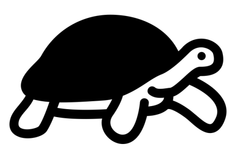
MATH 247 Notes



This page intentionally left blank.

Contents

This page intentionally left blank.

1.
Integration

Half-open Rectangles

(1.1)
Half-open Rectangles

For the development of calculus, it is important that integration theory can be pursued in the space $\mathbb{R}^n$ for an arbitrary dimension $n \in \mathbb{N}$. In order to build such an integration theory, we need a reliable collection of sets to which we know how to assign an n-dimensional volume. The family of sets we shall start with are the n-dimensional analogues of a rectangle, or of a parallelipiped. Theses are sets $P \subseteq \mathbb{R}^n$ obtained as Cartesian products of intervals of $\mathbb{R}$:

$$P = \prod_{i=1}^{n} J_i = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 \in J_1, \dots, x_n \in J_n\},$$

where $J_1, \dots, J_n \in \mathbb{R}$ are intervals. In order to assign to a finite volume to a set P as above, we shall ask that the intervals $J_1, \dots, J_n$ are bounded. Moreover, it would be convenient that we pick just one type of interval, among the various possibilities of closed, open, and half-open intervals, and use it consistently throughout proofs and calculations. Due to reasons which will be apparent as we proceed, we shall work with *half-open* intervals, intervals of the form $(a,b] \subseteq \mathbb{R}$ for some $a,b \in \mathbb{R}$.

Def'n. Half-open Rectangle in an Euclidean Space

A ***half-open rectangle*** P in $\mathbb{R}^n$ is a set of the form

$$P = \prod_{i=1}^{n} (a_i, b_i] = \{(x_1, \dots, x_n) : \forall i \in \{1, \dots, n\}\, a_i < x_i \le b_i\}.$$

Moreover, we shall denote the collection of every half-open intervals in $\mathbb{R}^n$ by $\mathcal{P}_n$.

(1.2)
vol_n Function

To assign an n-dimensional volume to each element of $\mathcal{P}_n$, we define $\mathrm{vol}_n : \mathcal{P}_n \to [0, \infty)$ such that

(a) $\mathrm{vol}_n(\varnothing) = 0$ and

(b) for every $a_1, b_1, \dots, a_n, b_n \in \mathbb{R}$ with $a_i < b_i$ for all i, we have

$$\mathrm{vol}_n((a_1, b_1], \dots, (a_n, b_n]) = \prod_{i=1}^{n} b_i - a_i.$$

(1.3)
Set Operations on $\mathcal{P}_n$

Here are some observations about set operations on $\mathcal{P}_n$.

(a) If $P, Q \in \mathcal{P}_n$, then $P \cap Q \in \mathcal{P}_n$.

Proof. Write $P = \prod_{i=1}^{n} (a_i, b_i], Q = \prod_{i=1}^{n} (c_i, d_i] \subseteq \mathbb{R}^n$. Then it is clear that

$$P \cap Q = \prod_{i=1}^{n} (a_i, b_i] \cap (c_i, d_i],$$

where each $(a_i, b_i] \cap (c_i, d_i]$ is a half-open interval. ∎

The above result can easily be extanded to a more general case. That is, given k half-open rectangles $A_1, \dots, A_k \in \mathcal{P}_n$,

$$\bigcap_{i=1}^{k} A_i$$

is also a member of $\mathcal{P}_n$.

(b) There is not much to say concerning unions.

(c) Concerning set differences, we run into an interesting situation. Certainly $\mathcal{P}_n$ is not closed under set difference, but we can still say the following: for every $P, Q \in \mathcal{P}_n$, there exists $k \in \mathbb{N}$ and disjoint $R_1, \ldots, R_k \in \mathcal{P}_n$ such that
$$\bigcup_{i=1}^{k} R_i = P \setminus Q.$$
In words, $P \setminus Q$ need not be in $\mathcal{P}_n$, but it can be *decomposed* as a finite disjoint union of members of $\mathcal{P}_n$. This is one of the properties of a *semiring of sets*. In fact, $\mathcal{P}_n$ turns out to be a semiring of sets.

The property of $\mathcal{P}_n$ as described in (a) is referred as a π-system.

