \leq \chi_A \leq \chi_Q$. It follows that

$$\nu(P) = \mathcal{I}_\sigma(\chi_P) \leq \underline{\mathcal{I}}_\sigma(\chi_A) \leq \overline{\mathcal{I}}_\sigma(\chi_A) \leq \overline{\mathcal{I}}_\sigma(\chi_Q) = \nu(Q). \tag{2.4.7}$$

Now since A is a Jordan set, (2.4.7) implies that

$$\begin{aligned}\nu(A) &= \sup\{\nu(P) \mid P \subseteq A \text{ and } P \in \mathcal{P}(\mathbb{R}^n)\}\\ &\leq \underline{\mathcal{I}}_\sigma(\chi_A) \leq \overline{\mathcal{I}}_\sigma(\chi_A)\\ &\leq \inf\{\nu(Q) \mid Q \supseteq A \text{ and } Q \in \mathcal{P}(\mathbb{R}^n)\} = \nu(A).\end{aligned} \tag{2.4.8}$$

This forces $\chi_A \in \mathcal{RS}_\sigma(\mathbb{R}^n)$ with $\mathcal{I}_\sigma(\chi_A) = \nu(A)$, as desired. □

Merely from the definition, we can establish via some ad-hoc method that many of the usual geometric object in $\mathbb{R}^n$, such as the n-dimensional balls, ellipsoids, n-simplicies(3), etc., are all Jordan sets. For theoretical efficiency, we do not pursue this at this stage, but instead turn to a very useful characterization of Jordan sets, which will give a more systematic treatment of these examples. For better understanding such a characterization, we introduce some preparation.

Lemma 2.4.7. — *Let $\square_a^b$ be a box in $\mathbb{R}^n$ (recall Definition 2.1.9). Then for any $\varepsilon > 0$, there exists a box $\square_{p_1}^{p_2}$ and a box $B_2 = \square_{q_1}^{q_2}$, such that*

(1) *$p_1, p_2, q_1, q_2 \in \mathbb{Q}^n$;*
(2) *$\square_{p_1}^{p_2} \subseteq \mathrm{Int}\left(\square_a^b\right) \subseteq \overline{\square_a^b} \subseteq \square_{q_1}^{q_2}$;*
(3) *we have*

$$\nu\left(\square_a^b \setminus \square_{p_1}^{p_2}\right) + \nu\left(\square_{q_1}^{q_2} \setminus \square_a^b\right) = \nu\left(\square_{q_1}^{q_2} \setminus \square_{p_1}^{p_2}\right) \leq \varepsilon.$$

Proof. — The lemma is trivial if $\square_a^b = \emptyset$, so we treat the non-empty case in the following, which means $a < b$ (see Notation 2.1.1). Let $a = (a_i)$ and $b = (b_i)$, $i = 1, \ldots, n$. For each i, since $a_i < b_i$, we may choose $p_{1i}, p_{2i}, q_{1i}, q_{2i} \in \mathbb{Q}$, such that

$$\forall 1 \leq i \leq n, \qquad q_{1i} < a_i < p_{1i} < p_{2i} < b_i < q_{2i}. \tag{2.4.9}$$

(3) The word "Simplicies" is the plural form of the word "simplex".

Set $p_1 = (p_{1i})$, $p_2 = (p_{2i})$, $q_1 = (q_{1i})$, $q_2 = (q_{2i})$, which immediately guarantees (1) and it follows from (2.4.9) that we have (2). Finally by choosing p_{1i}, q_{1i} close to a_i and p_{2i}, q_{2i} close to b_i for each i, we can make (2.4.9) hold and the lemma is proved. ☐

Definition 2.4.8. — For $N \in \mathbb{N}^+$ and $k_1, \ldots, k_n \in \mathbb{Z}$, we set

$$\Box(N; k_1, \ldots, k_n) := \Box^{((k_1+1)/N, \ldots, (k_n+1)/N)}_{(k_1/N, \ldots, k_n/N)}. \tag{2.4.10}$$

We call boxes of the form (2.4.10) a **grid box of level** N. A paved set is called a **grid paved set of level** N if it is a finite union of grid boxes of level N. More generally, boxes of the form $\Box^b_a$ with $a, b \in \mathbb{Q}^n$ are called **rational**, and paved sets in $\mathbb{R}^n$ are called **rational** if it is a finite union of rational boxes.

Corollary 2.4.9. — *For any $P \in \mathcal{P}(\mathbb{R}^n)$ and any $\varepsilon > 0$, there are* rational $P_1, P_2 \in \mathcal{P}(\mathbb{R}^n)$, *such that*

$$P_1 \subseteq \overline{P_1} \subseteq \mathrm{Int}(P) \subseteq P \subseteq \overline{P} \subseteq \mathrm{Int}(P_2) \subseteq P_2. \tag{2.4.11}$$

and

$$\nu(P_2 \setminus P_1) < \varepsilon. \tag{2.4.12}$$

Proof. — This follows from the definition and Lemma 2.4.7. We leave the details as an exercise. ☐

Lemma 2.4.10. — *The following hold.*

(1) *For every $N \in \mathbb{N}^+$, we have the decomposition*

$$\mathbb{R}^n = \bigsqcup_{(k_1, \ldots, k_n) \in \mathbb{Z}^n} \Box(N; k_1, \ldots, k_n) \tag{2.4.13}$$

of the whole space $\mathbb{R}^n$ into the disjoint union of all grid boxes of level N.

(2) *Each rational box is a finite disjoint union of grid boxes of level N for some $N \in \mathbb{N}^+$.*

(3) *If $N_1, N_2 \in \mathbb{N}$ and $N_1 \mid N_2$ (recall from elementary number theory course that this means N_1 divides N_2), then each grid box of level N_1 is the disjoint union of exactly $(N_2/N_1)^n$ grid boxes of level N_2.*

(4) *For finitely many rational $P_1, \ldots, P_m \in \mathcal{P}(\mathbb{R}^n)$, there exists an $N \in \mathbb{N}^+$ such that every P_i, $1 \le i \le m$ is a grid paved set of level N.*

Proof. — (1). For any real number $r \in \mathbb{R}$, there exists a unique $k \in \mathbb{Z}$, such that $k/N \le r < (k+1)/N$, namely $k = [Nr]$ ($[\cdot]$ is the Gauss floor function). It follows from this that for any $x = (x_i) \in \mathbb{R}$, the point $([Nx_1], \ldots, [Nx_n]) \in \mathbb{Z}^n$ is the unique point in $(k_1, \ldots, k_n)$ in $\mathbb{Z}^n$ such that $x \in \Box(N; k_1, \ldots, k_n)$, which yields the decomposition (2.4.13).

(2). For each rational box $\Box^b_a$ with $a = (a_i)$ and $b = (b_i)$, we know $a_i, b_i \in \mathbb{Q}$ for all $i = 1, \ldots, n$ by definition, hence we can write $a_i = r_i/s_i$ and $b_i = u_i/v_i$ with $r_i, u_i \in \mathbb{Z}$ and $s_i, v_i \in \mathbb{N}^+$ for each i. Taking $N = (\prod_{i=1}^n (s_i v_i)) \in \mathbb{N}^+$, it is easy to see that each grid box of level N either is contained in $\Box^b_a$ or is disjoint from $\Box^b_a$. Hence $\Box^b_a$ is a disjoint union of grid boxes of level N. Since each grid box of level N has the same positive volume (which is $1/N^n$) and $\Box^b_a$ has finite volumes, it follows that $\Box^b_a$ is a finite disjoint union of grid boxes of level N.

(3). Since $N_2/N_1 \in \mathbb{N}$, we see that for any grid box B_1 (resp. B_2) of level N_1 (resp. N_2), we have either $B_1 \cap B_2 = \emptyset$ or $B_2 \subseteq B_1$. It follows from the decomposition in (1) that each grid box B of level N_1 is a disjoint union of grid boxes of level N_2. The volume of each grid box of level N_2 is N_2^{-n}, while the volume of B is N_1^{-n}. Hence the number of grid boxes of level N_2 in the aforementioned decomposition of N_1 is exactly $N_1^{-n}/N_2^{-n} = (N_2/N_1)^n$.

(4). We know from the decomposition (2.4.13) that it suffices to show that for a suitable choice of $N \in \mathbb{N}^+$, each P_i, $1 \le i \le m$ is a finite union (necessarily disjoint) of grid boxes of level N. By definition of rational paved sets, there are decompositions

$$\forall i = 1, \ldots, m,\ P_i = \bigcup_{j=1}^{m(i)} B_{ij}, \tag{2.4.14}$$

where each $m(i) \in \mathbb{N}$ and each B_{ij} is a rational box. By (2), there exists an $N_{ij} \in \mathbb{N}^+$ for each (i,j) such that B_{ij} is a finite disjoint union of grid boxes of level N_{ij}. Let $N \in \mathbb{N}^+$ be the product of all these N_{ij}'s, then (4) follows (2.4.14) and (3). $\square$

Definition 2.4.11. — A bounded set $A \subseteq \mathbb{R}^n$ is said to be **Jordan negligible**, or **of Jordan content** zero, if for each $\varepsilon > 0$, there exists a $P \subseteq \mathcal{P}(R^n)$, such that $\nu(P) < \varepsilon$ and $A \subseteq P$.

Remark 2.4.12. — (Exercise) All Jordan negligible sets are Jordan sets with Jordan content 0. Finite union of Jordan negligible sets remains Jordan negligible and so is any subset of a Jordan negligible set.

We may now give our desired characterization of Jordan sets.

Proposition 2.4.13. — *Let $A \subseteq \mathbb{R}^n$ be a bounded set. The following are equivalent.*

(1) *The set A is a Jordan set.*
(2) *For each $\varepsilon > 0$, there exists* rational *paved sets $P, Q \in \mathcal{P}(\mathbb{R}^n)$, such that $P \subseteq \operatorname{Int}(A) \subseteq \overline{A} \subseteq Q$ and $\nu(Q \setminus P) < \varepsilon$.*
(3) *The boundary ∂A is of Jordan content zero.*

Proof. — We give a brief sketch and leave the details as an exercise. (1) implies (2). This follows from Corollary 2.4.9.

(2) implies (3). This follows directly form the definition of Jordan negligible sets.

(3) implies (1). Take any $\varepsilon > 0$. By definition of Jordan negligible sets, there exists $P_0 \in \mathcal{P}(\mathbb{R}^n)$ with $\nu(P_0) < \varepsilon$ and $\partial A \subseteq P_0$. By Corollary 2.4.9, we may assume P_0 is rational, hence is a grid paved set of level N for some $N \in \mathbb{N}^+$ by Lemma 2.4.10. Set

$$P := \bigcup_{\substack{(k_1,\ldots,k_n)\in\mathbb{Z}^n \\ \square(N;k_1,\ldots,k_n)\subseteq \operatorname{Int}(A)}} \square(N; k_1, \ldots, k_n). \tag{2.4.15}$$

Then the boundedness of A implies (2.4.15) is a finite union. On the other hand, for any grid box B of level N with $B \cap \overline{A} \neq \emptyset$, either $B \subseteq \operatorname{Int}(A)$, in which case $B \subseteq P$, or $B \cap \partial A \neq \emptyset$, in which case $B \subseteq P_0$. Set $Q := P \cup P_0$. From our construction, we see that $P \subseteq \operatorname{Int}(A) \subseteq \overline{A} \subseteq Q$, hence

$$\partial A = \overline{A} \setminus \operatorname{Int}(A) \subseteq Q \setminus P \subseteq P_0 \tag{2.4.16}$$

and $\nu(P_0) < \varepsilon$ implies that $\nu(Q \setminus P) < \varepsilon$ so $A \in \mathcal{J}(\mathbb{R}^n)$ as desired. $\square$

As a corollary, we may finally construct a theory of volumes "larger" that encompasses more than just paved sets.

Corollary 2.4.14. — *The pair $\big(\mathcal{J}(\mathbb{R}^n), \nu\big)$ is a theory of volumes. Moreover, if $\big(\mathcal{O}(\mathbb{R}^n), \mu\big)$ is a theory of volumes such that $\mathcal{O}(\mathbb{R}^n) \supseteq \mathcal{J}(\mathbb{R}^n)$, then there exists a unique constant $c \geq 0$, such that $\mu \restriction_{\mathcal{J}(\mathbb{R}^n)} = c\nu$.*

Proof. — We first check that $\big(\mathcal{J}(\mathbb{R}^n), \nu\big)$ is indeed a theory of volumes. It is clear that $\mathcal{J}(\mathbb{R}^n)$ is stable under all translations and $\square_{\mathbf{0}}^{\mathbf{1}} \in \mathcal{J}(\mathbb{R}^n)$. If $A, B \in \mathcal{J}(\mathbb{R}^n)$, then

$$\partial(A \cup B) = \overline{A \cup B} \setminus \operatorname{Int}(A \cup B) \subseteq (\overline{A} \cup \overline{B}) \setminus \big(\operatorname{Int}(A) \cup \operatorname{Int}(B)\big) \subseteq \partial A \cup \partial B. \tag{2.4.17}$$

It follows from Remark 2.4.12 and Proposition 2.4.13 that $\partial(A \cup B)$ is Jordan negligible. Similarly, we have

$$\partial(A \cap B) = \partial(A^c \cup B^c)^c \subseteq \partial A^c \cup \partial B^c = \partial A \cup \partial B, \tag{2.4.18}$$

and

$$\partial(A \cap B) = \partial(A \cap B^c) \subseteq \partial A \cup \partial B^c = \partial A \cup \partial B. \tag{2.4.19}$$

Hence $A \cap B, A \setminus B \in \mathcal{J}(\mathbb{R}^n)$. Now by Corollary 2.3.15 and Proposition 2.4.6, we see that $\big(\mathcal{J}(\mathbb{R}^n), \nu\big)$ is a theory of volumes.

If $\mu(\square_{\mathbf{0}}^{\mathbf{1}}) = 0$, then the theory $\big(\mathcal{O}(\mathbb{R}^n), \mu\big)$ is trivial since any bounded set is contained in a finite disjoint union of grid boxes of level 1. In this case, it is clear that $c = 0$ is the unique constant such that $\mu \restriction_{\mathcal{J}(\mathbb{R}^n)} = c\nu$.

Now suppose $c := \mu(\square_{\mathbf{0}}^{\mathbf{1}}) > 0$. Set $\mu_0 := c^{-1}\mu$. Then $\big(\mathcal{O}(\mathbb{R}^n), \mu_0\big)$ remains a theory of volumes with $\mu_0(\square_{\mathbf{0}}^{\mathbf{1}}) = 1$, so $\mu_0 \restriction_{\mathcal{P}(\mathbb{R}^n)} = \nu \restriction_{\mathcal{P}(\mathbb{R}^n)}$. Now for each $A \subseteq \mathcal{J}(\mathbb{R}^n)$, from $P \subseteq A \subseteq Q$ with $P, Q \in \mathcal{P}(\mathbb{R}^n)$, we see that

$$\nu(P) = \mu_0(P) \leq \mu_0(A) \leq \mu_0(Q) = \nu(Q). \tag{2.4.20}$$

Since we choose $P, Q \in \mathcal{P}(\mathbb{R}^n)$ so that $\nu(Q) - \nu(P) = \nu(Q \setminus P)$ can be as small as we want, (2.4.20) forces $\mu_0(A) = \nu(A)$ and $\mu_0 \restriction_{\mathcal{J}(\mathbb{R}^n)} = \nu$, i.e. $\mu \restriction_{\mathcal{J}(\mathbb{R}^n)} = c\nu$. The proof is now complete. $\square$

Corollary 2.4.15. — *If $A \in \mathcal{J}(\mathbb{R}^n)$, then for any $B \subseteq \mathbb{R}^n$ with $\mathrm{Int}(A) \subseteq B \subseteq \overline{A}$, we have $B \in \mathcal{J}(\mathbb{R}^n)$ and $\nu(B) = \nu(A)$.*

Proof. — Since $\partial A \in \mathcal{J}(\mathbb{R}^n)$ with $\nu(\partial A) = 0$ by Remark 2.4.12, we see from Corollary 2.4.14 that $\mathrm{Int}(A) = A \setminus \partial A \in \mathcal{J}(\mathbb{R}^n)$. Note that $A \setminus \mathrm{Int}(A) \subseteq \partial A$, we have

$$\nu(\mathrm{Int}(A)) \leq \nu(A) = \nu(\mathrm{Int}(A)) + \nu(A \setminus \mathrm{Int}(A)) \leq \nu(\mathrm{Int}(A)) + \nu(\partial A) = \nu(\mathrm{Int}(A)). \tag{2.4.21}$$

Hence $\nu(\mathrm{Int}(A)) = \nu(A)$. Moreover, for any $\mathrm{Int}(A) \subseteq B \subseteq \overline{A}$, we still have $B \setminus \mathrm{Int}(A) \subseteq \overline{A} \setminus \mathrm{Int}(A) = \partial A$, hence $B \setminus \mathrm{Int}(A)$ is a Jordan negligible set. It follows from Remark 2.4.12 and Corollary 2.4.14 again that $B = \mathrm{Int}(A) \sqcup \big(B \setminus \mathrm{Int}(A)\big)$ is a Jordan set, with $\nu(B) = \nu(\mathrm{Int}(A)) = \nu(A)$. $\square$

Definition 2.4.16. — A map $f : X \to Y$ between two metric spaces is called **Lipschitz**, if there exists a constant $k \geq 0$, such that

$$\forall x_1, x_2 \in X, \qquad d_Y\big(f(x_1), f(x_2)\big) \leq k \cdot d_X(x_1, x_2). \tag{2.4.22}$$

It is clear that Lipschitz maps are automatically continuous.

Proposition 2.4.17. — *If $U \subseteq \mathbb{R}^n$ is open, and $f : U \to \mathbb{R}^n$ is a Lipschitz map, then the following hold.*

(1) *The map f is still if we replace the usual Euclidean metric on U and $\mathbb{R}^n$ by the metric induced by $\|\cdot\|_\infty$.*
(2) *For any Jordan negligible set $A \subseteq \mathbb{R}^n$ with $\overline{A} \subseteq U$, the set $f(A)$ remains Jordan negligible.*
(3) *If f is also open, then for any $A \in \mathcal{J}(\mathbb{R}^n)$ with $\overline{A} \subseteq U$, we have $f(A) \in \mathcal{J}(\mathbb{R}^n)$. This applies in particular to the case where f is a Lipschitz homeomorphism from the open set U onto an open set in $\mathbb{R}^n$.*

Proof. — (1). On the finite dimensional vector space $\mathbb{R}^n$, the norms $\|\cdot\|_\infty$ and $\|\cdot\|$ are equivalent, i.e. there exists $c, C \in]0, +\infty[$, such that $c\|\cdot\| \leq \|\cdot\|_\infty \leq C\|\cdot\|$. Note that f is Lipschitz, there is a constant $k \in \mathbb{R}$ be a constant such that

$$\forall x_1, x_2 \in U, \qquad \|f(x_1) - f(x_2)\| \leq k\|x_1 - x_2\|. \tag{2.4.23}$$

We then have

$$\forall x_1, x_2 \in U, \qquad \|f(x_1) - f(x_2)\|_\infty \leq C\|f(x_1) - f(x_2)\| \leq Ck\|x_1 - x_2\| \leq Ckc^{-1}\|x_1 - x_2\|_\infty. \tag{2.4.24}$$

Hence f is still Lipschitz if we use the metric induced by $\|\cdot\|_\infty$ on $\mathbb{R}^n$ and on U.

(2). Take any non-empty $B = \overline{\square_a^b} \subseteq U$. Then set $c := (a+b)/2$, $l = \|b - a\|_\infty/2$ and $K = Ckc^{-1}$, then $B = \overline{B_\infty(c, l)}$ and by (2.4.24), we have

$$\forall x \in B, \qquad \|f(x) - f(c)\|_\infty \leq K\|x - c\|_\infty \leq Kl, \tag{2.4.25}$$

hence $f(B) \subseteq \overline{B_\infty(f(c), Kl)}$. It follows that

$$\nu\big(f(B)\big) \leq \nu\left(\overline{B_\infty\big(f(c), Kl\big)}\right) = K^n(2l)^n = K^n\nu(B). \tag{2.4.26}$$

Since $\overline{A} \subseteq U$ and A is compact, there is a $\delta > 0$ such that $\overline{B_\infty(x, l)} \subseteq U$ whenever $x \in \overline{A}$ and $0 < l \leq \delta$. Now fix an arbitrary $\varepsilon > 0$. By Corollary 2.4.9 and Lemma 2.4.10, we may find a grid paved set P with level N for an N large enough so that $1/N < \delta$, such that $\overline{A} \subseteq \overline{P}$ and $\nu(\overline{P}) < \varepsilon$, with P being a finite disjoint union of grid boxes of level N. If B is a grid box of level N, then $\overline{B}$ is a closed ball of radius

$1/(2N)$ (so of diameter $1/N$) with respect to $\|\cdot\|_\infty$, hence if $\overline{B} \cap \overline{A} \neq \emptyset$, then choose any $b \in \overline{B} \cap \overline{A}$, we have

$$\forall x \in \overline{B}, \qquad \|x-b\|_\infty \leq \frac{1}{N} < \delta. \tag{2.4.27}$$

Hence $\overline{B} \subseteq \overline{B_\infty(b,\delta)} \subseteq U$, and by throwing out grid boxes of level N that do not intersect with $\overline{A}$, we may assume $\overline{P} \subseteq U$. Writing $P = \sqcup_{i=1}^m B_i$ as a finite union of grid boxes of level N, we then have, by (2.4.25), that

$$f(A) \subseteq f(\overline{P}) \subseteq \bigcup_{i=1}^m f(\overline{B_i}) \subseteq \bigcup_{i=1}^m C_i, \tag{2.4.28}$$

where C_i is of the form $\overline{B_\infty(c_i, K/(2N))}$ with c_i being the image of the center of B_i under f. Being the closure of boxes, all C_i are Jordan set, hence $\cup_{i=1}^m C_i \in \mathcal{J}(\mathbb{R}^n)$. Finally, by (2.4.26), we have

$$\nu\left(\bigcup_{i=1}^m C_i\right) \leq \sum_{i=1}^m \nu(C_i) \leq K^n \sum_{i=1}^m \nu(\overline{B_i}) = K^n \sum_{i=1}^m \nu(B_i) = K^n \nu(P) < K^n\varepsilon. \tag{2.4.29}$$

Since $\varepsilon > 0$ is arbitrary, this shows that $f(A)$ is indeed Jordan negligible.

(3). Since equivalent norms give the same topology on $\mathbb{R}^n$, we see that $f : U \to \mathbb{R}^n$ is open regardless which norm on $\mathbb{R}^n$ we choose. On the other hand, $\overline{A}$ being compact, we see that $f(\overline{A})$ is compact, hence closed in $\mathbb{R}^n$. Thus $f(\overline{A}) \supseteq \overline{f(A)}$. Since f is continuous (all Lipschitz maps are), we have $\overline{f(A)} \supseteq f(\overline{A})$. Hence $\overline{f(A)} = f(\overline{A})$. On the other hand, since f is open, the set $f(\mathrm{Int}(A))$ is open in $\mathbb{R}^n$, hence $f(\mathrm{Int}(A)) \subseteq \mathrm{Int}(f(A))$. Hence

$$\partial(f(A)) = \overline{f(A)} \setminus \mathrm{Int}(f(A)) \subseteq f(\overline{A}) \setminus f(\mathrm{Int}(A)) \subseteq f(\overline{A} \setminus \mathrm{Int}(A)) = f(\partial A) \tag{2.4.30}$$

By (2), this implies that $\partial(f(A))$ is Jordan negligible and $f(A) \in \mathcal{J}(\mathbb{R}^n)$ by Proposition 2.4.17. □

Example 2.4.18. — (Exercise) Consider the map $f : \mathbb{R}^n \to \mathbb{R}^n$ defined by $f(0) = 0$ and $f(x) = \left(\frac{\|x\|}{\|x\|_\infty}\right) x$ if $x \neq 0$. Then f is a Lipschitz homeomorphism, and $f\left(\overline{B_\infty(0,1)}\right)$ is exactly the closed unit ball $B_n = \left\{x \in \mathbb{R}^n \mid \|x\| \leq 1\right\}$, which shows that $B_n \in \mathcal{J}(\mathbb{R}^n)$ by Proposition 2.4.17.

We now describe more ways to construct Jordan sets.

Lemma 2.4.19. — *Let $A \in \mathcal{J}(\mathbb{R}^n)$. If $f : A \to \mathbb{R}^k$ ($k \in \mathbb{N}^+$) is a bounded uniformly continuous map, then* $\mathrm{graph}(f)$ *is Jordan negligible in $\mathbb{R}^n \times \mathbb{R}^k = \mathbb{R}^{n+k}$.*

Proof. — To distinguish the Jordan contents on $\mathbb{R}^n$, on $\mathbb{R}^k$ and on $\mathbb{R}^{n+k}$, we use respectively the symbol ν_n, ν_k and ν_{n+k} to denote the corresponding Jordan contents. Take any $\varepsilon > 0$. Since $\|\cdot\|_\infty$ is equivalent to the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^n$ and $\mathbb{R}^k$, the map f remains uniformly continuous and bounded as a map from $A \subseteq (\mathbb{R}^n, \|\cdot\|_\infty)$ to $(\mathbb{R}^k, \|\cdot\|)$. By this uniform continuity, there exists $\delta > 0$, such that for all $a, b \in A$ with $\|a-b\|_\infty < \delta$, we have $\|f(a) - f(b)\|_\infty < \varepsilon$. By Lemma 2.4.7, 2.4.10, Proposition 2.4.13 and Corollary 2.4.15, since $A \in \mathcal{J}(\mathbb{R}^n)$, there exists a rational $P \in \mathcal{P}(\mathbb{R}^n)$ which is a grid box for some large $N \in \mathbb{N}^+$ with $1/N \leq \delta$, such that $P \subseteq \mathrm{Int}(A)$, and $\nu_n(\mathrm{Int}(A) \setminus P) < \varepsilon$. Hence

$$\nu_n(A \setminus P) = \nu_n(A) - \nu_n(P) = \nu_n(\mathrm{Int}(A)) - \nu_n(P) = \nu_n(\mathrm{A} \setminus P) < \varepsilon. \tag{2.4.31}$$

Hence by definition of the Jordan content, reasoning as above, by replacing N with a suitable integer multiple of N if necessary (so we still have $1/N < \delta$), we may find a grid paved set $P_0 \in \mathcal{P}(\mathbb{R}^n)$ of level N such that $\nu_n(P_0) < \varepsilon$ and $A \setminus P \subseteq P_0$. Now let $P = \sqcup_{i=1}^m B_i$ be the decomposition of P into the (disjoint) union of grid boxes of level N, we see that

$$\mathrm{graph}(f\restriction_P) = \bigsqcup_{i=1}^m \mathrm{graph}\left(f\restriction_{B_i}\right). \tag{2.4.32}$$

Since for each $i = 1, \ldots, m$ and all $x, y \in B_i$, we have $\|x-y\|_\infty < 1/N < \delta$, it follows that $\|f(x) - f(y)\|_\infty < \varepsilon$. This implies that $f(B_i)$ is contained in a box C_i in $\mathbb{R}^k$ with each side of C_i has length at most ε. In particular, $\nu_k(C_i) \le \varepsilon^k$. Clearly $f(B_i) \subseteq B_i \times C_i$, and $B_i \times C_i$ is a box in $\mathbb{R}^{n+k}$ with

$$\nu_{n+k}(B_i \times C_i) = \nu_n(B_i) \cdot \nu_k(C_i) \le \varepsilon^k \nu_n(B_i). \tag{2.4.33}$$

This holds for each i, hence by (2.4.32), we have

$$\operatorname{graph}(f \restriction_P) \subseteq Q_1 := \bigcup_{i=1}^{m} (B_i \times C_i) \tag{2.4.34}$$

with $Q_1 \in \mathcal{P}(\mathbb{R}^{n+k})$ such that

$$\nu(Q_1) \le \sum_{i=1}^{m} \nu_{n+k}(B_i \times C_i) \le \varepsilon^k \sum_{i=1}^{m} \nu_n(B_i) = \varepsilon^k \nu_n(P) \le \varepsilon^k \nu_n(A). \tag{2.4.35}$$

On the other hand, let $M > 0$ be a constant such that $\|f(x)\|_\infty < M$ for any $x \in A$. We have, in particular, that

$$\operatorname{graph}(f \restriction_{P_0}) \subseteq Q_0 := P_0 \times \square_{-\mathbf{M}}^{\mathbf{M}}, \tag{2.4.36}$$

where $\mathbf{M} = (M, \dots, M) \in \mathbb{R}^k$. Clearly, $Q_0 \in \mathcal{P}(\mathbb{R}^{n+k})$, and

$$\nu_{n+k}(Q_0) = (2M)^k \nu_n(P_0) \le \varepsilon (2M)^k. \tag{2.4.37}$$

Set $Q := Q_0 \cup Q_1 \in \mathcal{P}(\mathbb{R}^{n+k})$. We see from $A \subseteq P_0 \cup P$ that

$$\operatorname{graph}(f) \subseteq \operatorname{graph}(f \restriction_{P_0}) \cup \operatorname{graph}(f \restriction_{P_1}) \subseteq Q_0 \cup Q_1 = Q, \tag{2.4.38}$$

and by (2.4.35) and (2.4.37), we have

$$\nu_{n+k}(Q) \le \nu_{n+k}(Q_0) + \nu_{n+k}(Q_1) \le \varepsilon (2M)^k + \varepsilon^k \nu_n(A). \tag{2.4.39}$$

Since $\varepsilon > 0$ is arbitrary, (2.4.38) and (2.4.39) together implies that graph(f) is a Jordan content zero set in $\mathbb{R}^{n+k}$. □

Lemma 2.4.20. — *If $A \in \mathcal{J}(\mathbb{R}^n)$, $B \in \mathcal{J}(\mathbb{R}^k)$ $n, k \in \mathbb{N}^+$, and identify $\mathbb{R}^n \times \mathbb{R}^k$ canonically with $\mathbb{R}^{n+k}$ (we write this as $\mathbb{R}^n \times \mathbb{R}^k = \mathbb{R}^{n+k}$ in the following), then*

(1) *we have*

$$\partial(A \times B) \subseteq \big((\partial A) \times B\big) \cup (A \times \partial B), \tag{2.4.40}$$

and both sides of (2.4.40) are Jordan negligible sets in $\mathbb{R}^{n+k}$;

(2) *denoting the Jordan content on $\mathcal{J}(\mathbb{R}^l)$ by ν_l for $l = n, k, n+k$, we have*

$$A \times B \in \mathcal{J}(\mathbb{R}^{n+k}) \tag{2.4.41}$$

and

$$\nu_{n+k}(A \times B) = \nu_n(A) \nu_k(B). \tag{2.4.42}$$

Proof. — By now, this lemma should be a routine exercise, and we leave the details to reader. □

Proposition 2.4.21. — *Suppose $A \in \mathcal{J}(\mathbb{R}^n)$ is compact, and $f, g : A \to \mathbb{R}$ continuous. Consider the set*

$$\mathsf{S}(f, g) := \big\{(x, t) \in \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1} \;\big|\; f(x) \le t \le g(x)\big\} \subseteq \mathbb{R}^{n+1}. \tag{2.4.43}$$

Then the following hold.

(1) *There exists $M > 0$, such that $|f(x)| \le M, |g| \le M$ for all $x \in A$.*
(2) *The set $\mathsf{S}(f, g)$ is compact in $\mathbb{R}^{n+1}$.*
(3) *We have*

$$\partial \mathsf{S}(f, g) \subseteq \Big((\partial A) \times [-M, M]\Big) \cup \operatorname{graph}(f) \cup \operatorname{graph}(g), \tag{2.4.44}$$

and both sides of (2.4.44) are Jordan negligible sets in $\mathbb{R}^{n+1}$. In particular, $\mathsf{S}(f, g) \in \mathcal{J}(\mathbb{R}^{n+k})$.

Proof. — (1). This holds since A is compact and continuous functions are bounded on compact sets.

(2). Clearly $\mathsf{S}(f,g)$ is bounded. So it suffices to show that $\mathsf{S}(f,g)$ is closed, or equivalently, that $\mathbb{R}^{n+1} \setminus \mathsf{S}(f,g)$ is open, which follows easily from the continuity of f and g (exercise).

(3). We first establish (2.4.44). Consider any $(x_0,t_0) \in \partial\mathsf{S}(f,g)$ with $x_0 \in \mathbb{R}^n$ and $t_0 \in \mathbb{R}$. Let $\pi_1 : \mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be the projection $(x,t) \mapsto x$. Clearly, $\mathsf{S}(f,g) \subseteq A \times [-M,M]$. Hence,

$$\partial\mathsf{S}(f,g) \subseteq \mathsf{S}(f,g) \subseteq A \times [-M,M]. \tag{2.4.45}$$

We now have

$$x_0 = \pi_1(x_0,t_0) \in \pi_1(\partial\mathsf{S}(f,g)) \subseteq \pi_1(A \times [-M,M]) \subseteq A. \tag{2.4.46}$$

To establish (2.4.44), it suffices to show that either $x_0 \in \partial A$ or $t_0 \in \{f(x_0), g(x_0)\}$. We suppose that neither of these holds, and show that this leads to a contradiction. Indeed, since $(x_0,t_0) \in \partial\mathsf{S}(f,g) \subseteq \mathsf{S}(f,g)$, we have $f(x_0) \le t_0 \le g(x_0)$. Since $t_0 \notin \{f(x_0), g(x_0)\}$, we must have

$$f(x_0) < t_0 < g(x_0). \tag{2.4.47}$$

On the other hand, we have $x_0 \in A \setminus \partial A = \operatorname{Int}(A)$. Hence by continuity of f and g as well as the openness of $\operatorname{Int}(A)$, we see that there exists an open neighborhood U_0 of x_0 and an open interval I_0 containing t_0, such that $U_0 \subseteq \operatorname{Int}(A)$ and

$$\forall (x,t) \in U_0 \times I_0, \qquad f(x) < t < g(x). \tag{2.4.48}$$

This means

$$(x_0,t_0) \in U_0 \times I_0 \subseteq \operatorname{Int}(\mathsf{S}(f,g)),$$

which contradicts $(x_0,t_0) \in \partial\mathsf{S}(f,g)$. This contradiction finishes the proof of (2.4.44).

Note that continuous functions on compact sets are automatically uniformly continuous, by Lemma 2.4.19, we see that $\operatorname{graph}(f)$ and $\operatorname{graph}(g)$ are Jordan negligible in $\mathscr{J}(\mathbb{R}^n)$. By Lemma 2.4.20, we see that $(\partial A) \times [-M,M]$ is Jordan negligible. Since Jordan negligible sets are stable under taking finite union, we see that the right side of (2.4.44), hence both sides of it, are Jordan negligible. That $\mathsf{S}(f,g) \in \mathscr{J}(\mathbb{R}^{n+1})$ now follows from Proposition 2.4.13, which completes the proof. $\square$

Example 2.4.22. — (Exercise) Using Proposition 2.4.21, we can easily show (e.g. by induction), that some simple geometric figures such as the ball $\overline{B}(a,r) = \left\{x \in \mathbb{R}^n \;\middle|\; \|x-a\| \le r\right\}$ in $\mathbb{R}^n$, or more generally the ellipsoid

$$E(a; r_1, \dots, r_n) := \left\{ x = (x_i) \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^n \frac{(x_i - a_i)^2}{r_i^2} \le 1 \right\} \tag{2.4.49}$$

where $a \in \mathbb{R}^n$, $r, r_1, \dots, r_n > 0$, as well as the n-simplex

$$\Delta_n = \left\{ x \in \mathbb{R}^n_{\ge 0} \;\middle|\; \sum_{i=1}^n x_i \le 1 \right\} \tag{2.4.50}$$

are all Jordan sets.

2.5. The Effect of Linear Transforms

Still fix a dimension $n \in \mathbb{N}^+$ in this section, and consider a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$. Here, by transformation, we mean an invertible map, so $T \in \mathrm{GL}_n(\mathbb{R}^n)$. For convenience of discussion, in this section, we shall also abuse the notation, and use the same symbol S to denote the matrix under the standard basis for any linear operator $S : \mathbb{R}^n \to \mathbb{R}^n$, so Sx can be computed by matrix multiplication if we regard vectors in $\mathbb{R}^n$ as row vectors.

Proposition 2.5.1. — *For any linear transformation $T \in \mathrm{GL}_n(\mathbb{R}^n) = \mathrm{Aut}(\mathbb{R}^n) = \mathrm{Iso}(\mathbb{R}^n, \mathbb{R}^n)$, the map*

$$\begin{aligned} \Phi_T : \mathscr{J}(\mathbb{R}^n) &\to \mathscr{J}(\mathbb{R}^n) \\ A &\mapsto T(A) \end{aligned} \tag{2.5.1}$$

is a well-defined bijection, and $\nu(T(A)) = |\det T|\nu(A)$ *for any* $A \in \mathcal{J}(\mathbb{R}^n)$*.*

Proof. — Note first that T is a Lipschitz homeomorphism, since $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\|Tx\| \le \|T\| \cdot \|x\|$ with $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ as its inverse. It follows from Proposition 2.4.17 that $T(A) \in \mathcal{J}(\mathbb{R}^n)$ for any $A \in \mathcal{J}(\mathbb{R}^n)$, i.e. Φ_T is well-defined. By a routine check $\Phi_{T^{-1}}$ and Φ_T are inverses to each other and Φ_T is indeed a well-defined bijection.

Now for any $S \in \mathrm{Aut}(\mathbb{R}^n)$, define $\nu_S : \mathcal{J}(\mathbb{R}^n) \to \mathbb{R}$ by $\nu_S := \nu \circ \Phi_S$. Then a routine check shows that $\big(\mathcal{J}(\mathbb{R}^n), \nu_S\big)$ remains a theory of volumes. Hence by Corollary 2.4.14, there exists a unique constant c_S such that $\nu_S = c_S\nu$, i.e.

$$\forall A \in \mathcal{J}(\mathbb{R}^n), \qquad \nu\big(S(A)\big) = \nu_S(A) = c_S\nu(A). \tag{2.5.2}$$

If $S_1, S_2 \in \mathrm{Aut}(\mathbb{R}^n)$, we have

$$\forall A \in \mathcal{J}(\mathbb{R}^n), \qquad c_{S_1S_2}\nu(A) = \nu\big((S_1S_2)(A)\big) = \nu\Big(S_1\big(S_2(A)\big)\Big) = c_{S_1}\nu\big(S_2(A)\big) = c_{S_1}c_{S_2}\nu(A). \tag{2.5.3}$$

Taking $A = \square_{\mathbf{0}}^{\mathbf{1}}$ in (2.5.3) gives

$$\forall S_1, S_2 \in \mathrm{Aut}(\mathbb{R}^n), \qquad c_{S_1S_2} = c_{S_1}c_{S_2}. \tag{2.5.4}$$

With these preparation, we aim to show that $c_T = |\det T|$, which will finish the proof.

Consider first the strictly positive definite matrix T^tT. From linear algebra (or one of the homework exercise using Lagrange multipliers), there exists an $O \in O(n, \mathbb{R})$, such that $T^tT = O^tDO = O^{-1}DO$, where $D = \mathrm{Diag}(d_1, \dots, d_n)$ is a diagonal matrix with $d_1, \dots, d_n$ on its diagonal, and each $d_i > 0$ (since T^tT is strictly positive definite). Consider the diagonal matrix

$$D_{1/2} := \mathrm{Diag}(\sqrt{d_1}, \dots, \sqrt{d_n}). \tag{2.5.5}$$

Clearly $D_{1/2}^2 = D$. Set $|T| := O^{-1}D_{1/2}O$. Then $|T|$ is invertible, and

$$|T|^2 = O^{-1}D_{1/2}^2O = O^{-1}DO = T^tT. \tag{2.5.6}$$

We define $V := T|T|^{-1}$ so that $T = V|T|$. Now for each $y \in \mathbb{R}^n$, let $x = |T|^{-1}y$, we then have $y = |T|x$, and we have

$$\|Vy\|^2 = \|V|T|x\|^2 = \|Tx\|^2 = \|y\|^2. \tag{2.5.7}$$

So $V \in O(n, \mathbb{R})$.

Clearly $V(B_n) = O(B_n) = O^{-1}(B_n) = B_n \in \mathcal{J}(\mathbb{R}^n)$, where B_n is the closed unit ball (see Example 2.4.18). Moreover, B_n has non-empty interior, which implies $\nu(B_n) > 0$. Hence taking $A = B_n$ in (2.5.2), we see that

$$c_V = c_O = c_{O^{-1}} = 1. \tag{2.5.8}$$

On the other hand, it is clear that

$$D_{1/2}(\square_{\mathbf{0}}^{\mathbf{1}}) = \square_{\mathbf{0}}^{(\sqrt{d_1}, \dots, \sqrt{d_n})}. \tag{2.5.9}$$

So

$$c_{D_{1/2}} = \det D_{1/2} = \sqrt{\det D} = \sqrt{\det(T^tT)} = \sqrt{(\det T)^2} = |\det T|. \tag{2.5.10}$$

From $|T| = O^{-1}D_{1/2}O$, we see that

$$c_{|T|} = c_{O^{-1}}c_{D_{1/2}}c_O = |\det T|. \tag{2.5.11}$$

Finally $T = V|T|$ implies

$$c_T = c_Vc_{|T|} = |\det T|. \tag{2.5.12}$$

$\square$

Remark 2.5.2. — (Exercise) If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map that is *not* invertible, then $\det T = 0$ and $\mathcal{R}(T)$ has a dimension strictly less than n. It follows that $T(A)$ is a Jordan negligible set in $\mathbb{R}^n$ (check it) for any $A \in \mathcal{J}(\mathbb{R}^n)$. Hence we have (check it)

$$\forall T \in \mathcal{L}(\mathbb{R}^n), A \in \mathcal{J}(\mathbb{R}^n), \qquad T(A) \in \mathcal{J}(\mathbb{R}^n) \text{ and } \nu\big(T(A)\big) = |\det T|\nu(A). \tag{2.5.13}$$

2.6. Riemann Integrability

In this section, we continue to fix a dimension $n \in \mathbb{N}^+$ unless stated otherwise. Before systematically developing the theory of Riemann integrable maps and their Riemann integrals, we make some preparation.