Def'n. π-system
Let X be nonempty and let $\mathcal{C}$ be a collection of subsets of X. If $\mathcal{C}$ has the property that
$$A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C},$$
then we say $\mathcal{C}$ is a ***π-system***.

(1.4) Now that we introduced the idea of decomposing a half-open rectangle into a disjoint finite union of other half-open rectangles, let us show that such decompositions can always be made to consist of *small* pieces. Here, the notion of small refers to the diameter of sets considered. More precise way of stating this idea is as follows.

Proposition 1.1. *Let $P \in \mathcal{P}_n$ and let $\varepsilon > 0$. Then there exists a decomposition*
$$P = \bigcup_{i=1}^{k} P_i$$
with pairwise disjoint members $P_1, \ldots, P_k \in \mathcal{P}_n$ such that
$$\max\{\operatorname{diam}(P_1), \ldots, \operatorname{diam}(P_k)\} < \varepsilon.$$

(1.5) Observe that the above proposition does not hold if we define $\mathcal{P}_n$ to be the collection of *open* (or *closed*) rectangles in $\mathbb{R}^n$, since the union of two disjoint open (or closed) intervals is not an interval. We now proceed to show an important property of vol_n function.

Proposition 1.2. vol_n is decomposition-additive

Let $P \in \mathcal{P}_n$ and consider a decomposition
$$P = \bigcup_{i=1}^{k} P_i$$
with pairwise disjoint $P_1, \ldots, P_k \in \mathcal{P}_n$. Then it follows that
$$\sum_{i=1}^{k} \operatorname{vol}_n(P_i) = \operatorname{vol}_n(P).$$

We shall consistently refer to the above property by saying that vol_n is *decomposition-additive*, although this is not a standard term in the literature.

Def'n. Semiring of Subsets

Let X be nonempty and let $\mathcal{C} \subseteq \mathcal{P}(X)$ be a collection of subsets of X. We say $\mathcal{C}$ is a ***semiring*** of subsets of X if it satisfies the following conditions.

(a) $\varnothing \in \mathcal{C}$.

(b) $\mathcal{C}$ is a π-system.

(c) For any $A, B \in X$ there exist pairwise disjoint $C_1, \ldots, C_k \in \mathcal{C}$ such that

$$A \setminus B = \bigcup_{i=1}^{k} C_i.$$

(1.6) Ring of Subsets

If a semiring $\mathcal{C} \subseteq \mathcal{P}(X)$ has an additional property that

$$A, B \in \mathcal{C} \implies A \cup B \in \mathcal{C},$$

then we say $\mathcal{C}$ is a ***ring*** of subsets. Now, it is an easy induction argument to show that

$$C_1, \ldots, C_k \in \mathcal{C} \implies \bigcup_{i=1}^{k} C_i \in \mathcal{C},$$

if $\mathcal{C}$ is a ring of subsets. Then, it follows from (c) of the definition of subring that the set difference of any two members of $\mathcal{C}$ is also a member of $\mathcal{C}$. In other words, a ring of subsets is a nice system in the sense that closed under all three of basic set operations. However, we would not make much use of this notion, as our main example of interest is not a ring of sets.

(1.7) σ-additivity

We now turn our attention to an abstract notion of the volume function vol_n. In connection to this abstract notion of volume, we shall typically use the Greek letter μ.

Def'n. σ-additivity

Let X be nonempty and let $\mathcal{C}$ be a collection of subsets of X such that $\varnothing \in \mathcal{C}$. We say a function $\mu : \mathcal{C} \to [0, \infty)$ is ***σ-additive*** if, given pairwise disjoint $A_1, \ldots, A_k \in \mathcal{C}$ such that $\bigcup_{i=1}^{k} A_i \in \mathcal{C}$,

$$\mu\left(\bigcup_{i=1}^{k} A_i\right) = \sum\nolimits_{i=1}^{k} \mu\left(A_i\right).$$

It should be noted that the above definition does not assume that $\mathcal{C}$ is closed under finite unions. In fact, the only assumption about $\mathcal{C}$ is that $\varnothing \in \mathcal{C}$. One should read the above definition as follows: if $A_1, \ldots, A_k \in \mathcal{C}$ are pairwise disjoint and if the union of $A_1, \ldots, A_k$ is also a member of $\mathcal{C}$, then the measure of the union is the sum of the measures (we read $\mu(x)$ as the *measure* of x).