Definition 2.6.1. — Let E be a Banach space. A map $g : \mathbb{R}^n \to E$ is called **paved** if it is of the form

$$g = \sum_{i=1}^{m} c_i \chi_{B_i} \tag{2.6.1}$$

where $m \in \mathbb{N}$, each $c_i \in E$ and B_i is a box in $\mathbb{R}^n$ (we make the convention that the empty summation is 0, corresponding to the $m = 0$ case).

Roughly speaking, paved maps from $\mathbb{R}^n$ to E are the "linear combinations" of characteristic functions of boxes with coefficients in E. Intuitively (and we will make this rigorous later), paved maps are the obvious ones that one should be able to integrate. A natural idea for defining Riemann integrable maps is to consider the compactly supported maps that admit suitable approximations by paved maps. We now formulate this idea precisely.

Definition 2.6.2. — Let E be a Banach space, $f : \mathbb{R}^n \to E$ an arbitrary map, $g : \mathbb{R}^n \to E$ a paved map. For $\varepsilon, \delta > 0$, we say g is an (ε, δ)-**approximation** of f, if there exists an $A \in \mathcal{J}(\mathbb{R}^n)$, such that

$$\nu(A) \leq \delta, \qquad \text{and} \qquad \forall x \in \mathbb{R}^n \setminus A, \qquad \|f(x) - g(x)\| \leq \varepsilon. \tag{2.6.2}$$

We say the map f is **Jordan contented**, or simply **contented**, if f is bounded and of compact support, and for each $\varepsilon, \delta > 0$, there exists an (ε, δ)-approximation of f by a paved map.

Remark 2.6.3. — Even if $f : \mathbb{R}^n \to E$ admits an (ε, δ)-approximation by a paved map for any $\varepsilon, \delta > 0$, the map f may still not be bounded, nor must it be compactly supported. For such an example of f without boundedness, let $(x_m)_{m \geq 1} \subseteq \mathbb{R}^n$ be a sequence of distinct elements with $\lim_{m \to \infty} x_m = 0$ and define $f(x_m) = m$ and $f(x) = 0$ if $x \notin \{x_m \mid m \in \mathbb{N}^+\}$; for such an example of f with uncompact support, suppose B_N is the unique box whose interior is $B_\infty(0, N) := \{x \in \mathbb{R}^n \mid \|x\|_\infty < N\}$, fix any $\xi \in E \setminus \{0\}$, then the map

$$f_\xi = \sum_{N=1}^{\infty} \frac{\xi}{N} \left(\chi_{B_{N+1}} - \chi_{B_N}\right) \tag{2.6.3}$$

is a well-defined map from $\mathbb{R}^n$ to E that admits an (ε, δ)-approximation by a paved map for all $\varepsilon, \delta > 0$, but $\operatorname{supp}(f) = \mathbb{R}^n$, which is not compact. This means that the assumptions of being compactly supported and being bounded are both necessary in the definition of a contented map.

We track here some useful technical results concerning paved maps and (ε, δ)-approximations.

Lemma 2.6.4. — *Let E be a Banach space, $g : \mathbb{R}^n \to E$. The following are equivalent:*

(1) *g is paved;*
(2) *g is of the form*

$$\sum_{i=1}^{m} c_i \chi_{P_i} \tag{2.6.4}$$

where each $c_i \in E$ and each $P_i \in \mathcal{P}(\mathbb{R}^n)$;

(3) *there exists $m \in \mathbb{N}$ and a* disjoint *family $(B_i)_{i=1}^m$ of boxes ($m = 0$ corresponds to the empty family), and $(c_i)_{i=1}^m$ a finite sequence in E ($m = 0$ corresponds to the empty sequence), such that*

$$g = \sum_{i=1}^{m} c_i \chi_{B_i}. \tag{2.6.5}$$

Proof. — By now, this should be a routine exercise, and we leave the details as an exercise. □

Lemma 2.6.5. — *Let E be a Banach space. If $P \in \mathcal{P}(\mathbb{R}^n)$ and $g : \mathbb{R}^n \to E$ is paved, then the map $\chi_P g$ is also paved. More generally, if $s : \mathbb{R}^n \to \mathbb{R}$ is paved, then so is $sg : \mathbb{R}^n \to E$.*

Proof. — Write

$$g = \sum_{i=1}^{m} c_i \chi_{B_i}, \tag{2.6.6}$$

with $c_i \in E$ and B_i a box for each i. Then

$$\chi_P g = \sum_{i=1}^{m} c_i \chi_P \chi_{B_i} = \sum_{i=1}^{m} c_i \chi_{P\cap B_i}. \tag{2.6.7}$$

Since $P \cap B_i \in \mathcal{P}(\mathbb{R}^n)$ for each i, then $\chi_P g$ remains paved by Lemma 2.6.4. The last statement follows from the first by taking linear combinations. □

Lemma 2.6.6. — *Let E be a Banach space, $\varepsilon, \delta > 0$, $f : \mathbb{R}^n \to E$ any map, $g : \mathbb{R}^n \to E$ a paved map, and $P \in \mathcal{P}(\mathbb{R}^n)$. If g is an (ε, δ)-approximation of f, then $\chi_P g$ is an (ε, δ)-approximation of $\chi_P f$.*

Proof. — Since g is an (ε, δ)-approximation of f, there exists an $A \in \mathcal{J}(\mathbb{R}^n)$, such that $\nu(A) \leq \delta$, and

$$\big\{x \in \mathbb{R}^n \bigm| \|f(x) - g(x)\| > \varepsilon\big\} \subseteq A. \tag{2.6.8}$$

Clearly,

$$\forall x \in \mathbb{R}^n, \qquad \|(\chi_P f - \chi_P g)(x)\| = |\chi_P(x)|\|f(x) - g(x)\| \leq \|f(x) - g(x)\|, \tag{2.6.9}$$

it follows that

$$\big\{x \in \mathbb{R}^n \bigm| \|\chi_P f(x) - \chi_P g(x)\| > \varepsilon\big\} \subseteq \big\{x \in \mathbb{R}^n \bigm| \|f(x) - g(x)\| > \varepsilon\big\} \subseteq A, \tag{2.6.10}$$

which, by Lemma 2.6.5, implies that $\chi_P g$ is an (ε, δ)-approximation of $\chi_P f$. □

Besides giving a better idea of the (ε, δ)-approximation of a contented map by paved maps, the following result will also be used in the proof of one of the implications in Theorem 2.6.20.

Proposition 2.6.7. — *Let E be a Banach space, $f : \mathbb{R}^n \to E$ a compactly supported map, $g : \mathbb{R}^n \to E$ a paved map, $\varepsilon, \delta > 0$, and $P \in \mathcal{P}(\mathbb{R}^n)$ such that $\overline{P} \supseteq \operatorname{supp}(f)$. If g is an (ε, δ)-approximation of f, then $\chi_P g$ is also an (ε, δ)-approximation of f by a paved map.*

Proof. — We have seen in Lemma 2.6.6 that $\chi_P g$ is an (ε, δ)-approximation of $\chi_P f$, i.e.

$$\exists A \in \mathcal{J}(\mathbb{R}^n), \qquad \nu(A) \leq \delta \quad \text{ and } \quad \Big\{x \in \mathbb{R}^n \Bigm| \|(\chi_P f)(x) - (\chi_g P)(x)\| > \varepsilon\Big\} \subseteq A, \tag{2.6.11}$$

On the other hand, since $\overline{P} \supseteq \operatorname{supp}(f)$, we see that $(\chi_P f)(x) = 0 = f(x)$ for all $x \in \mathbb{R}^n \setminus \overline{P}$, while $(\chi_P f)(x) = f(x)$ for all $x \in P$ since in this case $\chi_P(x) = 1$. Hence

$$\Big\{x \in \mathbb{R}^n \Bigm| (\chi_P f)(x) \neq f(x)\Big\} \subseteq \overline{P} \setminus P = \partial P. \tag{2.6.12}$$

Set $B := A \cup \partial P \in \mathcal{J}(\mathbb{R}^n)$. Clearly,

$$\nu(A) \leq \nu(B) \leq \nu(A) + \nu(\partial P) = \nu(A), \tag{2.6.13}$$

so

$$\nu(B) = \nu(A) \leq \delta, \tag{2.6.14}$$

while by (2.6.11) and (2.6.12), we have

$$\Big\{x \in \mathbb{R}^n \Bigm| \|f(x) - (\chi_P g)(x)\| > \varepsilon\Big\} \subseteq B. \tag{2.6.15}$$

The proposition is now established by (2.6.14), (2.6.15) and Lemma 2.6.5. □

As mentioned before Definition 2.6.2, the principle reason that we introduce contented maps is that we want them to be a class of compactly supported maps that can be approximated by paved maps in a good sense that is suitable for developing integration theory. Motivated by measuring the difference of Riemann sums, we introduce the following terminologies and notation.

Definition 2.6.8. — Let $f : \mathbb{R}^n \to E$ be a map with $n \in \mathbb{N}^+$ and E a Banach space. The **amplitude** of f on any non-empty $S \subseteq \mathbb{R}^n$, denoted by $W(f, S)$, is defined to be the number

$$W(f,S) := \sup\big\{\|f(x) - f(y)\| \;\big|\; x, y \in S\big\} \in [0, +\infty]. \tag{2.6.16}$$

The **total oscillation** of f on a Jordan set $A \in \mathcal{J}(\mathbb{R}^n)$ with $\nu(A) > 0$, denoted by $\omega(f, A)$, is defined to be the number

$$\omega(f, A) := W(f, A)\nu(A) \in [0, +\infty]. \tag{2.6.17}$$

If $A \subseteq \mathbb{R}^n$ is Jordan negligible and $f(A)$ is bounded[(4)], then we also define $\omega(f, A) := 0$, and call it the **total oscillation** of f on A, this applies even to the case of $A = \emptyset$.

Definition 2.6.9. — Let $A \in \mathcal{J}(\mathbb{R}^n)$. By a **Jordan partition**, we mean a finite disjoint family $\Delta = (A_i)_{i=1}^m$ of Jordan sets such that $A = \sqcup_{i=1}^m A_i$. In this case, for any map $f : \mathbb{R}^n \to E$ to a Banach space E, we set

$$\omega(f, \Delta) := \sum_{i=1}^{m} \omega(f, A_i), \tag{2.6.18}$$

provided that for each i, $\nu(A_i) = 0$ whenever $f(A_i)$ is unbounded. For this f and Δ, elements of the form[(5)]

$$\sum_{A_i \neq \emptyset} f(x_i)\nu(A_i), \qquad x_i \in A_i \tag{2.6.19}$$

are called **Riemann sums** of f with respect to Δ. We say another Jordan partition $\Delta' = (A'_j)_{j=1}^p$ of A is **finer than** Δ, or Δ is **coarser than** Δ', if each A'_j is contained in some A_i.

Notation 2.6.10. — *We use the notation $S(f, \Delta)$ to denote the set of all Riemann sums of f with respect to Δ, where f and Δ are as in Definition 2.6.9.*

Proposition 2.6.11. — *Let E be a Banach space and $f : \mathbb{R}^n \to E$ any map. If Δ_1 and Δ_2 are two Jordan partitions of any fixed $A \in \mathcal{J}(\mathbb{R}^n)$ such that Δ_2 is finer than Δ_1, then*

$$\omega(f, \Delta_2) \leq \omega(f, \Delta_1). \tag{2.6.20}$$

Proof. — Let $\Delta_1 = (A_i)_{i=1}^m$ and $\Delta_2 = (B_j)_{j=1}^l$. Since Δ_2 is finer than Δ_1, we see that

$$\forall i \in \{1, \ldots, m\}, \qquad A_i = \sqcup_{B_j \subseteq A_i} B_j. \tag{2.6.21}$$

Clearly, $W(f, B_j) \leq W(f, A_i)$ if $B_j \subseteq A_i$. Hence

$$\begin{aligned} \omega(f, A_i) &= W(f, A_i)\nu(A_i) = W(f, A_i) \sum_{B_j \subseteq A_i} \nu(B_j) \\ &\geq \sum_{B_j \subseteq A_i} W(f, B_j)\nu(B_j) = \sum_{B_j \subseteq A_i} \omega(f, B_j). \end{aligned} \tag{2.6.22}$$

Taking the sum of both sides of (2.6.22) over i, we obtain $\omega(f, \Delta_1) \geq \omega(f, \Delta_2)$. □

Remark 2.6.12. — (Exercise) Intuitively speaking, for any $f : \mathbb{R}^n \to E$, and any Jordan partition Δ of a Jordan set $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$, the quantity $\omega(f, \Delta)$ measures the margin of error of the corresponding Riemann sums. More precisely, if $s_1, s_2 \in S(f, \Delta)$, then

$$\|s_1 - s_2\| \leq \omega(f, \Delta). \tag{2.6.23}$$

Then Proposition 2.6.11 says that for any $\varepsilon > 0$, if there exists a Jordan partition Δ of A such that $\omega(f, \Delta) \leq \varepsilon$, then for any Jordan partition Δ' of A that is finer than Δ and any $s'_1, s'_2 \in S(f, \Delta')$, we will have

$$\|s'_1 - s'_2\| \leq \omega(f, \Delta') \leq \omega(f, \Delta) \leq \varepsilon. \tag{2.6.24}$$

---

(4) So we have $W(f, A)$ is finite in the case $A \neq \emptyset$, and we won't have to deal with $\infty \cdot 0$ when defining $\omega(f, A) := W(f, A)\nu(A)$.
(5) We make the usual convention that the empty sum is 0.

This gives us a precise way to control the "uncertainty" of Riemann sums, which will be important when we define Riemann integrals of Riemann integrable maps.

Traditionally, instead of the more general and more flexible Jordan partition, the theory of Riemann integration is closely related to special Jordan partition where each member is a (closed) box.

Definition 2.6.13. — An **open (resp. closed) box** in $\mathbb{R}^n$ is defined to be the interior (resp. closure) of a (unique) box in $\mathbb{R}^n$.

Note that as Jordan sets, all non-empty (open/closed) boxes are non-degenerate in the following sense.

Definition 2.6.14. — We call a Jordan set **non-degenerate** if its volume is non-zero. We call a Jordan partition **non-degenerate** if each of its member is a non-degenerate Jordan set. If $P \in \mathcal{P}(\mathbb{R}^n)$, a Jordan partition $(B_i)_{i=1}^m$ of P is called a **box partition** if each B_i is a non-empty box.

Lemma 2.6.15. — *Let E be a Banach space, $n \in \mathbb{N}^+$, and $f : \mathbb{R}^n \to E$ a map. Then for any* finite *non-empty collection $(A_i)_{i=1}^m$ of non-degenerate Jordan sets, if f is unbounded on $\cup_{i=1}^m A_i$, then*

$$\sum_{i=1}^m \omega(f, A_i) = +\infty. \tag{2.6.25}$$

Proof. — Since f is unbounded on $\cup_{i=1}^m A_i$, there is a sequence $(x_k)_{k\geq 1}$ in $\cup_{i=1}^m A_i$ such that $\lim_{k\to+\infty}\|f(x_k)\| = +\infty$. Since m is finite, there exists an index $j \in \{1, \dots, m\}$ and an infinite subsequence $(y_k)_{k\geq 1}$ of $(x_k)_{k\geq 1}$ such that $y_k \in A_j$ for all $k \geq 1$. It is then clear that $W(f, A_j) = +\infty$ since we also have $\lim_{k\to+\infty}\|f(y_k)\| = +\infty$. Consequently, by non-degenerateness of A_j, we must have $\omega(f, A_j) = +\infty$ for this j and (2.6.25) holds. □

As a final piece of our preparatory work, we shall also need a good theory of negligible sets to deal with the points at which a Riemann integrable map is not continuous, just as in the familiar one-dimensional case.

Proposition 2.6.16. — *For any $A \subseteq \mathbb{R}^n$ $(n \in \mathbb{N}^+)$. The following are equivalent.*

(1) *For any $\varepsilon > 0$, there exists a countable cover $(B_j)_{j\in\mathbb{N}^+}$ of A by boxes, such that*

$$\sum_{j=1}^\infty \nu(B_j) \leq \varepsilon. \tag{2.6.26}$$

(2) *For any $\varepsilon > 0$, there exists a countable cover $(B_j)_{j\in\mathbb{N}^+}$ of A by* open *boxes, such that*

$$\sum_{j=1}^\infty \nu(B_j) \leq \varepsilon. \tag{2.6.27}$$

(3) *For any $\varepsilon > 0$, there exists a countable cover $(B_j)_{j\in\mathbb{N}^+}$ of A by* closed *boxes, such that*

$$\sum_{j=1}^\infty \nu(B_j) \leq \varepsilon. \tag{2.6.28}$$

(4) *For any $\varepsilon > 0$, there exists a countable cover $(B_j)_{j\in\mathbb{N}^+}$ of A by Jordan sets, such that*

$$\sum_{j=1}^\infty \nu(B_j) \leq \varepsilon. \tag{2.6.29}$$

Proof. — Exercise. □

Definition 2.6.17. — A set $A \subseteq \mathbb{R}^n$ is called **(Lebesgue) negligible** if it satisfies any of the equivalent condition in Proposition 2.6.16. We denote by $\mathcal{N}(\mathbb{R}^n)$ the collection of all negligible sets in $\mathbb{R}^n$.

Proposition 2.6.18. — *The collection $\mathcal{N}(\mathbb{R}^n)$ is stable under taking subsets and countable unions.*

Proof. — This is a routine exercise. □

Remark 2.6.19. — Note that all Jordan negligible sets are Lebesgue negligible. The converse, however, is not true. As an easy example, consider $\mathbb{Z}^n \subseteq \mathbb{R}^n$, which is unbounded, hence $\mathbb{Z}^n$ is not a Jordan set, and consequently not Jordan negligible. But being countable, $\mathbb{Z}^n$ is negligible. Even for bounded sets, there are negligible sets that are not Jordan negligible, e.g. $\mathbb{Q}^n \cap \square_{\mathbf{0}}^{\mathbf{1}}$.

We are finally ready to establish the following theorem that incorporating various perspectives of what will later be called Riemann integrable maps, which is fundamental in the whole theory of Riemann integration theory.

Theorem 2.6.20. — *Let E be a Banach space, $n \in \mathbb{N}^+$, and $f : \mathbb{R}^n \to E$ a compactly supported map. The following are equivalent.*

(1) *For any non-degenerate $A \in \mathcal{J}(\mathbb{R}^n)$ containing $\operatorname{supp}(f)$, and any $\varepsilon > 0$, there exists a non-degenerate Jordan partition Δ of A, such that $\omega(f, \Delta) \leq \varepsilon$.*

(2) *For any non-degenerate $A \in \mathcal{J}(\mathbb{R}^n)$ containing $\operatorname{supp}(f)$, any $\delta > 0$ and any $\varepsilon > 0$, there exists a non-degenerate Jordan partition $\Delta = (A_i)_{i=0}^m$ of A, such that $\nu(A_0) < \delta$, A_i is open for $i = 1, \ldots, m$, and $\omega(f, \Delta) \leq \varepsilon$.*

(3) *For any non-empty paved set $P \subseteq \mathbb{R}^n$ containing $\operatorname{supp}(f)$, and any $\varepsilon > 0$, there exists a box partition Δ of P such that $\omega(f, \Delta) \leq \varepsilon$.*

(4) *For any non-empty paved set $P \subseteq \mathbb{R}^n$ with $\overline{P} \supseteq \operatorname{supp}(f)$, and any $\varepsilon > 0$, there exists a box partition $(B_i)_{i=1}^p$ of P, such that*

$$\sum_{i=1}^{m} \omega(f, \overline{B_i}) \leq \varepsilon. \tag{2.6.30}$$

(5) *The map f is Jordan contented.*

(6) *The set $D(f)$ of discontinuous points of f is negligible in $\mathbb{R}^n$ and f is bounded.*

Proof. — The structure of our proof goes as follows.

$$(1) \implies (2) \implies (6) \implies (5) \implies (4) \implies (3) \implies (1).$$

(1) implies (2). Assume (1) holds, then Lemma 2.6.15 implies the boundedness of f, i.e. there exists an $M > 0$ such that $M \geq \|f(x)\|$ for all $x \in \mathbb{R}^n$. Take any $\varepsilon > 0$ and any non-degenerate $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$. By (1), there exists a non-degenerate Jordan partition $\Delta' = (A'_i)_{i=1}^m$ of A, such that

$$\omega(f, \Delta') = \sum_{i=1}^{m} \omega(f, A'_i) \leq \frac{\varepsilon}{2}. \tag{2.6.31}$$

For each $i = 1, \ldots, m$, we know $\operatorname{Int}(A'_i) \in \mathcal{J}(\mathbb{R}^n)$ and $\nu(\operatorname{Int}(A'_i)) = \nu(A_i) > 0$. We set $A_i := \operatorname{Int}(A_i)$ for $i = 2, \ldots, m$, and choose a small non-empty box $B \in \mathcal{P}(\mathbb{R}^n)$ with $\overline{B} \subseteq \operatorname{Int}(A'_1)$ and

$$\nu(B) \leq \min\left(\frac{\varepsilon}{4M}, \delta\right), \tag{2.6.32}$$

and set $A_1 := \operatorname{Int}(A'_1) \setminus \overline{B}$. Finally, set $A_0 := A \setminus \sqcup_{i=1}^m A_i$. Then $\nu(A_0) \geq \nu(B) > 0$ since $B \subseteq A_0$, A_i is non-empty open for $i = 1, \ldots, m$, and $\Delta := (A_i)_{i=0}^m$ is a non-degenerate partition of A. Moreover, since $\nu(A_i) = \nu(A'_i)$ for all $i = 2, \ldots, m$ and

$$\nu(A_1) = \nu(\operatorname{Int}(A'_1)) - \nu(\overline{B}) = \nu(A'_1) - \nu(B), \tag{2.6.33}$$

we have

$$\begin{aligned}
\nu(A_0) &= \nu(A) - \sum_{i=1}^{m} \nu(A_i) \\
&= \nu(A) - \nu(A_1) - \sum_{i=2}^{m} \nu(A'_i) \\
&= \left(\sum_{i=1}^{m} \nu(A'_i)\right) - \Big(\nu(A'_1) - \nu(B)\Big) - \sum_{i=2}^{m} \nu(A'_i) \\
&= \nu(A'_1) - \nu(A_1) \\
&= \nu(B) \leq \min\left(\frac{\varepsilon}{4M}, \delta\right).
\end{aligned} \tag{2.6.34}$$

It follows that

$$\omega(f, A_0) = W(f, A_0)\nu(A_0) \leq 2M \frac{\varepsilon}{4M} = \frac{\varepsilon}{2}. \tag{2.6.35}$$

For $i = 1, \ldots, m$, we have $A_i \subseteq A'_i$, so

$$\sum_{i=1}^{m} \omega(f, A_i) = \sum_{i=1}^{m} W(f, A_i)\nu(A_i) \le \sum_{i=1}^{m} W(f, A'_i)\nu(A'_i) = \sum_{i=1}^{m} \omega(f, A'_i) = \omega(f, \Delta') \le \frac{\varepsilon}{2}. \tag{2.6.36}$$

It follows from (2.6.35) and (2.6.36) that

$$\omega(f, \Delta) = \omega(f, A_0) + \sum_{i=1}^{m} \omega(f, A_i) \le \varepsilon. \tag{2.6.37}$$

This proves (1) implies (2).

(2) implies (6). The boundedness of f follows from Lemma 2.6.15.

For each $k \in \mathbb{N}^+$, set

$$D_k(f) := \left\{ x \in \mathbb{R}^n \;\middle|\; \omega_x(f) \ge \frac{1}{k} \right\}. \tag{2.6.38}$$

Then $D(f) = \cup_{k \in \mathbb{N}^+} D_k(f)$. To show that $D(f)$ is negligible, it suffices to show that for each fixed $k \in N^+$, the set $D_k(f)$ is negligible.

Now take any non-degenerate $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$ (this is possible since $\operatorname{supp}(f)$ is compact, hence bounded, so we may e.g. take A to be a large closed box). By (2), there exists a non-degenerate Jordan partition $(A_i)_{i=0}^m$ of A, such that A_i is open for all $i = 1, \ldots, m$, $\nu(A_0) \le \frac{\varepsilon}{2}$ and $\omega(f, \Delta) \le \frac{\varepsilon}{2k}$. Define

$$I_k := \left\{ i \in \{1, \ldots, m\} \;\middle|\; D_k(f) \cap A_i \ne \emptyset \right\}. \tag{2.6.39}$$

Clearly, I_k is finite. Since $f \restriction_{\mathbb{R}^n \setminus \operatorname{supp}(f)} = 0$, and $\mathbb{R}^n \setminus \operatorname{supp}(f)$ is open (because $\operatorname{supp}(f)$ is compact, so closed), we see that $D_k(f) \subseteq D(f) \subseteq \operatorname{supp}(f) \subseteq A = \sqcup_{i=0}^m A_i$, which implies

$$D_k(f) \subseteq A_0 \cup \left(\cup_{i \in I_k} A_i \right). \tag{2.6.40}$$

Moreover, for any $i \in I_k$, choose any $x \in A_i \cap D_k(f)$, note that A_i is open neighborhood of x and $\omega_x(f) \ge \frac{1}{k}$, we have $W(f, A_i) \ge \frac{1}{k}$. Consequently,

$$\forall i \in I_k, \qquad \omega(f, A_i) = W(f, A_i)\nu(A_i) \ge \frac{\nu(A_i)}{k}. \tag{2.6.41}$$

Hence

$$\frac{\varepsilon}{2k} \ge \omega(f, \Delta) \ge \sum_{i \in I_k} \omega(f, A_i) \ge \frac{1}{k} \sum_{i \in I_k} \nu(A_i). \tag{2.6.42}$$

This implies

$$\nu(A_0) + \sum_{i \in I_k} \nu(A_i) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{2.6.43}$$

By (2.6.40) and (2.6.43), since $\varepsilon > 0$ is arbitrary, $D_k(f)$ is indeed negligible(6) for each $k \in \mathbb{N}^+$.

(6) implies (5). It suffices to find, for any fixed $\varepsilon > 0$ and $\delta > 0$, an (ε, δ)-approximation of f by a paved map g. For this purpose, we will choose a suitable open cover of the compact set $\operatorname{supp}(f)$, and extract from this cover a finite sub-cover.

Since $D(f)$ is negligible, there exists a countable cover $(B_k)_{k \in \mathbb{N}^+}$ of $D(f)$ by open boxes with total volumes bounded by δ. For any $x \in \operatorname{supp}(f) \setminus D(f)$, since f is continuous at x, there exists a rational box C_x with $x \in \operatorname{Int}(C_x)$, such that $W(f, C_x) \le \varepsilon$. Clearly, we have the following cover of the compact set $\operatorname{supp}(f)$ by open sets:

$$\operatorname{supp}(f) = D(f) \cup \big(\operatorname{supp}(f) \setminus D(f)\big) \subseteq \left(\cup_{n \in \mathbb{N}^+} B_n\right) \cup \left(\cup_{x \in \operatorname{supp}(f) \setminus D(f)} \operatorname{Int}(C_x)\right). \tag{2.6.44}$$

---

(6) The proof actually shows the stronger result that each $D_k(f)$ is Jordan negligible. However, $D(f) = \cup_{k=1}^{\infty} D_k(f)$ itself may well be merely negligible but not Jordan negligible. See Remark 2.6.19.

Hence, there exists a finite sub-cover, i.e. there exists l distinct $n_1, \dots, n_l \in \mathbb{N}^+$, and m distinct $x_1, \dots, x_m \in \operatorname{supp}(f) \setminus D(f)$, such that

$$\operatorname{supp}(f) \subseteq \left(\cup_{i=1}^{l} B_{n_i}\right) \cup \left(\cup_{j=1}^{m} \operatorname{Int}(C_{x_j})\right) \subseteq \left(\cup_{i=1}^{l} B_{n_i}\right) \cup \left(\cup_{j=1}^{m} C_{x_j}\right). \tag{2.6.45}$$

Clearly, $\cup_{j=1}^{m} C_{x_j} \in \mathcal{P}(\mathbb{R}^n)$, so there exists a decomposition

$$\cup_{j=1}^{m} C_{x_j} = \sqcup_{k=1}^{p} D_k \tag{2.6.46}$$

where each D_k is a non-degenerate box contained in some C_{x_j}. Take any $y_k \in D_k$, and set

$$g := \sum_{j=1}^{m} f(y_k)\chi_{D_k},$$

which is clearly a paved map. We also set $A := \cup_{i=1}^{l} B_{n_i} \in \mathcal{J}(\mathbb{R}^n)$. By our choice of $(B_k)_{k\in\mathbb{N}^+}$, we have

$$\nu(A) \le \sum_{i=1}^{l} \nu(B_{n_i}) \le \sum_{k=1}^{\infty} \nu(B_k) \le \delta. \tag{2.6.47}$$

The proof is complete if we could show

$$\forall x \in \mathbb{R}^n \setminus A, \qquad \|f(x) - g(x)\| \le \varepsilon, \tag{2.6.48}$$

which we now establish. Take any $x \in \mathbb{R}^n \setminus A$. If $x \in \sqcup_{k=1}^{p} D_k = \cup_{j=1}^{m} C_{x_j}$, then $x \in D_k \subseteq C_{x_j}$ for some k and j. Since we also have $y_k \in D_k \subseteq C_{x_j}$ for the same k and j and $W(f, C_{x_j}) \le \varepsilon$, it follows that $\|f(x) - f(y_k)\| \le \varepsilon$. But $g(x) = f(y_k)$ from our definition of g, we see that

$$\forall x \in \sqcup_{k=1}^{p} D_k, \qquad \|f(x) - g(x)\| \le \varepsilon. \tag{2.6.49}$$

On the other hand,

$$\forall x \in (\mathbb{R}^n \setminus A) \setminus \sqcup_{k=1}^{p} D_k \subseteq \mathbb{R}^n \setminus \operatorname{supp}(f), \qquad f(x) = g(x) = 0. \tag{2.6.50}$$

Now (2.6.48) follows from (2.6.49) and (2.6.50).

(5) implies (4). Since f is contented, f is bounded. We may then take $M > 0$ so that $\|f(x)\| \le M$ for all $x \in \mathbb{R}^n$. Let $\varepsilon > 0$ be the constant in (4). Take any $\varepsilon_0 > 0$ and $\delta_0 > 0$, which will be to determined later, we can then take an $(\varepsilon_0, \delta_0)$-approximation of f by a paved map $g_0 : \mathbb{R}^n \to E$.

Now take any non-empty $P \in \mathcal{P}(\mathbb{R}^n)$ with $\overline{P} \supseteq \operatorname{supp}(f)$. By Proposition 2.6.7 and Lemma 2.6.5, we know that $g_1 := \chi_P g_0$ remains paved and an $(\varepsilon_0, \delta_0)$-approximation of f. Hence there exists $A_0 \in \mathcal{J}(\mathbb{R}^n)$, such that $\nu(A_0) \le \delta_0$, and

$$\forall x \in \mathbb{R}^n \setminus A_0, \qquad \|f(x) - g_1(x)\| \le \varepsilon_0. \tag{2.6.51}$$

Since $A_0 \in \mathcal{J}(\mathbb{R}^n)$, we can choose $A \in \mathcal{P}(\mathbb{R}^n)$ with $A \supseteq A_0$ and $\nu(A) \le 2\delta_0$. Finally, set

$$g := \chi_{\mathbb{R}^n \setminus A} g_1 = \chi_{P \setminus A} g_0, \tag{2.6.52}$$

we see from (2.6.51) that

$$\begin{aligned} \forall x \in \mathbb{R}^n \setminus A \subseteq \mathbb{R}^n \setminus A_0, \qquad \|f(x) - g(x)\| &= \left\|\chi_{\mathbb{R}^n \setminus A}(x)(f(x) - g_1(x))\right\| \\ &\le \|f(x) - g_1(x)\| \le \varepsilon_0. \end{aligned} \tag{2.6.53}$$

We are now ready to find a suitable box partition of P as in (4). For this, set $P_0 := P \cap A \in \mathcal{P}(\mathbb{R}^n)$ (recall that we have chosen $A \in \mathcal{P}(\mathbb{R}^n)$) and $P_1 := P \setminus P_0 = P \setminus A$, so that $P = P_0 \sqcup P_1$ and $g = \chi_{P_1} g_1$ from our construction. Since g is paved, we may write

$$g = \sum_{i=1}^{m} c_i \chi_{C_i}, \tag{2.6.54}$$

where $c_1, \dots, c_m \in E \setminus \{0\}$ and $(C_i)_{i=1}^{m}$ is a disjoint family of boxes. If $x \in C_i$ for some i, then $0 \ne c_i = g(x) = \chi_{P_1}(x) g_1(x)$, so $\chi_{P_1}(x) \ne 0$ and $x \in P_1$. This implies that $\sqcup_{i=1}^{m} C_i \subseteq P_1$. Clearly $P_1 \setminus \sqcup_{i=1}^{m} C_i \in \mathcal{P}(\mathbb{R}^n)$, so there is a finite disjoint family of non-empty boxes $(C_i)_{i=m+1}^{l}$ with $m \le l$ (the case $l = m$ corresponds to the empty family), such that $P_1 \setminus \sqcup_{i=1}^m C_i = \sqcup_{i=m+1}^l C_i$, and P_1 is the disjoint union $\sqcup_{i=1}^l C_i$. By (2.6.53) and the fact that $P_1 = P \setminus P_0 = P \setminus A \subseteq \mathbb{R}^n \setminus A$, we see

$$\begin{aligned} \forall i = 1, \dots, m,\ \forall x, y \in C_i, \qquad & \|f(x) - f(y)\| \\ & \le \|f(x) - g(x)\| + \|g(x) - g(y)\| + \|f(y) - g(y)\| \\ & \le \varepsilon_0 + \|c_i - c_i\| + \varepsilon_0 = 2\varepsilon_0, \end{aligned} \tag{2.6.55}$$

and

$$\begin{aligned} \forall i = m+1, \dots, l,\ \forall x, y \in C_i, \qquad & \|f(x) - f(y)\| \le \|f(x)\| + \|f(y)\| \\ & = \|f(x) - g(x)\| + \|f(y) - g(y)\| \le 2\varepsilon_0. \end{aligned} \tag{2.6.56}$$

Hence

$$\forall i = 1, \dots, l, \qquad W(f, C_i) \le 2\varepsilon_0. \tag{2.6.57}$$

Use Lemma 2.4.7, we may choose, for each $i = 1, \dots, l$, a box B_i such that

$$\overline{B_i} \subseteq C_i \qquad \text{and} \qquad \nu(C_i \setminus \overline{B_i}) = \nu(C_i) - \nu(\overline{B_i}) = \nu(C_i) - \nu(B_i) \le \frac{\delta_0}{l+1}. \tag{2.6.58}$$

Clearly $(B_i)_{i=1}^l$ remains a finite disjoint family, and

$$\sqcup_{i=1}^l B_i \subseteq \sqcup_{i=1}^l C_i = P_1. \tag{2.6.59}$$

Set $Q_0 := P_0 \sqcup (\sqcup_{i=1}^l (C_i \setminus B_i)) \in \mathcal{P}(\mathbb{R}^n)$. Then

$$P = \left(\sqcup_{i=1}^l B_i\right) \sqcup Q_0, \tag{2.6.60}$$

and noting that $\nu(P_0) \le \nu(A) \le 2\delta_0$ since $P_0 = A \cap P \subseteq A$, we have

$$\nu(Q_0) = \nu(P_0) + \sum_{i=1}^l \nu(C_i \setminus B_i) \le 2\delta_0 + l\frac{\delta_0}{l+1} < 3\delta_0. \tag{2.6.61}$$

Now we choose a box decomposition $(B_i)_{i=l+1}^p$ of Q_0 with $p \ge m$ (again $p = l$ corresponds to the case where $Q_0 = \emptyset$). By (2.6.60), we see that

$$P = \sqcup_{i=1}^p B_i, \tag{2.6.62}$$

i.e. $(B_i)_{i=1}^p$ is a box decomposition of P. Moreover, for $i = 1, \dots, l$, by $\overline{B_i} \subseteq C_i$ and (2.6.57), we have

$$\forall i = 1, \dots, l, \qquad \omega(f, \overline{B_i}) = W(f, \overline{B_i})\nu(\overline{B_i}) \le W(f, C_i)\nu(B_i) \le 2\varepsilon_0 \nu(B_i), \tag{2.6.63}$$

while by our choice of $M > 0$, $W(f, \mathbb{R}^n) \le 2M$, so

$$\forall i = l+1, \dots, p, \qquad \omega(f, \overline{B_i}) = W(f, \overline{B_i})\nu(\overline{B_i}) \le 2M\nu(B_i). \tag{2.6.64}$$

Hence

$$\begin{aligned} \sum_{i=1}^p \omega(f, \overline{B_i}) &= \sum_{i=1}^l \omega(f, \overline{B_i}) + \sum_{i=l+1}^p \omega(f, \overline{B_i}) \\ &\le 2\varepsilon_0 \sum_{i=1}^l \nu(B_i) + 2M \sum_{i=l+1}^p \nu(B_i) \\ &\le 2\varepsilon_0 \sum_{i=1}^l \nu(C_i) + 2M\nu(Q_0) \\ &= 2\varepsilon_0 \nu(P_1) + 2M\nu(Q_0) \\ &\le 2\varepsilon_0 \nu(P) + 6M\delta_0. \end{aligned} \tag{2.6.65}$$

Now it is clear that by choosing ε_0 and δ_0 suitably, e.g. let $\varepsilon_0 = \frac{\varepsilon}{4\nu(P)}$ and $\delta_0 = \frac{\varepsilon}{12M}$, we get (2.6.30), and (4) holds.

(4) implies (3). This is trivial since $\omega(f, B_i) \le \omega(f, \overline{B_i})$ for any box B_i.

(3) implies (1). By Lemma 2.6.15, we see that f is bounded. So there is a constant $M > 0$ such that $\|f(x)\| \le M$ for all $x \in \mathbb{R}^n$. Take any non-degenerate $A \in \mathcal{J}(\mathbb{R}^n)$. Since A is bounded, we may choose

an arbitrary $P \in \mathcal{P}(\mathbb{R}^n)$ with $P \supseteq A \supseteq \operatorname{supp}(f)$. Clearly P is nonempty since $\nu(P) \geq \nu(A) > 0$. By (3), there exists a box partition $\Delta' = (B_i)_{i=1}^p$ of P such that $\omega(f, \Delta) \leq \varepsilon/2$. Set $A_i' := B_i \cap A \in \mathcal{J}(\mathbb{R}^n)$. Clearly $A = \sqcup_{i=1}^p A_i'$, so

$$0 < \nu(A) = \sum_{i=1}^{p} \nu(A_i'). \tag{2.6.66}$$

Hence there is some non-degenerate A_i'. Relabeling if necessary, we may suppose $A_1', \ldots, A_{m-1}'$ are non-degenerate with $2 \leq m \leq p+1$, and $A_{m+1}', \ldots, A_p'$ are all Jordan negligible. Now since $A_1' \in \mathcal{J}(\mathbb{R}^n)$ is non-degenerate, so is $\operatorname{Int}(A_1')$. We may then choose a small non-empty box B so that $B \subseteq \overline{B} \subseteq \operatorname{Int}(A_1') \subseteq A_1'$, and $\nu(A_i' \setminus B) > 0$ and $\nu(B) \leq \varepsilon/(4M)$. Finally, set $A_1 := A_1' \setminus B$, $A_i = A_i'$ for $i = 2, \ldots, m-1$, and $A_m = (\sqcup_{i=m}^p A_i') \sqcup B$. Then clearly $\Delta := (A_i)_{i=1}^m$ is a non-degenerate Jordan partition of A.

Moreover, it is clear that $\omega(f, A_i') \leq \omega(f, B_i)$ for $i = 1, \ldots, m-1$ since $A_i' \subseteq B_i$. We then have

$$\sum_{i=1}^{m-1} \omega(f, A_i) \leq \sum_{i=1}^{m-1} \omega(f, B_i) \leq \omega(f, \Delta') \leq \frac{\varepsilon}{2}. \tag{2.6.67}$$

Since

$$\nu(A_m) = \nu(B) + \sum_{i=m}^{p} \nu(A_i') = \nu(B) \leq \frac{\varepsilon}{4M}, \tag{2.6.68}$$

we also have

$$\omega(f, A_m) = W(f, A_m)\nu(A_m) \leq 2M\frac{\varepsilon}{4M} = \frac{\varepsilon}{2}. \tag{2.6.69}$$

Taking the sum of (2.6.67)(2.6.69), we see that $\omega(f, \Delta) \leq \varepsilon$, so (1) holds. The proof of the whole theorem is now complete. □

Definition 2.6.21. — Let E be a Banach space. A map $f : \mathbb{R}^n \to E$ is called **Riemann integrable** if it is compactly supported and satisfies any of the equivalent conditions in Theorem 2.6.20.