(EX 1.8)

Let X to be an infinite set (say $X = \mathbb{N}$ or $X = \mathbb{R}$), and suppose that we are given a function $w : X \to [0, \infty)$, which *assigns weight* to each point $x \in X$. Let $\mathcal{C} = \{A \subseteq X : A \text{ is finite}\}$ and let $\mu : \mathcal{C} \to [0, \infty)$ be defined by

$$\mu(A) = \sum\nolimits_{x \in A} w(x).$$

In the case $A = 0$, we invoke the usual convention that an empty sum of real numbers is equal to zero, thus giving $\mu(\varnothing) = 0$.

(EX 1.9)

Here is a pathological example. Let $X = \mathbb{R}$ and let

$$\mathcal{C} = \{(a, b) : a, b \in \mathbb{R}, a < b\} \cup \{\varnothing\}.$$

Let $\mu : \mathcal{C} \to [0, \infty)$ be such that $\mu(\varnothing) = 0$ and $\mu((a,b)) = b - a$ for every open interval $(a,b) \subseteq \mathbb{R}$. Then $\mathcal{C}$ is a π-system and μ is σ-additive. The pathological aspect of this example comes from the fact that, if we are to decompose any $A \in \mathcal{C}$, as

$$A = \bigcup_{i=1}^{k} A_i$$

for some pairwise disjoint $A_1, \ldots, A_k \in \mathcal{C}$, then there must exist $j \in \{1, \ldots, k\}$ such that $A_j = A$ and $A_i = \varnothing$ for all $i \neq j$. This is because the union of two disjoint open intervals is not interval.

(1.10) Division and Refinement

To generalize the decomposition into disjoint nonempty sets, we shall discuss the notion of division. First, fix a nonempty set X, fix $\mathcal{C}$, a collection of subsets of X with $\varnothing \in \mathcal{C}$, and fix a nonempty $A \in \mathcal{C}$.

Def'n. Division, Refinement

A ***division*** Δ of A is a set

$$\Delta = \{A_1, \ldots, A_p\},$$

where $A_1, \ldots, A_p \subseteq$ are pairwise disjoint and satisfy that

$$\bigcup_{i=1}^{p} A_i = A.$$

Moreover, let $\Gamma = \left\{B_1, \ldots, B_q\right\}$ be another division of A. We say Γ is a ***refinement*** of Δ (or alternatively, Γ ***refines*** Δ), denoted by $\Gamma \preceq \Delta$, if for every $j \in \{1, \ldots, q\}$ there exists $i \in \{1, \ldots, p\}$ such that $B_j \subseteq A_i$.

(1.11) Reindexing a Refinement

Given a division Δ and its refinement Γ, we can always reindex the elements of Γ as follows. First, write $\Delta = \{A_1, \ldots, A_p\}$ and $\Gamma = \left\{B_1, \ldots, B_q\right\}$. Then, one can reindex the elements of Γ in the form

$$\Gamma = \left\{B_{1,1}, \ldots, B_{1,q_1}, \ldots, B_{p,1}, \ldots, B_{p,q_p}\right\}$$

such that $\left\{B_{i,1}, \ldots, B_{i,q_i}\right\}$ is a division of A_1 for all $i \in \{1, \ldots, p\}$.

Proof. Exercise. ∎

Def'n. Meet of Two Refinements

Let $\Delta' = \left\{A'_1, \ldots, A'_p\right\}$ and $\Delta'' = \left\{A''_1, \ldots, A''_q\right\}$ be divisions of A. We call

$$\Delta' \wedge \Delta'' = \left\{A'_i \cap A''_j : i \in \{1, \ldots, p\}, j \in \{1, \ldots, q\}, A'_i \cap A''_j \neq \varnothing\right\}$$

the ***meet*** of Δ' and Δ''.