Notation 2.6.22. — *We use $\mathcal{R}(\mathbb{R}^n, E)$ to denote the space of all Riemann integrable maps from $\mathbb{R}^n$ to the Banach space E. If $E = \mathbb{R}$, we abbreviate $\mathcal{R}(\mathbb{R}^n, \mathbb{R})$ simply by $\mathcal{R}(\mathbb{R}^n)$.*

Corollary 2.6.23. — *Let E be a Banach space. The space $\mathcal{R}(\mathbb{R}^n, E)$ is a linear subspace of the vector space of all maps from $\mathbb{R}^n$ to E.*

Proof. — This is an easy consequence of (6) of Theorem 2.6.20. The details are left as an exercise. □

Corollary 2.6.24. — *Let E, F, G be Banach spaces, and $\beta : E \times F \to G$ a continuous bilinear map. Then for any $f \in \mathcal{R}(\mathbb{R}^n, E)$ and $g \in \mathcal{R}(\mathbb{R}^n, F)$, we have $\beta(f, g) \in \mathcal{R}(\mathbb{R}^n, G)$, where $\beta(f, g) : \mathbb{R}^n \to G$ is the map sending $x \in \mathbb{R}^n$ to $\beta(f(x), g(x)) \in G$. In particular, if $f \in \mathcal{R}(\mathbb{R}^n)$ and $g \in \mathcal{R}(\mathbb{R}^n, F)$, then $fg \in \mathcal{R}(\mathbb{R}^n, G)$.*

Proof. — Again, this is an easy consequence of (6) of Theorem 2.6.20, and left as an exercise. □

2.7. Riemann Integrals

We still fix a dimension $n \in \mathbb{N}^+$ throughout this section unless otherwise stated.

We are now well-prepared to treat Riemann integrals. Let E be a Banach space and $f \in \mathcal{R}(\mathbb{R}^n, E)$. Consider any $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$ and any Jordan partition Δ of A. Recall the notion of Riemann sums (Definition 2.6.9) and Notation 2.6.10. Remark 2.6.12 strongly indicates that the Riemann sums of f with respect to Δ should be close to the Riemann integral of f, if the latter can be defined, as long as $\omega(f, \Delta)$ is small enough. Now observe that $f \in \mathcal{R}(\mathbb{R}^n, E)$ implies in particular that $\omega(f, \Delta)$ can be made arbitrarily small. We thus adopt the approach of defining the Riemann integral of f via taking a suitable limit of the corresponding Riemann sums. It is precisely this limiting process that demands the completeness of E.

We now describe this limiting process, which is slightly involved. To facilitate our discussion and motivated by the consideration in Remark 2.6.12, we give the following definition.

Definition 2.7.1. — Let E be a Banach space and $f \in \mathcal{R}(\mathbb{R}^n, E)$. For $A \in \mathcal{J}(\mathbb{R}^n)$, by a **limiting sequence** of Jordan partitions of A with respect to f, we mean a sequence $(\Delta_k)_{k\geq 1}$, where each Δ_k is a Jordan partition of A, such that

$$\lim_{k\to\infty} \omega(f, \Delta_k) = 0. \tag{2.7.1}$$

Fix for the moment an $f \in \mathcal{R}(\mathbb{R}^n, E)$ (E being an arbitrary Banach space). Now take any $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$. By Theorem 2.6.20, we may find a limiting sequence (Δ_k) of partitions of A with respect to f. For each k, take $s_k \in S(f, \Delta_k)$. We wish to define the **Riemann integral** $\int f$ as the limit of the sequence $(s_k)_{k\geq 1}$ in E. In order for this to be well-defined, we need to establish the following:

(1) the sequence (s_k) is Cauchy (so it has a limit in E);
(2) for any $B \in \mathcal{J}(\mathbb{R}^n)$ with $B \supseteq \operatorname{supp}(f)$, any limiting sequence of Jordan partitions $(\widetilde{\Delta}_k)$ of B with respect to f, and any $\widetilde{s}_k \in S(f, \widetilde{\Delta}_k)$ for each k, the two Cauchy sequence $(\widetilde{s}_k)$ and (s_k) have the same limit.

Lemma 2.7.2. — *For any $A \in \mathcal{J}(\mathbb{R}^n)$, suppose Δ_1 and Δ_2 are two Jordan partitions of A, then there is a Jordan partition Δ of A that is a common refinement of Δ_1 and Δ_2.*

Proof. — Let $\Delta_1 = (A_i)_{i=1}^l$ and $\Delta_2 = (B_j)_{j=1}^m$, just take $\Delta = (A_i \cap B_j)_{1\leq i\leq l,\, 1\leq j\leq m}$. □

Lemma 2.7.3. — *Let E be a Banach space, $g : \mathbb{R}^n \to E$ a* bounded *map. If Δ_1 and Δ_2 are two Jordan partitions of the same $A \in \mathcal{J}(\mathbb{R}^n)$ such that Δ_2 is finer than Δ_1, then for any $\sigma_1 \in S(g, \Delta_1)$ and $\sigma_2 \in S(g, \Delta_2)$, we have*

$$\|\sigma_1 - \sigma_2\| \leq \omega(g, \Delta_1). \tag{2.7.2}$$

Proof. — Let $\Delta_1 = (A_i)_{i=1}^m$, $\Delta_2 = (B_j)_{j=1}^l$. It suffices to treat the case where all A_i and B_j are non-empty. Since Δ_2 is finer than Δ_1, we have

$$\forall i \in \{1, \ldots, m\}, \qquad A_i = \bigsqcup_{K_j \subseteq J_i} B_j. \tag{2.7.3}$$

By definition, we have $a_i \in A_i$ and $b_j \in B_j$ for all i and j, such that

$$\sigma_1 = \sum_{i=1}^m g(a_i)\nu(A_i) \qquad \text{and} \qquad \sigma_2 = \sum_{j=1}^l g(b_j)\nu(B_j). \tag{2.7.4}$$

Using the decomposition (2.7.3), we conclude that

$$\sigma_1 - \sigma_2 = \sum_{i=1}^m \sum_{K_j\subseteq A_i} \big[g(a_i) - g(b_j)\big]\nu(B_j), \tag{2.7.5}$$

hence

$$\|\sigma_1 - \sigma_2\| \leq \sum_{i=1}^m \sum_{K_j\subseteq A_i} \|g(a_i) - g(b_j)\|\nu(B_j) \leq \sum_{i=1}^m \sum_{K_j\subseteq A_i} W(g, A_i)\nu(B_j) = \omega(g, \Delta_1), \tag{2.7.6}$$

where the last equality follows again from the decomposition (2.7.3). □

Remark 2.7.4. — Lemma 2.7.3 can be viewed as a variant of Proposition 2.6.11.

Corollary 2.7.5. — *Let E be a Banach space, $g : \mathbb{R}^n \to E$ a* bounded *map. If Δ_1 and Δ_2 are Jordan partitions of the same $A \in \mathcal{J}(\mathbb{R}^n)$, then for any $\sigma_1 \in S(g, \Delta_1)$ and $\sigma_2 \in S(g, \Delta_2)$, we have*

$$\|\sigma_1 - \sigma_2\| \leq \omega(g, \Delta_1) + \omega(g, \Delta_2). \tag{2.7.7}$$

Proof. — By Lemma 2.7.2, we may take a Jordan partition Δ of A that refines both Δ_1 and Δ_2. Now take any $\sigma \in S(g, \Delta)$, then by Lemma 2.7.3, we have

$$\|\sigma_1 - \sigma_2\| \leq \|\sigma_1 - \sigma\| + \|\sigma_2 - \sigma\| \leq \omega(g, \Delta_1) + \omega(g, \Delta_2). \tag{2.7.8}$$

□

Remark 2.7.6. — The estimation in Corollary 2.7.5 is not as fine as the estimation in Lemma 2.7.3, but has the advantage of being applicable for any two Jordan partitions Δ_1 and Δ_2 of the same Jordan set, even when Δ_1 and Δ_2 are not comparable.

Now back to our plan of defining $\int f$ via taking limit of Riemann sums.

Proposition 2.7.7. — *Let E be a Banach space, $f \in \mathcal{R}(\mathbb{R}^n, E)$. The following hold.*

(1) *There is $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$.*
(2) *For each $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$, there exists a limiting sequence (Δ_k) of Jordan partitions of A with respect to f.*
(3) *For any $A \in \mathcal{J}(\mathbb{R}^n)$ with $A \supseteq \operatorname{supp}(f)$, any limiting sequence (Δ_k) of Jordan partitions of A with respect to f and $s_k \in S(f, \Delta_k)$ for each k, the sequence (s_k) in E is Cauchy.*
(4) *Let $A, \widetilde{A} \in \mathcal{J}(\mathbb{R}^n)$ with $A \cap \widetilde{A} \supseteq \operatorname{supp}(f)$. Suppose (Δ_k) (resp. $\widetilde{\Delta}_k$) is a limiting sequence of Jordan partitions of A (resp. $\widetilde{A}$) with respect to f. For each k, choose any $s_k \in S(f, \Delta_k)$ (resp. $\widetilde{s}_k \in S(f, \widetilde{\Delta}_k)$), the Cauchy sequences (s_k) and $(\widetilde{s}_k)$ have the same limit.*

Proof. — (1) follows from the compactness of $\operatorname{supp}(f)$ and (2) from Theorem 2.6.20.

(3). Take any $\varepsilon > 0$. By Definition 2.7.1, there exists $N \in \mathbb{N}$, such that

$$m \geq N \implies \omega(f, \Delta_k) \leq \frac{\varepsilon}{2}. \tag{2.7.9}$$

Hence by Corollary 2.7.5, we have

$$\forall m, l \geq N, \qquad \|s_m - s_l\| \leq \omega(f, \Delta_m) + \omega(f, \Delta_l) \leq \varepsilon, \tag{2.7.10}$$

which implies that (s_k) is indeed Cauchy.

(4). Define $A' := A \cup \widetilde{A} \in \mathcal{J}(\mathbb{R}^n)$. Then $A' \supseteq \operatorname{supp}(f)$. For each k, define $\Delta'_{2k} = \Delta_k \cup \{A' \setminus A\}$ and $\Delta'_{2k-1} = \widetilde{\Delta}_k \cup \left\{A' \setminus \widetilde{A}\right\}$; correspondingly, set $s'_{2k} = s_k$ and $s'_{2k-1} = \widetilde{s}_k$. Since $A \cap \widetilde{A} \supseteq \operatorname{supp}(f)$, we have $f(A' \setminus \widetilde{A}) = f(A' \setminus A) = \{0\}$. It follows that $\omega(f, A' \setminus A) = \omega(f, A' \setminus \widetilde{A}) = 0$, so $\omega(f, \Delta'_{2k}) = \omega(f, \Delta_k)$ and $\omega(f, \Delta'_{2k-1}) = \omega(f, \widetilde{\Delta}_k)$, which implies that (Δ'_k) is a limiting sequence of Jordan partitions of A'; moreover, from our construction $s'_k \in S(f, \Delta'_k)$ for each k. Now being sub-sequences of the Cauchy sequence (s'_k), the Cauchy sequence (s_k) and $(\widetilde{s}_k)$ have the same limit. □

Definition 2.7.8. — Let E be a Banach space, $f \in \mathcal{R}(\mathbb{R}^n, E)$. The **Riemann integral** of f, denoted by $\int f$, $\int_{\mathbb{R}^n} f$ or $\int_{\mathbb{R}^n} f(x)\mathrm{d}x$, is the common limit in part (4) of Proposition 2.7.7.

Closely related to this notion is the Riemann integral over a Jordan set. For $A \in \mathcal{J}(\mathbb{R}^n)$, we have $\chi_A \in \mathcal{R}(\mathbb{R}^n)$ since the set $D(\chi_A)$ of discontinuous points of χ_A is precisely ∂A, which is Jordan negligible, hence negligible (χ_A is clearly bounded). Thus if $f \in \mathcal{R}(\mathbb{R}^n, E)$, then $\chi_A f \in \mathcal{R}(\mathbb{R}^n, E)$.

Definition 2.7.9. — Using the above notation, we define $\int_A f$ to be the integral $\int \chi_A f$, and call it the **Riemann integral** of f **over** A.

***Proposition 2.7.10* (Basic properties of Riemann integrals).** — *Let E be a Banach space. The following hold.*

(1) *If $A \in \mathcal{J}(\mathbb{R}^n)$ is Jordan negligible, then we have $\int_A f = 0$, for any $f \in \mathcal{R}(\mathbb{R}^n, E)$.*
(2) *We have $\int_A f = \int_{\mathbb{R}^n} f$ whenever $f \in \mathcal{R}(\mathbb{R}^n, E)$, $A \in \mathcal{J}(\mathbb{R}^n)$ and $\overline{A} \supseteq \operatorname{supp}(f)$.*
(3) *For any $f \in \mathcal{R}(\mathbb{R}^n, E)$, any $A \in \mathcal{J}(\mathbb{R}^n)$ with $\overline{A} \supseteq \operatorname{supp}(f)$ and any Jordan partition Δ of A, we have*

$$\forall s \in S(f, \Delta), \qquad \left\| s - \int f \right\| \leq \omega(f, \Delta). \tag{2.7.11}$$

(4) *The map*

$$\begin{aligned} \int : \mathcal{R}(\mathbb{R}^n, E) &\to E \\ f &\mapsto \int f \end{aligned} \tag{2.7.12}$$

is linear.

(5) *For any $f \in \mathcal{R}(\mathbb{R}^n, E)$, we have $\|f\| \in \mathcal{R}(\mathbb{R}^n)$ where $\|f\| : \mathbb{R}^n \to \mathbb{R}$ is the norm map sending $x \in \mathbb{R}^n$ to $\|f(x)\|$, and*

$$\left\| \int f \right\| \leq \int \|f\|. \tag{2.7.13}$$

(6) *If $A, B \in \mathcal{J}(\mathbb{R}^n)$ with $A \cap B$ being Jordan negligible, then for any $f \in \mathcal{R}(\mathbb{R}^n, E)$, we have*

$$\int_{A \cup B} f = \int_A f + \int_B f. \tag{2.7.14}$$

Proof. — (1). Replacing f by $\chi_A f$ and A by $\overline{A}$ if necessary, we may assume without loss of generality that $\operatorname{supp}(f) \subseteq A$. Since $\nu(A) = 0$ and $f \in \mathcal{R}(\mathbb{R}^n, E)$ (in particular, f is bounded), we have $B \in \mathcal{J}(\mathbb{R}^n)$ and $\nu(B) = 0$ for any $B \subseteq A$. In particular, $S(f, \Delta) = \{0\}$ for any Jordan partition Δ of A. Consequently, $\int f = 0$.

(2). Let (Δ_k) be a limiting sequence of Jordan partitions of A with respect to $\chi_A f$. For each k, set $\widetilde{\Delta}_k = \Delta_k \cup \{\overline{A} \setminus A\}$, which is a Jordan partition of $\overline{A}$. Since $\overline{A} \setminus A \subseteq \partial A$, the set $\overline{A} \setminus A$ is Jordan negligible. Note that $(\chi_A f) \restriction_A = f \restriction_A$, we have

$$\omega(\chi_A f, \Delta_k) = \omega(f, \Delta_k) = \omega(f, \widetilde{\Delta}_k) \tag{2.7.15}$$

where the last equality follows from $\nu(\overline{A} \setminus A) = 0$. Hence $(\widetilde{\Delta}_k)$ is a limiting sequence of Jordan partitions of $\overline{A}$ with respect to f. Choose a $s_k \in S(\chi_A f, \Delta_k) = S(f, \Delta_k)$ for each k. For any fixed k, since $\nu(\overline{A} \setminus A) = 0$, we also have $s_k \in S(f, \widetilde{\Delta}_k)$, and

$$\int_A f = \int \chi_A f = \lim_{k \to \infty} s_k = \int f. \tag{2.7.16}$$

(3). Since $\overline{A} \setminus A$ is Jordan negligible, we have

$$s \in S(f, \Delta) \subseteq S(f, \widetilde{\Delta}), \qquad \text{where} \qquad \widetilde{\Delta} = \Delta \cup \{\overline{A} \setminus A\}. \tag{2.7.17}$$

Let (Δ_k) be an arbitrary limiting sequence of Jordan partitions of $\overline{A}$. For each k, take a Jordan partition $\widetilde{\Delta}_k$ of $\overline{A}$ that refines both $\widetilde{\Delta}$ and Δ_k, then by Proposition 2.6.11, we have

$$\omega(f, \widetilde{\Delta}_k) \leq \omega(f, \Delta_k) \tag{2.7.18}$$

so $(\widetilde{\Delta}_k)$ is still a limiting sequence of Jordan partitions of $\overline{A}$ with respect to f. Take any $\widetilde{s}_k \in S(f, \widetilde{\Delta}_k)$, we have, by $\widetilde{\Delta}_k$ is finer than $\widetilde{\Delta}$ and Lemma 2.7.3, that

$$\forall k, \qquad \|s - s_k\| \leq \omega(f, \widetilde{\Delta}) = \omega(f, \Delta). \tag{2.7.19}$$

Since $\int f = \lim_{k \to \infty} s_k$, (2.7.11) follows from (2.7.19) by taking limit along $k \to \infty$.

(4). Take any $f, g \in \mathcal{R}(\mathbb{R}^n, E)$ and any $\alpha, \beta \in \mathbb{R}$. Fix an arbitrary $\varepsilon > 0$. Choose a box B with $B \supseteq \operatorname{supp}(f) \cup \operatorname{supp}(g)$. Taking refinement if necessary, we may find a Jordan partition $\Delta = (B_i)_{i=1}^m$ of B with each $B_i \neq \emptyset$, such that

$$\omega(f, \Delta) + \omega(g, \Delta) \leq \varepsilon. \tag{2.7.20}$$

Choose $x_i \in B_i$ for each i. Set

$$s_f := \sum_{i=1}^m f(x_i) \nu(B_i) \in S(f, \Delta), \tag{2.7.21}$$

and

$$s_g := \sum_{i=1}^m g(x_i) \nu(B_i) \in S(g, \Delta). \tag{2.7.22}$$

Then (2.7.20) implies that $\omega(f, \Delta) \leq \varepsilon$, so

$$\left\| s_f - \int f \right\| \leq \varepsilon \tag{2.7.23}$$

by (3). Similarly,

$$\left\| s_g - \int g \right\| \leq \varepsilon. \tag{2.7.24}$$

Moreover, by definition, it is easy to see that

$$W(\alpha f + \beta g, C) \leq |\alpha| W(f, C) + |\beta| W(g, C), \tag{2.7.25}$$

so

$$\omega(\alpha f+\beta g,\Delta)\le|\alpha|\omega(f,\Delta)+|\beta|\omega(g,\Delta)\le(|\alpha|+|\beta|)\varepsilon. \tag{2.7.26}$$

But

$$\alpha s_f+\beta s_g=\sum_{i=1}^m(\alpha f+\beta g)(x_i)\nu(B_i)\in S(\alpha f+\beta g,\Delta). \tag{2.7.27}$$

By Lemma 2.7.3 and (2.7.26), we have

$$\left\|\alpha s_f+\beta s_g-\int(\alpha f+\beta g)\right\|\le(|\alpha|+|\beta|)\varepsilon. \tag{2.7.28}$$

It is easy to see from (2.7.28) as well as (2.7.23) and (2.7.24) that

$$\left\|\alpha\int f+\beta\int g-\int(\alpha f+\beta g)\right\|\le 2(|\alpha|+|\beta|)\varepsilon. \tag{2.7.29}$$

Since $\varepsilon>0$ is arbitrary, we must have

$$\int(\alpha f+\beta g)=\alpha\int f+\beta\int g, \tag{2.7.30}$$

and $\int:\mathcal{R}(\mathbb{R}^n,E)\to E$ is indeed linear.

(5). Since the norm function $\|\cdot\|:E\to\mathbb{R}$ is continuous, it follows that $D(\|f\|)\subseteq D(f)$, so $D(\|f\|)$ is negligible because $D(f)$ is so. Clearly $\|f\|$ is bounded since f is. It now follows from Theorem 2.6.20 that $\|f\|\in\mathcal{R}(\mathbb{R}^n)$. Now take any $\varepsilon>0$ and any Jordan partition Δ of some $A\in\mathcal{J}(\mathbb{R}^n)$ with $A\supseteq\operatorname{supp}(f)$, such that $\omega(f,\Delta)\le\varepsilon$. Without loss of generality, we may assume each member in Δ is non-empty. Choose any $s\in S(f,\Delta)$. For each $a,b\in\mathbb{R}^n$, we have

$$\max(\|f(a)\|-\|f(b)\|,\|f(b)\|-\|f(a)\|)\le\|f(a)-f(b)\|. \tag{2.7.31}$$

It follows from (2.7.31) that

$$\forall C\subseteq\mathbb{R}^n,\qquad W(\|f\|,C)\le W(f,C). \tag{2.7.32}$$

Hence

$$\omega(\|f\|,\Delta)=\sum_{i=1}^m W(\|f\|,A_i)\nu(A_i)\le\sum_{i=1}^m W(f,A_i)\nu(A_i)=\omega(f,\Delta)\le\varepsilon. \tag{2.7.33}$$

Let $s=\sum_{i=1}^m f(a_i)\nu(A_i)$ with each $a_i\in A_i$. Then the number $r:=\sum_{i=1}^m\|f(a_i)\|\nu(A_i)\in S(\|f\|,\Delta)$. Clearly,

$$\|s\|\le r. \tag{2.7.34}$$

On the other hand, by (3), we see that

$$\left\|s-\int f\right\|\le\omega(f,\Delta)\le\varepsilon, \tag{2.7.35}$$

and (applied to the case $E=\mathbb{R}$)

$$\left|r-\int\|f\|\right|\le\omega(\|f\|,\Delta)\le\omega(f,\Delta)\le\varepsilon. \tag{2.7.36}$$

From the above inequalities, we have

$$\left\|\int f\right\|\le\|s\|+\left\|s-\int f\right\|\le r+\varepsilon\le 2\varepsilon+\int\|f\|, \tag{2.7.37}$$

which implies (2.7.13), since $\varepsilon>0$ is arbitrary.

(6). This is a consequence of $\chi_{A\cup B}=\chi_A+\chi_B-\chi_{A\cap B}$. More precisely, from (1) and (4), we have

$$\begin{aligned}\int_{A\cup B}f&=\int\chi_{A\cup B}f=\int(\chi_A f+\chi_B f)-\int\chi_{A\cap B}f\\&=\int_A f+\int_B f-\int_{A\cap B}f=\int_A f+\int_B f.\end{aligned} \tag{2.7.38}$$

The proof of the whole proposition is now complete. $\square$

Sometimes it is convenient to be able to integrate maps that are only defined on a Jordan set. This can be reduced to the case that we have treated.

Notation 2.7.11. — *Let E be a Banach space, $A \in \mathcal{J}(\mathbb{R}^n)$, and $f : A \to E$ be any map. We may extend f to a map $\widetilde{f}$ from $\mathbb{R}^n$ to E, by setting $\widetilde{f}(\mathbb{R}^n \setminus A) = \{0\}$. We also set*

$$\widetilde{\mathcal{R}}(A, E) := \Big\{ f : A \to E \;\Big|\; \widetilde{f} \in \mathcal{R}(\mathbb{R}^n, E) \Big\}. \tag{2.7.39}$$

Definition 2.7.12. — Using Notation 2.7.11, elements of $\widetilde{R}(A, E)$ are also called **Riemann integrable**, i.e. a map $f : A \to E$ is Riemann integrable if and only if $\widetilde{f} \in \mathcal{R}(\mathbb{R}^n, E)$. In this case, we call $\int_{\mathbb{R}^n} \widetilde{f}$ the **Riemann integral** of f, and denote it by $\int_A f$, $\int_A f(x)\mathrm{d}x$ or simply $\int f$ when A is clear from context.

Lemma 2.7.13. — *Use Notation 2.7.11, we have $D(f) \cup \partial A = D(\widetilde{f}) \cup \partial A$, where $D(f)$ (resp. $D(\widetilde{f})$) is the set of discontinuous points of f in A (resp. $\mathbb{R}^n$).*

Proof. — Since $\widetilde{f} \restriction_{\mathbb{R}^n \setminus \overline{A}} \equiv 0$, we have $D(\widetilde{f}) \subseteq \overline{A}$. On the other hand, $\widetilde{f} \restriction_{\mathrm{Int}(A)} = f \restriction_{\mathrm{Int}(A)}$, so $D(f) \cap \mathrm{Int}(A) = D(\widetilde{f}) \cap \mathrm{Int}(A)$. The lemma now follows since $\partial A = \overline{A} \setminus \mathrm{Int}(A)$. $\square$

Proposition 2.7.14. — *Using Notation 2.7.11, f is Riemann integrable if and only if f is bounded and $D(f)$ is negligible.*

Proof. — Clearly f is bounded if and only if $\widetilde{f}$ is so. Now the result follows from Lemma 2.7.13 and Theorem 2.6.20. $\square$

Remark 2.7.15. — One may actually mimic Theorem 2.6.20 and establishes a corresponding version of that theorem for maps defined on a fixed Jordan set $A \subseteq \mathbb{R}^n$. Then one can proceed to define Riemann sums and Riemann integrals as the above $A = \mathbb{R}^n$ case. We will *not* adopt this approach, as Proposition 2.7.14 already suffices for the purpose of developing the Riemann integration theory for maps defined on A.

Proposition 2.7.16. — *Let E be a Banach space and $A \in \mathcal{J}(\mathbb{R}^n)$. The following hold.*

(1) *(Linearity of the Riemann integral) The space $\widetilde{\mathcal{R}}(A, E)$ is a linear subspace of the vector space of all maps from A to E, and for any $f, g \in \widetilde{\mathcal{R}}(A, E)$ and $\alpha, \beta \in \mathbb{R}$, we have*

$$\int_A (\alpha f + \beta g) = \alpha \int_A f + \beta \int_A g. \tag{2.7.40}$$

(2) *For any Riemann integrable map f from A to E, we have*

$$\left\| \int_A f \right\| \le \int_A \|f\|. \tag{2.7.41}$$

(3) *For any Riemann integrable map f from A to E, and for each $B \in \mathcal{J}(\mathbb{R}^n)$ with $B \subseteq A$, the map $f \restriction_B : B \to E$ is Riemann integrable. Moreover, using $\int_B f$ to denote $\int_B f \restriction_B$ for any such B, we have*

$$\forall B_1, B_2 \in \mathcal{J}(\mathbb{R}^n), \qquad B_1 \cup B_2 \subseteq A \text{ and } \nu(B_1 \cap B_2) = 0 \implies \int_{B_1 \cup B_2} f = \int_{B_1} f + \int_{B_2} f. \tag{2.7.42}$$

Proof. — The proof can be seen as a consequence of Proposition 2.7.10. Using Notation 2.7.11, for (1), that $\widetilde{R}(A, E)$ forms a linear subspace follows from Lemma 2.7.13, and we have

$$\begin{aligned} \int_A (\alpha f + \beta g) &= \int_{\mathbb{R}^n} \widetilde{\alpha f + \beta g} \\ &= \int_{\mathbb{R}^n} (\alpha \widetilde{f} + \beta \widetilde{g}) \\ &= \alpha \int_{\mathbb{R}^n} \widetilde{f} + \beta \int_{\mathbb{R}^n} \widetilde{g} \\ &= \alpha \int_A f + \beta \int_A g. \end{aligned} \tag{2.7.43}$$

For (2), we have

$$\left\|\int_A f\right\| = \left\|\int_{\mathbb{R}^n} \widetilde{f}\right\| \leq \int_{\mathbb{R}^n} \left\|\widetilde{f}\right\| = \int_{\mathbb{R}^n} \widetilde{\|f\|} = \int_A \|f\|. \tag{2.7.44}$$

Finally, we prove (3). In this case, the Riemann integrability of $f\restriction_B$ follows from its obvious boundedness and the fact that $D(f\restriction_B) \subseteq D(f)$. In this case, it is easy to see that $\int_B f = \int_A \chi_B f = \int_{\mathbb{R}^n} \chi_B \widetilde{f}$.

For (2.7.42), we have

$$\begin{aligned}\int_{B_1\cup B_2} f &= \int_{\mathbb{R}^n} \chi_{B_1\cup \chi_{B_2}} \widetilde{f} = \int_{\mathbb{R}^n} \chi_{B_1}\widetilde{f} + \int_{\mathbb{R}^n} \chi_{B_2}\widetilde{f} - \int_{\mathbb{R}^n} \chi_{B_1\cap B_2}\widetilde{f}\\ &= \int_{B_1} \widetilde{f} + \int_{B_2} \widetilde{f} - \int_{B_1\cap B_2} \widetilde{f} = \int_{B_1} f + \int_{B_2} f,\end{aligned} \tag{2.7.45}$$

where the last equality follows since $\int_{B_1\cap B_2} \widetilde{f} = 0$ by Proposition 2.7.10 and $\nu(B_1 \cap B_2) = 0$. □

It is often convenient to have a parallel theory with $\mathbb{R}^n$ replaced by an open set U in $\mathbb{R}^n$.

Notation 2.7.17. — *For a Banach space E, the symbol $\mathcal{R}(U, E)$ denotes the subspace of $\widetilde{\mathcal{R}}(U, E)$ consisting of all Riemann integrable maps with support in U, i.e. (see Notation 2.7.11)*

$$\mathcal{R}(U, E) := \left\{ f \in \widetilde{\mathcal{R}}(U, E) \;\middle|\; \operatorname{supp}(f) \subseteq U \right\}. \tag{2.7.46}$$

When $E = \mathbb{R}$, we often simply write $\mathcal{R}(U)$ instead of $\mathcal{R}(U, \mathbb{R})$.

Proposition 2.7.18. — *Using Notation 2.7.17, $\mathcal{R}(U, E)$ is a linear subspace of $\widetilde{\mathcal{R}}(U, E)$.*

Proof. — It is clear that for any $f, g : \mathbb{R}^n \to E$ and any $\alpha, \beta \in \mathbb{R}$, we have

$$\operatorname{supp}(\alpha f + \beta g) \subseteq \operatorname{supp}(f) + \operatorname{supp}(g). \tag{2.7.47}$$

Clearly, this implies that if $\operatorname{supp}(f), \operatorname{supp}(g) \subseteq U$, then so is $\operatorname{supp}(\alpha f + \beta g)$, which means $\mathcal{R}(U, E)$ is a linear subspace of $\mathcal{R}(\mathbb{R}^n, E)$. □

For an open $U \subseteq \mathbb{R}^n$, the following result says that $\mathcal{R}(U, E)$ can be identified with the space of all maps from U to E that are bounded, compactly supported with support in U, and whose set of discontinuous points is negligible.

Proposition 2.7.19. — *Using the above notation. The following hold.*

(1) *The space*

$$\mathcal{R}_U(\mathbb{R}^n, E) := \{f \in \mathcal{R}(\mathbb{R}^n, E) \mid \operatorname{supp}(f) \subseteq U\} \tag{2.7.48}$$

is a linear subspace of $\mathcal{R}(\mathbb{R}^n, E)$.

(2) *The map*

$$\begin{aligned}\widetilde{(\cdot)} : \mathcal{R}(U, E) &\to \mathcal{R}_U(\mathbb{R}^n, E)\\ f &\mapsto \widetilde{f}\end{aligned} \tag{2.7.49}$$

is bijective, where $\widetilde{f} : \mathbb{R}^n \to E$ is the extension of f with $\widetilde{f}(x) = 0$ for all $x \in \mathbb{R}^n \setminus U$; and the inverse of $\widetilde{(\cdot)}$ is given by

$$\begin{aligned}(\cdot)\restriction_U : \mathcal{R}_U(\mathbb{R}^n, E) &\to \mathcal{R}(U, E)\\ g &\mapsto g\restriction_U .\end{aligned} \tag{2.7.50}$$

Proof. — By now, the proof should be a routine verification by carefully unwinding the definitions and is left as an exercise. □

Remark 2.7.20. — From now on, we will often make the identification described in Proposition 2.7.19, i.e. we will often identify $f \in \mathcal{R}(U, E)$ with $\widetilde{f} \in \mathcal{R}_U(\mathbb{R}^n, E)$ and vice versa.

2.8. Successive Integrals

Fix three dimensions n, k and l in $\mathbb{N}^+$ with $n = k + l$, so that we have the canonical identification of $\mathbb{R}^n$ with $\mathbb{R}^k \times \mathbb{R}^l$. The goal of this section is to study the possibility of representing a Riemann integral over $\mathbb{R}^n$ as a successive integral along the decomposition $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$.

Note that for the special case $E = \mathbb{R}$, we have seen in the homework exercise that

Theorem 2.8.1 **(Fubini's Theorem for Real Valued Riemann Integrals)**

For each $f \in \mathcal{R}(\mathbb{R}^n)$, which we view as a function of variables $(x_1, \ldots, x_n)$, and $\sigma \in S_n$. For each $k \in \{1, \ldots, n\}$, let $\overline{\mathcal{I}}_k$ (resp. $\underline{\mathcal{I}}_k$) denote the operation of taking integrals of the variable x_k over $\mathbb{R}$. Then

$$\int_{\mathbb{R}^n} f = (\mathcal{J}_{\sigma(1)} \circ \cdots \circ \mathcal{J}_{\sigma(n)})(f) \tag{2.8.1}$$

where each $\mathcal{J}_k$ is either $\overline{\mathcal{I}}_k$ or $\underline{\mathcal{I}}_k$. In particular,

$$\begin{aligned}
\int_{\mathbb{R}^n} f &= \overline{\int}_{\mathbb{R}^k} \left(\overline{\int}_{\mathbb{R}^l} f(x,y)\mathrm{d}x \right) \mathrm{d}y \\
&= \overline{\int}_{\mathbb{R}^k} \left(\underline{\int}_{\mathbb{R}^l} f(x,y)\mathrm{d}x \right) \mathrm{d}y \\
&= \underline{\int}_{\mathbb{R}^k} \left(\overline{\int}_{\mathbb{R}^l} f(x,y)\mathrm{d}x \right) \mathrm{d}y \\
&= \underline{\int}_{\mathbb{R}^k} \left(\underline{\int}_{\mathbb{R}^l} f(x,y)\mathrm{d}x \right) \mathrm{d}y,
\end{aligned} \tag{2.8.2}$$

where $\overline{\int}$ and $\underline{\int}$ are respectively the upper and lower Riemann integrals via Darboux upper/lower sums as introduced in the homework exercise. □

Suppose E is a general Banach space, for any $f \in \mathcal{R}(\mathbb{R}^n, E)$, integrating out the first (resp. second) variable with respect to the decomposition $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$, we obtain a map from $\mathbb{R}^l$ (resp. $\mathbb{R}^k$) to E, denoted respectively by

$$\begin{aligned}
\int_{\mathbb{R}^k} f(x,\cdot)\mathrm{d}x : \mathbb{R}^l &\to E \\
y &\mapsto \int_{\mathbb{R}^k} f(x,y)\mathrm{d}x,
\end{aligned} \tag{2.8.3}$$

and

$$\begin{aligned}
\int_{\mathbb{R}^l} f(\cdot,y)\mathrm{d}y : \mathbb{R}^k &\to E \\
x &\mapsto \int_{\mathbb{R}^l} f(x,y)\mathrm{d}y.
\end{aligned} \tag{2.8.4}$$

Definition 2.8.2. — Using the above notation, we say f is **successively contented with respect to the decomposition** $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$, if it is in $\mathcal{R}(\mathbb{R}^n, E)$ and the following conditions hold:

(1) $f(\cdot, y) : \mathbb{R}^k \to E$ is in $\mathcal{R}(\mathbb{R}^k, E)$ for all $y \in \mathbb{R}^l$;
(2) $f(x, \cdot) : \mathbb{R}^l \to E$ is in $\mathcal{R}(\mathbb{R}^l, E)$ for all $x \in \mathbb{R}^k$;
(3) the map (2.8.3) is in $\mathcal{R}(\mathbb{R}^l, E)$;
(4) the map (2.8.4) is in $\mathcal{R}(\mathbb{R}^k, E)$.

Remark 2.8.3. — We remark that not all Riemann integrable maps are successively contented. As a counter-example, we treat the case $E = \mathbb{R}$ (counter-examples for general E can then be easily obtained since E contains a copy of $\mathbb{R}$, unless $E = 0$ in which case everything is trivial). Consider the case of $A = \square_0^1 \cap \mathbb{Q}^k \subseteq \mathbb{R}^k$, $B = \{0\} \subseteq \mathbb{R}^l$ and $C = A \times B \subseteq \mathbb{R}^n$. It is easy to see that C is Jordan negligible in $\mathbb{R}^n$, so $\chi_C \in \mathcal{R}(\mathbb{R}^n)$. However, $\chi_C(\cdot, 0) = \chi_A \notin \mathcal{R}(\mathbb{R}^k)$.

Our next goal is to establish Theorem 2.9.7, for which we need some preparation.

Lemma 2.8.4. — *The space of all successively contented with respect to the decomposition $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$ form a linear subspace of $\mathcal{R}(\mathbb{R}^n, E)$.*

Proof. — This follows from the fact that inside the vector space $\mathcal{R}(\mathbb{R}^n, E)$, each of the conditions in Definition 2.8.2 is stable under taking linear combinations. □

Lemma 2.8.5. — *Every paved map $g : \mathbb{R}^n \to E$ is successively contented with respect to $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$.*

Proof. — By Lemma 2.8.4, it suffices to show the special case $g = c\chi_B$ where B is a box in $\mathbb{R}^n$. Clearly, $B = B_1 \times B_2$ where B_1 (resp. B_2) is a box in $\mathbb{R}^k$ (resp. $\mathbb{R}^l$). Then for each $x \in \mathbb{R}^k$ (resp. $y \in \mathbb{R}^l$), we have

$$g(x, \cdot) = c\chi_{B_1}(x)\chi_{B_2} \in \mathcal{R}(\mathbb{R}^l, E) \qquad \text{and} \qquad g(\cdot, y) = c\chi_{B_2}(y)\chi_{B_1} \in \mathcal{R}(\mathbb{R}^k, E). \tag{2.8.5}$$

This implies that

$$\int_{\mathbb{R}^l} g(\cdot, y)\mathrm{d}y = c\nu_l(B_2)\chi_{B_1} \in \mathcal{R}(\mathbb{R}^k, E) \qquad \text{and} \qquad \int_{\mathbb{R}^k} g(x, \cdot)\mathrm{d}y = c\nu_k(B_1)\chi_{B_2} \in \mathcal{R}(\mathbb{R}^l, E). \tag{2.8.6}$$

Hence g is indeed successively contented with respect to $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$. □

Lemma 2.8.6. — *Let E be a Banach space, $f \in \mathcal{R}(\mathbb{R}^n, E)$, then for each $\varepsilon > 0$, there exists a paved map $g : \mathbb{R}^n \to E$, such that $\|g\|_\infty \le \|f\|_\infty$ and*

$$\int_{\mathbb{R}^n} \|f - g\| \le \varepsilon. \tag{2.8.7}$$

Proof. — Take any non-empty $P \in \mathcal{P}(\mathbb{R}^n)$ with $P \supseteq \operatorname{supp}(f)$. By Theorem 2.8.7, there exists a box partition $\Delta = (B_i)_{i=1}^m$ of P with

$$\omega(f, \Delta) \le \varepsilon. \tag{2.8.8}$$

Choose $b_i \in B_i$ for each i, and set

$$g := \sum_{i=1}^m f(b_i)\chi_{B_i}. \tag{2.8.9}$$

Then g is clearly paved with $\|g\|_\infty \le \|f\|_\infty$. Moreover,

$$\forall i \in \{1, \ldots, m\}, \forall x \in B_i, \qquad \|f(x) - g(x)\| = \|f(x) - f(b_i)\| \le W(f, B_i). \tag{2.8.10}$$

Hence noting that $\chi_P f = f$ and $\chi_P g = g$, we have

$$\begin{aligned} \int_{\mathbb{R}^n} \|f - g\| &= \int_P \|f - g\| = \sum_{i=1}^m \int_{B_i} \|f - g\| \\ &\le \sum_{i=1}^m \int_{B_i} W(f, B_i) = \sum_{i=1}^m W(f, B_i)\nu(B_i) \\ &= \omega(f, \Delta) \le \varepsilon, \end{aligned} \tag{2.8.11}$$

just as desired. □

Theorem 2.8.7 **(Fubini's Theorem for Banach Space Valued Riemann Integrals)**
Let E be a Banach space and $f \in \mathcal{R}(\mathbb{R}^n, E)$ is successively contented with respect to the decomposition $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$, then

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^l} f(x, y)\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^l} \left(\int_{\mathbb{R}^k} f(x, y)\mathrm{d}x \right) \mathrm{d}y. \tag{2.8.12}$$

Proof. — Note that f is bounded, we set $M := \|f\|_\infty < \infty$. Take any $\varepsilon > 0$, then by Lemma 2.8.6, there exists a paved $g : \mathbb{R}^n \to E$ such that $\|g\|_\infty \le \|f\|_\infty = M$, and

$$\int_{\mathbb{R}^n} \|f - g\| \le \varepsilon. \tag{2.8.13}$$

By Lemma 2.8.5, every integral in the following estimation makes sense:

$$
\begin{aligned}
&\left\|\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} f(x,y)\mathrm{d}y\right)\mathrm{d}x-\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} g(x,y)\mathrm{d}y\right)\mathrm{d}x\right\|\\
&=\left\|\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l}(f-g)(x,y)\mathrm{d}y\right)\mathrm{d}x\right\|\\
&\leq\int_{\mathbb{R}^k}\left\|\int_{\mathbb{R}^l}(f-g)(x,y)\mathrm{d}y\right\|\mathrm{d}x\\
&\leq\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l}\|f-g\|(x,y)\mathrm{d}y\right)\mathrm{d}x\\
&=\int_{\mathbb{R}^n}\|f-g\|\leq\varepsilon,
\end{aligned}
\tag{2.8.14}
$$

where the last equality follows from Theorem 2.8.1. Similarly,

$$
\begin{aligned}
&\left\|\int_{\mathbb{R}^l}\left(\int_{\mathbb{R}^k} f(x,y)\mathrm{d}x\right)\mathrm{d}y-\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} g(x,y)\mathrm{d}x\right)\mathrm{d}y\right\|\\
&\leq\int_{\mathbb{R}^l}\left(\int_{\mathbb{R}^k}\|f-g\|(x,y)\mathrm{d}x\right)\mathrm{d}y\\
&=\int_{\mathbb{R}^n}\|f-g\|\leq\varepsilon.
\end{aligned}
\tag{2.8.15}
$$

When g is of the form $c\chi_B$, the calculation in the proof of Lemma 2.8.5, especially (2.8.6), imply that

$$
\int_{\mathbb{R}^n} g=\int_{\mathbb{R}^l}\left(\int_{\mathbb{R}^k} g(x,y)\mathrm{d}x\right)\mathrm{d}y=\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} g(x,y)\mathrm{d}y\right)\mathrm{d}x.
\tag{2.8.16}
$$

From finite additivity of Riemann integrals, we see that (2.8.16) also holds for general paved map g. It follows from (2.8.13), (2.8.14), and (2.8.16) that

$$
\begin{aligned}
&\left\|\int_{\mathbb{R}^n} f-\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} f(x,y)\mathrm{d}y\right)\mathrm{d}x\right\|\\
&\leq\left\|\int_{\mathbb{R}^n} f-\int_{\mathbb{R}^n} g\right\|+\left\|\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} f(x,y)\mathrm{d}y\right)\mathrm{d}x-\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}^l} g(x,y)\mathrm{d}y\right)\mathrm{d}x\right\|\\
&\leq 2\int_{\mathbb{R}^n}\|f-g\|\leq 2\varepsilon.
\end{aligned}
\tag{2.8.17}
$$

Similarly, by (2.8.13), (2.8.15) and (2.8.14), we have

$$
\left\|\int_{\mathbb{R}^n} f-\int_{\mathbb{R}^l}\left(\int_{\mathbb{R}^k} f(x,y)\mathrm{d}x\right)\mathrm{d}y\right\|\leq 2\varepsilon.
\tag{2.8.18}
$$

Since $\varepsilon>0$ is arbitrary, (2.8.12) follows from (2.8.17) and (2.8.18). □

In practice, most examples of Riemann integrable maps that one usually encounters are successively contented. We have seen that this is the case for paved maps in Lemma 2.8.5. The following result says that this is also true for continuous maps with compact support.