Proposition 1.3. Meet Is a Division

Let $\Delta' = \left\{A'_1, \ldots, A'_p\right\}$ and $\Delta'' = \left\{A''_1, \ldots, A''_q\right\}$ be divisions of A. Then $\Delta' \wedge \Delta''$ is also a division of A.

Proof. Exercise. ∎

Corollary 1.3.1.

For any two divisions Δ', Δ'' of A, there exists a division Γ which refines both Δ' and Δ''.

Integrable Functions

(1.12) The goal of this section is to describe how to integrate a function defined on rectangles in $\mathbb{R}^n$, in the sense of Riemann. But what are the properties of rectangles and vol_n function that we require to construct the integral? It turns out that the construction of the integral simply needs the π-system property.

(1.13) For convenience, fix

(a) a nonempty set X and

(b) a collection $\mathcal{C} \subseteq \mathcal{P}(X)$ of subsets of X, such that $\varnothing \in \mathcal{C}$ and such that $\mathcal{C}$ is a π-system

throughout this section. We also assume that we are given a functiohn $\mu : \mathcal{C} \to [0, \infty)$ which is decomposition-additive. We will discuss about integrability for functions of the form $f : A \to \mathbb{R}$, where $\text{domain}(f) = A \in \mathcal{C}$ is nonempty and f is assumed to be bounded (i.e. there exists $c \in \mathbb{R}$ sucht that $|f| \leq c$ for all $x \in \mathbb{R}$). Given such f, we shall consistently denote

$$\sup_B (f) = \sup\{f : x \in B\}, \inf_B (f) = \inf\{f : x \in B\}$$

for any nonempty $B \subseteq A$, and we shall also use the notation

$$\operatorname*{osc}_B (f) = \sup_B (f) - \inf_B (f)$$

to denote the *oscillation* of f. Similar to how Riemann integral is constructed for functions which domain is a subset of $\mathbb{R}$, we shall proceed in the following steps: construct

(a) upepr and lower sums;

(b) upper and lower integrals; and

(c) notion of integrability.

Def'n. Upper Sum, Lower Sum for a Function

Let $A \in \mathcal{C}$ be nonempty, let $f : A \to \mathbb{R}$ be a bounded function, and let Δ be a division of A. We define the ***upper sum*** (also known as ***upper Darboux sum***) for f and Δ to be

$$U(f, \Delta) = \sum_{i=1}^{p} \mu(A_i) \sup_{A_i} (f),$$

where $\Delta = \{A_1, \ldots, A_p\}$. Similarly, we define the ***lower sum*** (or ***lower Darboux sum***) for f and Δ to be

$$L(f, \Delta) = \sum_{i=1}^{p} \mu(A_i) \inf_{A_i} (f).$$

(1.14) It immediately follows from the definition that, given a bounded function $f : A \to \mathbb{R}$ and any division Δ of A,

$$U(f, \Delta) \geq L(f, \Delta).$$

To find the difference, one sees that the term-by-term subtraction gives

$$U(f, \Delta) - L(f, \Delta) = \sum_{i=1}^{p} \mu(A_i) \operatorname*{osc}_{A_i} (f).$$

The next proposition brings to attention the useful fact that the difference between the upper and lower sums is sure to become smaller when we refine the division we are working with.

Proposition 1.4. Refinement

Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. Let Γ, Δ be divisions of A such that Γ refines Δ. Then

$$U(f,\Gamma) \leq U(f,\Delta), L(f,\Gamma) \geq L(f,\Delta).$$

Consequently,

$$U(f,\Gamma) - L(f,\Gamma) \leq U(f,\Delta) - L(f,\Delta).$$

Proof. We are only going to show the first inequality

$$U(f,\Gamma) \leq U(f,\Delta),$$

as the proof of $L(f,\Gamma) \geq L(f,\Gamma)$ can be done analogously. To do so, first write Δ and Γ explicitely,

$$\Delta = \{A_1, \ldots, A_p\}, \Gamma = \left\{B_{11}, \ldots, B_{1q_1}, \ldots, B_{n1}, \ldots, B_{nq_n}\right\}$$

such that

$$\left\{B_{i1}, \ldots, B_{iq_i}\right\}$$

is a division of A_i, as discussed in (1.11). Then the upper sum for Γ can be written as