Proposition 2.8.8. — *Let E be a Banach space. If $f:\mathbb{R}^n\to E$ is a compactly supported* continuous *map and $k,l\in\mathbb{N}^+$ with $n=k+l$, then the following hold:*

(1) *the map $\int_{\mathbb{R}^k} f(x,\cdot)\mathrm{d}x:\mathbb{R}^l\to E$ is continuous and compactly supported for all $x\in\mathbb{R}^k$;*
(2) *the map $\int_{\mathbb{R}^l} f(\cdot,y)\mathrm{d}y:\mathbb{R}^k\to E$ is continuous and compactly supported for all $y\in\mathbb{R}^l$;*
(3) *the map f is successively contented with respect to the decomposition $\mathbb{R}^n=\mathbb{R}^k\times\mathbb{R}^l$.*

Proof. — Let $\pi_1:\mathbb{R}^n\to\mathbb{R}^k$ and $\pi_2:\mathbb{R}^n\to\mathbb{R}^l$ be the corresponding projections corresponding to the decomposition $\mathbb{R}^n=\mathbb{R}^k\to\mathbb{R}^l$. Since $\mathrm{supp}(f)$ is compact, so is the continuous images $K_1:=\pi_1(\mathrm{supp}(f))\subseteq\mathbb{R}^k$ and $K_2:=\pi_2(\mathrm{supp}(f))\subseteq\mathbb{R}^l$. Clearly, $f(x,\cdot)$ and $f(\cdot,y)$ are continuous for all $x\in\mathbb{R}^k$ and $y\in\mathbb{R}^l$ with support contained respectively in K_2 and K_1. Hence these maps are Riemann integrable. By Definition 2.8.2, (3) now follows from (1) and (2). Since f is compactly supported, it follows that f is uniformly continuous (why?). Hence for any $\varepsilon>0$, there exists a $\delta>0$ such that for all $z_1,z_2\in\mathbb{R}^n$,

$\|z_1 - z_2\| \leq \delta$ implies $\|f(z_1) - f(z_2)\| \leq \varepsilon$. Now choose any $B_1 \in \mathcal{J}(\mathbb{R}^k)$ (resp. $B_2 \in \mathcal{J}(\mathbb{R}^l)$) such that $B_1 \supseteq K_1$ (resp. $B_2 \supseteq K_2$), we have

$$
\begin{aligned}
\Big(x_1, x_2 \in \mathbb{R}^k \text{ and } \|x_1 - x_2\| \leq \delta\Big) \implies \quad & \left\|\int_{\mathbb{R}^l} f(x_1, y)\mathrm{d}y - \int_{\mathbb{R}^l} f(x_2, y)\mathrm{d}y\right\| \\
& \leq \int_{\mathbb{R}^l} \|f(x_1, y) - f(x_2, y)\|\mathrm{d}y \\
& = \int_{B_2} \|f(x_1, y) - f(x_2, y)\|\mathrm{d}y \leq \varepsilon\nu(B_2).
\end{aligned}
\tag{2.8.19}
$$

This implies that $\int_{\mathbb{R}^l} f(\cdot, y)\mathrm{d}y$ is continuous. On the other hand,

$$
\forall x \in \mathbb{R}^k \setminus K_1, \qquad \int_{\mathbb{R}^k} f(x, y)\mathrm{d}y = 0 \tag{2.8.20}
$$

since $f(x, y) = 0$ for all $y \in \mathbb{R}^n$ for this $x \in \mathbb{R}^k \setminus K_1$. Hence (2) follows (2.8.19) and (2.8.20). Similarly, (1) holds and the proof is complete. □

2.9. The Change of Variables Formula

As usual, we fix a dimension $n \in \mathbb{N}^+$.

Recall the notion of C^1-diffeomorphism between two open sets of $\mathbb{R}^n$ (Definition 1.12.3). We regard such a C^1-diffeomorphism as a change of variable. Our goal in this section is to study how Riemann integrals behave under such change of variables.

Notation 2.9.1. — *In order to avoid overloading the same differentiation symbol "d", in this section, we write φ' instead of $\mathrm{d}\varphi$ for the derivative of φ provided of course that φ is differentiable. For technical reasons which will be clear soon, we will prefer to work with the norm $\|\cdot\|_\infty$ on $\mathbb{R}^n$, and use $B_\infty(a, r)$ to denote the open ball with center $a \in \mathbb{R}^n$ and radius r in $\mathbb{R}^n_\infty := \big(\mathbb{R}^n, \|\cdot\|_\infty\big)$. The corresponding closed ball is of course its closure*

$$
\overline{B_\infty(a, r)} = \big\{x \in \mathbb{R}^n \;\big|\; \|x - a\|_\infty \leq r\big\}. \tag{2.9.1}
$$

To begin, let us establish a result concerning negligible sets that will be useful to treat the set of discontinuous points when studying change of variables of Riemann integrable maps.

Lemma 2.9.2. — *Let $U \subseteq \mathbb{R}^n$ be open and $\varphi : U \to \mathbb{R}^n$ a C^1-map. If $A \subseteq U$ is negligible, then $\varphi(A)$ is also negligible.*

Proof. — It is clear that the countable set

$$
\mathfrak{B} := \{B_\infty(a, r) \mid a \in \mathbb{Q}^n,\ r \in \mathbb{Q}_{>0}\} \tag{2.9.2}
$$

is a topological basis for $\mathbb{R}^n$. Hence there exists a sequence $(B_m)_{m\geq1}$ in $\mathfrak{B}$ such that $U = \cup_{m\geq1} B_m$. For any $a \in \mathbb{Q}^n$ and $r \in \mathbb{Q}_{>0}$, it follows from

$$
B_\infty(a, r) = \bigcup_{s \in \mathbb{Q} \cap]0, r[} \overline{B_\infty(a, s)} = \bigcup_{s \in \mathbb{Q} \cap]0, r[} B_\infty(a, s) \tag{2.9.3}
$$

that we may suppose each B_m satisfies $\overline{B_m} \subseteq U$ and we also have $U = \cup_{m\geq1} \overline{B_m}$. Clearly,

$$
\varphi(A) = \bigcup_{m\geq1} \varphi(A \cap \overline{B_m}). \tag{2.9.4}
$$

Since negligible sets are stable under taking countable unions, to finish the proof, it suffices to show that $\varphi(A \cap \overline{B_m})$ is negligible for each $m \in \mathbb{N}^+$.

Let c_m and r_m be the center and radius of $\overline{B_m} \subseteq U$. Since φ' is continuous on U and the norm map $\|\cdot\|_{\mathcal{L}(\mathbb{R}^n_\infty)} : \mathcal{L}(\mathbb{R}^n_\infty) \to \mathbb{R}$ is always continuous, on the compact set $\overline{B_m} \subseteq U$, we may set

$$
M_m := \max_{x \in \overline{B_m}} \|\varphi'(x)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \in [0, +\infty[. \tag{2.9.5}
$$

By Corollary 1.4.3, we have

$$\forall x, y \in \overline{B_m}, \qquad \|\varphi(x) - \varphi(y)\| \le M\|x - y\|. \tag{2.9.6}$$

Clearly, $\overline{B_m}$ is the closure of a unique rational box[7] $\widetilde{B_m}$ in $\mathbb{R}^n$ and $\overline{B_m} \setminus \widetilde{B_m}$ is negligible. Thus to show $A \cap \overline{B_m}$, it suffices to show that $A \cap \widetilde{B_m}$ is negligible, which we now do.

Take any $\varepsilon > 0$. Since A is negligible, so is $A \cap \widetilde{B_m}$. There are countably many boxes $(C_l)_{l \ge 1}$ in $\mathbb{R}^n$, such that $A \cap \widetilde{B_m} \subseteq \cup_{l \ge 1} C_l$ and $\sum_{l \ge 1} \nu(C_l) \le \varepsilon$. Enlarging each C_l slightly, we may suppose each C_l is rational. By replacing each C_l with the grid boxes in the grid box decomposition of the rational paved set $\widetilde{B_m} \cap C_l$ (for some level $N_{l,m}$) if necessary, we may suppose for each l, we have $C_l \subseteq \widetilde{B_m} \subseteq \overline{B_m}$ and $\overline{C_l}$ is a closed ball in $\mathbb{R}^n_\infty$ of radius r_l and center c_l, with $\overline{C_l} \subseteq \overline{B_m}$. By (2.9.6), we see that

$$\forall x \in \overline{C_l}, \qquad \|\varphi(x) - \varphi(c_l)\|_\infty \le M\|x - c_l\|_\infty \le M r_l \tag{2.9.7}$$

so $\varphi(C_l) \subseteq \varphi(\overline{C_l}) \subseteq \overline{B_\infty(\varphi(c_l), M r_l)}$. Clearly,

$$\varphi(A \cap \widetilde{B_m}) \subseteq \bigcup_{l \ge 1} \varphi(C_l) \subseteq \bigcup_{l \ge 1} \overline{B_\infty(\varphi(c_l), M r_l)} \tag{2.9.8}$$

and

$$\nu\left(\overline{B_\infty(\varphi(c_l), M r_l)}\right) = M^n (2r_l)^n = M^n \nu(C_l). \tag{2.9.9}$$

Hence

$$\sum_{l \ge 1} \nu\left(\overline{B_\infty(\varphi(c_l), M r_l)}\right) \le M^n \sum_{l \ge 1} \nu(C_l) \le M^n \varepsilon. \tag{2.9.10}$$

By (2.9.8) and (2.9.10), and note that $\varepsilon > 0$ is arbitrary, we conclude that $\varphi(A \cap \widetilde{B_m})$ is indeed negligible, and the proof is complete. □

Corollary 2.9.3. — *Let $U \subseteq \mathbb{R}^n$ be open and $\varphi : U \to \mathbb{R}^n$ a C^1-map. If $B \in \mathcal{J}(\mathbb{R}^n)$ with $\overline{B} \subseteq U$, then $\varphi(B) \in \mathcal{J}(\mathbb{R}^n)$.*

Proof. — Now take any such B, so $\partial B = \overline{B} \setminus \operatorname{Int}(B) = B \setminus \operatorname{Int}(B) \subseteq U$ is compact and Jordan negligible, hence negligible. By Lemma 2.9.2, $\varphi(\partial B) = \partial\varphi(B)$ is negligible and compact (since φ is a homeomorphism). Any compact negligible set is Jordan negligible, hence $\partial\varphi(B)$ is negligible. Being compact, $\varphi(B)$ is bounded. Hence $\varphi(B) \in \mathcal{J}(\mathbb{R}^n)$, as desired. □

When dealing with change of variable, the following topological result is useful for finding a lot suitable small (closed) boxes that will be useful for approximating the Riemann integrals using Riemann sums.

Lemma 2.9.4. — *Let U be an open set in $\mathbb{R}^n$, and K a compact subset of U. There exists a $\delta > 0$, such that for any closed ball $B = \overline{B_\infty(a, r)}$ with $a \in \mathbb{R}^n$ and $r \in]0, \delta]$ in $\mathbb{R}^n_\infty$, we have*

$$B \cap K \ne \emptyset \implies B \subseteq U. \tag{2.9.11}$$

Proof. — For each $x \in K \subseteq U$, choose an $r_x > 0$ such that $B_\infty(x, 2r_x) \subseteq U$. Clearly, $\{B_\infty(x, r_x) \mid x \in K\}$ is an open cover of the compact set K, and there are finitely points $x_1, \dots, x_l$ in K, such that

$$K \subseteq \bigcup_{i=1}^{l} B_\infty(x_i, r_{x_i}). \tag{2.9.12}$$

Set $\delta := \min_{1 \le i \le l} r_{x_i} > 0$. For any $B = \overline{B_\infty(a, r)}$ with $a \in \mathbb{R}^n$ and $r \in]0, \delta]$, if $B \cap K \ne \emptyset$, we may take $b \in B \cap K$, then $b \in B_\infty(x_j, r_{x_j})$ for some $j \in \{1, \dots, l\}$, and

$$\forall x \in B, \qquad \|x - x_j\|_\infty \le \|x - b\|_\infty + \|b - x_j\|_\infty < r + r_{x_j} \le \delta + r_{x_j} \le 2r_{x_j}, \tag{2.9.13}$$

which implies

$$B \subseteq B_\infty(x_j, 2r_{x_j}) \subseteq U. \tag{2.9.14}$$

□

[7] We actually have $\widetilde{B_m} = [c_{1,m} - r_m, c_{1,m} + r_m[\times \cdots \times [c_{n,m} - r_m, c_{n,m} + r_m[$, where $(c_{k,m})_{1 \le k \le n} \in \mathbb{R}^n$ is the coordinate of c_m.

We also need a good estimation of the volume of image of many small boxes under a change of variable. Our main idea for this purpose is to use affine maps to approximate a differentiable map locally. Since we are dealing with C^1-diffeomorphism as our change of variable, to be more precise, we want to, at least locally, use *invertible* affine map that looks like our change variable, such that the inverse of this affine map looks like the inverse of the change of variable.

Using the above idea, our intuition tells us that the C^1-diffeomorphisms α and β in Lemma 2.9.5 below should be close to a translation, and this lemma provides an estimation of the deviation.

Lemma 2.9.5. — *Let $\varphi : U \to V$ be a C^1-diffeomorphism between open sets in $\mathbb{R}^n$ with inverse $\psi : V \to U$. Suppose $a \in U$. Set $A := \varphi'(a)$, $b = \varphi(a) \in V$ and $B = \psi'(b) = \varphi'(a)^{-1}$. Consider the C^1-diffeomorphism $\alpha := A^{-1}\varphi = B\varphi$ from U onto $B(V)$ and its inverse $\beta := \psi A = \psi B^{-1}$ from $B(V)$ onto U. Then the following hold:*

(1) *If $x \in U$ and $\{a + t(x-a) \mid t \in [0,1]\} \subseteq U$, then*

$$\|\alpha(x) - \alpha(a) - (x-a)\|_\infty \leq \sup_{t\in]0,1[} \big\|B\big(\varphi'(a+t(x-a)) - A\big)\big\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \cdot \|x-a\|_\infty. \tag{2.9.15}$$

(2) *If $y \in B(V)$ and $\big\{A^{-1}b + t(y - A^{-1}b) \bigm| t \in [0,1]\big\} \subseteq B(V)$, then*

$$\big\|\beta(y) - \beta(A^{-1}b) - (y - A^{-1}b)\big\|_\infty \leq \sup_{t\in]0,1[} \big\|\big(\psi'(b+t(Ay-b)) - B\big)A\big\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \cdot \big\|y - A^{-1}b\big\|_\infty. \tag{2.9.16}$$

(3) *Suppose $\varepsilon \in]0,1[$, $r > 0$, such that*

$$B_\infty(a,r) \subseteq U, \tag{2.9.17}$$

and

$$B_\infty(A^{-1}b, (1-\varepsilon)r) \subseteq B(V). \tag{2.9.18}$$

If

$$\forall x \in B_\infty(a,r), \qquad \|B(\varphi'(x) - A)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \leq \varepsilon \tag{2.9.19}$$

and

$$\forall y \in B_\infty(A^{-1}b, (1-\varepsilon)r), \qquad \|(\psi'(Ay) - B)A\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \leq \varepsilon, \tag{2.9.20}$$

then

$$A\left(B_\infty\big(A^{-1}b, (1-\varepsilon)r\big)\right) \subseteq \varphi\left(B_\infty(a,r)\right) \subseteq A\left(B_\infty\big(A^{-1}b, (1+\varepsilon)r\big)\right). \tag{2.9.21}$$

Proof. — (1). This follows from Corollary 1.4.3 applied to $\alpha_0(x) = \alpha(x) - x$ and note that

$$\alpha_0'(x) = A^{-1}\varphi'(x) - \mathrm{Id} = A^{-1}\big(\varphi'(x) - A\big) = B\big(\varphi'(x) - A\big).$$

(2). Similarly, this follows from Corollary 1.4.3 applied to $\beta_0(y) = \beta(y) - y$ and note that

$$\beta_0'(y) = \psi'(Ay)B^{-1} - \mathrm{Id} = \big(\psi'(Ay) - B\big)B^{-1} = \big(\psi'(Ay) - B\big)A. \tag{2.9.22}$$

(3). By (1) and (2.9.19), we have

$$\begin{aligned}\forall x \in B_\infty(a,r), \qquad &\|\alpha(x) - \alpha(a)\|_\infty \\ &\leq \|x-a\| + \|\alpha(x) - \alpha(a) - (x-a)\| \\ &< (1+\varepsilon)\|x-a\|_\infty.\end{aligned} \tag{2.9.23}$$

This implies

$$\alpha\left(B_\infty(a,r)\right) \leq B_\infty\big(\alpha(a), (1+\varepsilon)r\big), \tag{2.9.24}$$

which, considering $\alpha = A^{-1}\varphi$ and $\alpha(a) = A^{-1}\varphi(a) = A^{-1}b$, implies

$$\varphi\left(B_\infty(a,r)\right) \subseteq A\left(B_\infty\big(A^{-1}b, (1+\varepsilon)r\big)\right). \tag{2.9.25}$$

On the other hand, by (2) and (2.9.20), we have

$$\begin{aligned}\forall y \in B_\infty\big(A^{-1}b, (1-\varepsilon)r\big), \qquad &\big\|\beta(y) - \beta(A^{-1}b)\big\| \\ &\le \big\|\beta(y) - \beta(A^{-1}b) - (y - A^{-1}b)\big\| + \big\|y - A^{-1}b\big\| \\ &< (\varepsilon+1)\big\|y - A^{-1}b\big\| \\ &\le (1+\varepsilon)(1-\varepsilon)r = (1-\varepsilon^2)r < r.\end{aligned} \tag{2.9.26}$$

Note that $\beta = \psi A$, so $\beta(A^{-1}b) = \psi(b) = a$, this implies that

$$\beta\left(B_\infty(A^{-1}b, (1-\varepsilon)r)\right) \subseteq B_\infty(a, r). \tag{2.9.27}$$

Apply $A\alpha = \varphi$ on both sides of (2.9.27) and note that $\alpha = \beta^{-1}$, we obtain

$$A\left(B_\infty(A^{-1}b, (1-\varepsilon)r)\right) \subseteq \varphi\left(B_\infty(a, r)\right). \tag{2.9.28}$$

Now (2.9.21) follows from (2.9.28) and (2.9.25). □

Combined with Proposition 2.5.1, the last statement of Lemma 2.9.5 provides a good way to estimate the volume of $\varphi\left(B_\infty(a,r)\right)$. The following result uses this idea to simultaneously estimate $\nu\big(\varphi(B)\big)$ for a large number of boxes B.

Lemma 2.9.6. — *Let $\varphi : U \to V$ be a C^1-diffeomorphism between open sets in $\mathbb{R}^n$, and K a compact subset of U. For each $\varepsilon \in\,]0,1[$, there exists a $\eta > 0$, such that for any* cubic *box B in $\mathbb{R}^n$, i.e. $\overline{B} = \overline{B_\infty(a,r)}$ for some $a \in \mathbb{R}^n$ and $r \in\,]0,\eta]$ (or equivalently, $\mathrm{Int}(B) = B_\infty(a,r)$), if $B \cap K \neq \emptyset$ then $\varphi(B) \in \mathcal{J}(\mathbb{R}^n)$, and*

$$(1-\varepsilon)|\det \varphi'(a)|\nu(B) \le \nu\big(\varphi(B)\big) \le (1+\varepsilon)|\det \varphi'(a)|\nu(B). \tag{2.9.29}$$

Proof. — To facilitate discussion, we still use $\psi : V \to U$ to denote the inverse of φ in this proof.

By Lemma 2.9.4, there exists $\delta_U > 0$, such that

$$K_{\delta_U} := \left\{x \in \mathbb{R}^n \;\middle|\; d_\infty(x, K) \le \delta_U\right\} \subseteq U, \tag{2.9.30}$$

where d_∞ is the metric on $\mathbb{R}^n$ induced by the norm $\|\cdot\|_\infty$. Clearly, K_{δ_U} is compact. Since φ is a homeomorphism, set $L := \varphi(K) \subseteq V$, then L is compact, and there exists, by the same Lemma 2.9.4, a $\delta_V > 0$, such that

$$L_{\delta_V} := \left\{y \in \mathbb{R}^n \;\middle|\; d_\infty(y, L) \le \delta_V\right\} \subseteq V, \tag{2.9.31}$$

and L_{δ_V} is compact. By compactness of K_{δ_U} and L_{δ_V} as well as the continuity of φ' and $(\varphi^{-1})'$, we may set

$$M_U := \max_{x \in K_{\delta_U}} \|\varphi'(x)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \in\,]0, +\infty[. \tag{2.9.32}$$

Now suppose $\varepsilon_0 > 0$, which is to be determined later. Since K_{δ_U} and L_{δ_V} are compact, the restrictions $\varphi\restriction_{K_{\delta_U}}$, $\varphi'\restriction_{K_{\delta_U}}$, $\psi'\restriction_{L_{\delta_V}}$ and $(\psi' \circ \varphi)\restriction_{K_{\delta_U}}$ are all uniformly continuous. We may therefore choose an $\eta \in\,]0, \delta_U[$, such that

$$M_U \eta \le \frac{\delta_V}{2}, \tag{2.9.33}$$

that,

$$\forall x_1, x_2 \in K_{\delta_U}, \qquad \|x_1 - x_2\|_\infty \le \eta \implies \begin{cases} \|\varphi'(x_1) - \varphi'(x_2)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \cdot M_{\psi'} \le \varepsilon_0 \\ \|\varphi(x_1) - \varphi(x_2)\|_\infty \le \dfrac{\delta_V}{2}, \end{cases} \tag{2.9.34}$$

and that

$$\forall z_1, z_2 \in L_{\delta_V}, \qquad \|z_1 - z_2\|_\infty \le M_{\varphi'}\eta \implies \|\psi'(z_1) - \psi'(z_2)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \cdot M_{\varphi'} \le \varepsilon_0, \tag{2.9.35}$$

where

$$M_{\varphi'} := \max_{x \in K_\delta} \|\varphi'(x)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \in\,]0, +\infty[, \tag{2.9.36}$$

and

$$M_{\psi'} := \max_{z\in\varphi(K_\delta)} \|\psi'(z)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \in]0,+\infty[. \tag{2.9.37}$$

Suppose B is a cubic box as in the statement of the lemma, and let $B_0 = B_\infty(a,r)$ be the interior of B. Since $\emptyset \neq B\cap K \subseteq \overline{B}\cap K$ and $r \leq \eta < \delta_U$, it follows from (2.9.30) that

$$B_0 = \mathrm{Int}(B) = B_\infty(a,r) \subseteq \overline{B} = \overline{B_\infty(a,r)} \subseteq K_{\delta_U} \subseteq U. \tag{2.9.38}$$

Now set $S = \varphi'(a) \in \mathcal{L}(\mathbb{R}^n) = \mathrm{Aut}(\mathbb{R}^n)$, $b = \varphi(a)$ and $T = \psi'(b) = S^{-1}$. Choose $x_0 \in B\cap K$ and set

$$y_0 = \varphi(x_0) \in \varphi(B)\cap\varphi(K) = \varphi(B)\cap L.$$

Since $\|x_0 - a\|_\infty \leq r < \eta$, we have, by (2.9.34) and (2.9.38),

$$\|y_0 - b\|_\infty = \|\varphi(x_0) - \varphi(a)\| \leq \frac{\delta_V}{2}. \tag{2.9.39}$$

We claim that

$$\forall z \in S\left(B_\infty(S^{-1}b,(1-\varepsilon_0)r)\right), \qquad \|z-y_0\|_\infty \leq \delta_V. \tag{2.9.40}$$

Indeed, take any such z, set $y := S^{-1}z \in B_\infty(S^{-1}b,(1-\varepsilon_0)r)$, we have

$$\begin{aligned}
\|z-y_0\| &= \|Sy - y_0\|_\infty \\
&\leq \|Sy-b\|_\infty + \|y_0 - b\|_\infty \\
&\leq \left\|S(y-S^{-1}b)\right\|_\infty + \frac{\delta_V}{2} \\
&\leq \|\varphi'(a)\|_{\mathcal{L}(\mathbb{R}^n_\infty)}\cdot\left\|y-S^{-1}b\right\| + \frac{\delta_V}{2} \\
&\leq M_U(1-\varepsilon_0)r + \frac{\delta_V}{2} \\
&\leq M_U r + \frac{\delta_V}{2} \leq M_U\eta + \frac{\delta_V}{2} \leq \delta_V,
\end{aligned} \tag{2.9.41}$$

where the last inequality follows from (2.9.33).

Combining (2.9.40) with (2.9.31) and recall $T = \psi'(b) = S^{-1}$, we see that that

$$B_\infty(S^{-1}b,(1-\varepsilon_0)r) \subseteq T(V). \tag{2.9.42}$$

Since $B\cap K \neq \emptyset$, we must have $a \in K_{\delta_U}$ and consequently $b = \varphi(a) \in \varphi(K_{\delta_U})$, so $\|S\|_{\mathcal{L}(\mathbb{R}^n_\infty)} = \|\varphi'(a)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \leq M_{\varphi'}$ and $\|T\|_{\mathcal{L}(\mathbb{R}^n_\infty)} = \|\psi'(b)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \leq M_{\psi'}$. It now follows from (2.9.34) again that

$$\begin{aligned}
\forall x \in B_\infty(a,r) \subseteq U, \qquad & \left\|T\big((\varphi'(x)) - S\big)\right\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \\
&\leq \|T\|_{\mathcal{L}(\mathbb{R}^n_\infty)}\cdot\|\varphi'(x)-\varphi'(a)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \\
&\leq M_{\psi'}\|\varphi'(x)-\varphi'(a)\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \leq \varepsilon_0.
\end{aligned} \tag{2.9.43}$$

Similarly, take any $y \in B_\infty(S^{-1}b,(1-\varepsilon_0)r) \subseteq T(V)$, then we have $Sy \in L_{\delta_V}$ by (2.9.40). In particular, $b = S(S^{-1}b) \in L_{\delta_V}$. Moreover,

$$\begin{aligned}
\|Sy-b\|_\infty &= \left\|S(y-S^{-1}b)\right\|_\infty \\
&\leq \|S\|_{\mathcal{L}(\mathbb{R}^n_\infty)}\cdot\left\|y-S^{-1}b\right\|_\infty \\
&\leq M_{\varphi'}\cdot(1-\varepsilon_0)r \leq M_{\varphi'}\eta,
\end{aligned} \tag{2.9.44}$$

hence by (2.9.35),

$$\begin{aligned}
\forall y \in B_\infty(S^{-1}b,(1-\varepsilon_0)r) \subseteq T(V), \qquad & \left\|\big(\psi'(Sy)-T\big)S\right\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \\
&\leq \left\|\big(\psi'(Sy)-\psi'(b)\big)\right\|_{\mathcal{L}(\mathbb{R}^n_\infty)}\cdot\|S\|_{\mathcal{L}(\mathbb{R}^n_\infty)} \\
&\leq \left\|\big(\psi'(Sy)-\psi'(b)\big)\right\|_{\mathcal{L}(\mathbb{R}^n_\infty)}\cdot M_{\varphi'} \leq \varepsilon_0.
\end{aligned} \tag{2.9.45}$$

By (2.9.38), (2.9.42), (2.9.43) and (2.9.45), we may apply Lemma 2.9.5 to the open cube $B_0 = B_\infty(a,r)$, and obtain

$$S\left(B_\infty(S^{-1}b,(1-\varepsilon_0)r)\right) \subseteq \varphi(B_0) \subseteq S\left(B_\infty(S^{-1}b,(1+\varepsilon_0)r)\right). \tag{2.9.46}$$

Clearly $B_\infty(S^{-1}b,r)$ is a translation of B_0, so these two open cubes have the same volume. It follows from Proposition 2.5.1 and (2.9.46) that

$$(1-\epsilon_0)^n\nu(B_0) \leq \nu\big(\varphi(B_0)\big) \leq (1+\varepsilon_0)^n\nu(B_0). \tag{2.9.47}$$

Since $B \setminus B_0 \subseteq \partial B_0$ is Jordan negligible, we have $\nu(B) = \nu(B_0)$. Now that $\varphi(B\setminus B_0) \subseteq \phi(\partial B_0)$ and $\phi(\partial B_0)$ is both compact and negligible (Lemma 2.9.2), we see immediately from the open box cover definition of negligible sets that $\phi(\partial B_0) = \partial\varphi(B_0)$ must be Jordan negligible. Hence the subset $\varphi(B)\setminus\varphi(B_0) = \phi(B\setminus B_0) \subseteq \varphi(B_0)$ is also Jordan negligible and we also have $\nu\big(\varphi(B)\big) = \nu\big(\varphi(B_0)\big)$. Combining with $\nu(B) = \nu(B_0)$ above and (2.9.47), we have

$$(1-\epsilon_0)^n\nu(B) \leq \nu\big(\varphi(B)\big) \leq (1+\varepsilon_0)^n\nu(B). \tag{2.9.48}$$

This finishes the proof by taking ε_0 so small that

$$1-\varepsilon < (1-\varepsilon_0)^n < (1+\varepsilon_0)^n < 1+\varepsilon. \tag{2.9.49}$$

□

We are now well prepared for the following key result.

Theorem 2.9.7 **(Change of Variable Formula for Riemann Integrals)**

Let U,V be open sets in $\mathbb{R}^n$, E a Banach space and $\varphi : U \to V$ a C^1-diffeomorphism. Then we have

$$f \in \mathcal{R}(V,E) \iff f\circ\varphi \in \mathcal{R}(U,E) \iff (f\circ\varphi)|\det\varphi'| \in \mathcal{R}(U,E) \tag{2.9.50}$$

and in this case,

$$\int_V f = \int_U (f\circ\varphi)(x)|\det\varphi'(x)|\mathrm{d}x. \tag{2.9.51}$$

Proof. — We first establish (2.9.50). Since $\varphi : U \to V$ is in particular a homeomorphism, we have $\varphi\big(D(f)\big) = D(f\circ\varphi)$. By Lemma 2.9.2 and note that $\varphi : U \to V$ is a C^1-diffeomorphism, we see that $D(f)$ is negligible if and only if $D(f\circ\varphi) = D\big((f\circ\varphi)|\det\varphi'|\big)$ is so. On the other hand, using the fact that $|\det\varphi'| : U \to \mathbb{R}$ never vanishes and continuous, we see that

$$\operatorname{supp}\Big((f\circ\varphi)|\varphi'|\Big) = \operatorname{supp}(f\circ\varphi) = \varphi\big(\operatorname{supp}(f)\big), \tag{2.9.52}$$

where the last equality follows since $\varphi : U \to V$ is a homeomorphism. Hence the three maps in (2.9.50) are simultaneously compactly supported or simultaneously not compactly supported. To finish the proof of (2.9.50), it remains to establish the equivalence of the boundedness of the three maps in (2.9.50) under the assumption that all three maps are compactly supported (see Notation 2.7.17). But it is clear that f and $f\circ\varphi$ have the same range, so they are either both bounded or both unbounded. Note that $|\det\varphi'|$ has its range in some compact interval in $]0,+\infty[$ on the compact set $\operatorname{supp}(f\circ\varphi)$, we see that the boundedness of $f\circ\varphi$ and $(f\circ\varphi)|\det\varphi'|$ are equivalent. This finishes the proof of (2.9.50).

Now assume $f \in \mathcal{R}(V,E)$. It remains to establish (2.9.51). Set $g := (f\circ\varphi)|\det\varphi'| \in \mathcal{R}(U,E)$. Since $\operatorname{supp}(g)$ is a compact set in U, use the grid decomposition (2.4.13) for a large enough level $N_0 \in \mathbb{N}^+$, by Lemma 2.9.4, we may find a rational paved set $P_0 \in \mathcal{P}(\mathbb{R}^n)$ (of level N_0, but this is not important since we might raise the level later), such that

$$\operatorname{supp}(g) \subseteq P_0 \subseteq \overline{P_0} \subseteq U. \tag{2.9.53}$$

Fix an arbitrary $\varepsilon > 0$. By Theorem 2.6.20, we may find a box partition $\Delta_0 = (B_i^{(0)})_{i=1}^{m_0}$ of P_0, such that $\omega(g,\Delta_0) < \varepsilon$. By replacing each $B_i^{(0)}$ with a rational box $B_i^{(1)}$ inside $B_i^{(0)}$ whose volume is close to $B_i^{(0)}$, and set

$$B_0^{(1)} := \bigsqcup_{i=1}^{m_0}\left(B_i^{(0)}\setminus B_i^{(1)}\right) = P_0\setminus\left(\sqcup_{i=1}^{m_0}B_i^{(1)}\right), \tag{2.9.54}$$

we may make $\nu(B_0^{(1)})$ as small as we wish, while keeping $B_i^{(1)}$ being a rational paved set for each $i = 0, \ldots, m_0$ (for the rationality of $B_0^{(1)}$, use (2.9.54) and the rationality of P_0). Since $\omega(g, B_i^{(1)}) \leq \omega(g, B_i^{(0)})$ for $i \geq 1$ and g is bounded, $\Delta_1 = \left(B_i^{(1)}\right)_{i=0}^{m_0}$ is a Jordan partition of P_0 by rational paved set, and, by making $\nu(B_0^{(1)})$ small enough as described above, we have

$$\omega(g, \Delta_1) \leq 2M\nu\left(B_0^{(1)}\right) + \sum_{i=1}^{m_0} \omega(g, B_i^{(1)}) \leq 2M\nu\left(B_0^{(1)}\right) + \omega(g, \Delta_0) \leq \varepsilon, \tag{2.9.55}$$

where

$$M := \sup_{x \in U} \|g(x)\| < +\infty. \tag{2.9.56}$$

Using the partition Δ_1, we may find (see Lemma 2.4.10) an $N_2 \in \mathbb{N}^+$ and a finer box partition $\Delta_2 = (B_i^{(2)})_{i=1}^{m_2}$ of P_0, where each box in Δ_2 is a grid box of level N_2. Now let $\eta > 0$ be as in Lemma 2.9.6 applied in the case $K := \overline{P_0}$, we may, by Lemma 2.4.10 again, raise the level N_2 to a large integer multiple of N_2 if necessary, and find a box partition $\Delta = (B_i)_{i=1}^m$ of P_0 that is finer than Δ_2, where each B_i is a grid box of level N for some large $N \in \mathbb{N}^+$ with $1/(2N) < \eta$ and (use the uniform continuity of $|\det \varphi'|^{-1}$ on the compact set $\overline{P_0} \subseteq U$)

$$\forall s, t \in \overline{P_0}, \quad \|s - t\|_\infty \leq \frac{1}{N} \implies \left| |\varphi'(s)|^{-1} - |\varphi'(t)|^{-1} \right| \leq \varepsilon. \tag{2.9.57}$$

Let c_i be the center of B_i for each i, and we have $\overline{B_i} = \overline{B_\infty(c_i, 1/(2N))}$. Now Lemma 2.9.6 says that

$$\forall i \in \{1, \ldots, m\}, \qquad (1 - \varepsilon)|\det \varphi'(c_i)|\nu(B_i) \leq \nu\big(\varphi(B_i)\big) \leq (1 + \varepsilon)|\det \varphi'(c_i)|\nu(B_i). \tag{2.9.58}$$

Set $\Delta_f = \big(\varphi(B_i)\big)_{i=1}^m$. Then by Corollary 2.9.3, we see that Δ_f is a Jordan partition of the relatively compact Jordan set $Q_0 := \varphi(P_0) \subseteq V$ that contains $\operatorname{supp}(f)$. Now let

$$s_f := \sum_{i=1}^m f(\varphi(c_i))\nu\big(\varphi(B_i)\big) \in S(f, \Delta_f). \tag{2.9.59}$$

and

$$s_g := \sum_{i=1}^m g(c_i)\nu(B_i) \in S(g, \Delta). \tag{2.9.60}$$

To relate the Riemann sums s_f and s_g to $\int_V f$ and $\int_U g$, on the one hand, we have

$$\omega(g, \Delta) \leq \omega(g, \Delta_2) \leq \omega(g, \Delta_1) \leq \varepsilon. \tag{2.9.61}$$

So

$$\left\| s_g - \int_U g \right\| \leq \omega(g, \Delta) \leq \varepsilon. \tag{2.9.62}$$

On the other hand, we have

$$\begin{aligned}
&\forall i \in \{1, \ldots, m\}, \forall a, b \in B_i \subseteq P_0 \subseteq \overline{P_0}, \\
&\quad f(\varphi(a)) - f(\varphi(b)) = g(a)|\det \varphi'(a)|^{-1} - g(b)|\det \varphi'(b)|^{-1} \\
&= \big(g(a) - g(b)\big)|\det \varphi'(a)|^{-1} + g(b)\left(|\det \varphi'(a)|^{-1} - |\det \varphi'(b)|^{-1}\right).
\end{aligned} \tag{2.9.63}$$

Note that $|\det \varphi'|^{-1}$ is continuous on the compact set $\overline{P_0} \subseteq U$, we have

$$L_1 := \sup_{x \in \overline{P_0}} |\det \varphi'(x)|^{-1} = \max_{x \in \overline{P_0}} |\det \varphi'(x)|^{-1} < +\infty. \tag{2.9.64}$$

Similarly,

$$L_2 := \sup_{x \in \overline{P_0}} |\det \varphi'(x)| = \max_{x \in \overline{P_0}} |\det \varphi'(x)| < +\infty. \tag{2.9.65}$$

Recall (2.9.56), note that any two points in B_i has distance (with respect to $\|\cdot\|_\infty$) less than $1/N$ so (2.9.57) is applicable, (2.9.63) implies that

$$W(f, \varphi(B_i)) \le L_1 \cdot W(g, B_i) + M\varepsilon. \tag{2.9.66}$$

Hence by (2.9.58) and note that $|\det \varphi'(c_i)| \le L_2$ by the definition of L_2, we have

$$\begin{aligned}\omega(f, \varphi(B_i)) &\le L_1 W(g, B_i)\nu(\varphi(B_i)) + M\varepsilon\nu(\varphi(B_i)) \\ &\le L_1 W(g, B_i)(1+\varepsilon)|\det \varphi'(c_i)|\nu(B_i) + M\varepsilon\nu(\varphi(B_i)) \\ &\le L_1 L_2(1+\varepsilon)\omega(B_i) + M\varepsilon\nu(\varphi(B_i)).\end{aligned} \tag{2.9.67}$$

Summing (2.9.67) over ε, we obtain

$$\omega(f, \Delta_f) \le L_1 L_2(1+\varepsilon)\omega(g, \Delta) + \varepsilon M\nu(Q_0) \le L_1 L_2(1+\varepsilon)\varepsilon + \varepsilon M\nu(Q_0). \tag{2.9.68}$$

So

$$\left\| s_f - \int_V f \right\| \le \omega(f, \Delta_f) \le L_1 L_2(1+\varepsilon)\varepsilon + \varepsilon M\nu(Q_0). \tag{2.9.69}$$

Setting

$$M_f := \sup_{x \in V} \|f(x)\| < +\infty, \tag{2.9.70}$$

we also have, by (2.9.58), that

$$\begin{aligned}\|s_f - s_g\| &= \left\| \sum_{i=1}^m f(\varphi(c_i))\Big(\nu(\varphi(B_i)) - |\varphi'(c_i)|\nu(B_i)\Big) \right\| \\ &\le \sum_{i=1}^m \|f(\varphi(c_i))\| \cdot \varepsilon|\varphi'(c_i)|\nu(B_i) \\ &\le \varepsilon M_f L_2 \sum_{i=1}^m \nu(B_i) = M_f L_2 \nu(P_0)\varepsilon.\end{aligned} \tag{2.9.71}$$

Since $\varepsilon > 0$ is arbitrary, by (2.9.62), (2.9.69) and (2.9.71), we must have $\int_U g = \int_V f$, as desired. The proof is now complete. ☐

2.10. Polar, Cylindrical and Spherical Coordinates

2.11. Improper Riemann Integrals, the Gaussian Integral and the Volumes of Balls

2.12. A Rapid Peek into the Lebesgue Theory: Bourbaki's Integral First Approach