$$U(f,\Gamma) = \sum_{i=1}^{p} \sum_{j=1}^{q} \mu(B_{ij}) \sup_{B_{ij}}(f).$$

We also note that, for every $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, q_i\}$, we have

$$\sup_{A_i}(f) \geq \sup_{B_{ij}}(f),$$

since $B_{ij} \subseteq A_i$. Thus, one sees that

$$\begin{aligned} U(f,\Gamma) &= \sum_{i=1}^{p} \sum_{j=1}^{q_i} \mu(B_{ij}) \sup_{B_{ij}}(f) \\ &\leq \sum_{i=1}^{p} \sum_{j=1}^{q_i} \mu(B_{ij}) \sup_{A_i}(f) = \sum_{i=1}^{p} \left(\sum_{j=1}^{q_i} \mu(B_{ij})\right) \sup_{A_i}(f) \\ &= \sum_{i=1}^{p} \mu(A_i) \sup_{A_i}(f) = U(f,\Delta), \end{aligned}$$

where the second to last equality utilizes the decomposition-additivity of μ. ∎

Corollary 1.4.1.

Let $A \in \mathcal{C}$ be nonempty, let $f : A \to \mathbb{R}$ be a bounded function, and let Δ', Δ'' be any divisions of A. Then

$$U\left(f,\Delta'\right) \geq L\left(f,\Delta''\right).$$

In words, what Corollary 1.4.1 means is that the lower sum for any division of domain(f) can never exceed the upper sum for any division of domain(f), even when we assume no relation between them.

Proof. Let Γ be the common refinement of Δ' and Δ'' as discussed in Corollary 1.3.1. Then one obtains

$$U\left(f,\Delta'\right) \leq U(f,\Gamma) \leq L(f,\Gamma) \leq L\left(f,\Delta''\right).$$ ∎

(1.15) Now we are ready to define upper and lower integrals. But first, consider the following proposition.

Proposition 1.5. Properties of Upper and Lower Integrals

Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded.

(a) *The set*
$$T = \{U(f,\Delta) : \Delta \text{ is a division of } A\}$$
is lower bounded, and we denote
$$\overline{\int_A} f = \inf(T).$$

(b) *The set*
$$S = \{L(f,\Delta) : \Delta \text{ is a division of } A\}$$
is upper bounded, and we denote
$$\underline{\int_A} f = \sup(T).$$

(c) *One has*
$$\underline{\int_A} f \leq \overline{\int_A} f.$$

Proof. To verify (a), observe that $L(f,\Delta')$ for any division Δ' of A is a lower bound of T, so by the completeness of $\mathbb{R}$, one can obtain
$$\overline{\int_A} f = \inf(f).$$
We verify (b) and (c) simultaneously. Observe that $\overline{\int_A} f$ is an upper bound for S, for, if
$$\overline{\int_A} f < L(f,\Delta'')$$
for any division Δ'' of A, we have
$$\overline{\int_A} f < L(f,\Delta'') \leq U(f,\Delta')$$
for any division Δ' of A, so $\overline{\int_A} f$ is not the infimum of T, which is a contraidction. Therefore, $\underline{\int_A} f = \sup(S)$ is well-defined, and, in particular, one sees that
$$\underline{\int_A} f \leq \overline{\int_A} f,$$
as required. ∎

Def'n. Upper Integral, Lower Integral of a Function

Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. We define the ***upper integral*** of f on A to be
$$\overline{\int_A} f = \inf\{U(f,\Delta) : \Delta \text{ is a division of } A\}$$
and we define the ***lower integral*** of f on A to be
$$\underline{\int_A} f = \sup\{U(f,\Delta) : \Delta \text{ is a division of } A\}.$$

We can now state what integrability is.

Def'n. Integrable Function
Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. We say f is ***integrable*** if

$$\underline{\int_A} f = \overline{\int_A} f.$$

If f is integrable, we define the ***integral*** of f to be

$$\int_A f = \underline{\int_A} f = \overline{\int_A} f.$$

(1.16) We now proceed to prove a very useful criterion for integrability which will be used over and over when one wants to prove a function is integrable.

Theorem 1.6. ε-Δ Criterion of Integrability

Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. Then the following are equivalent.

(a) f is integrable.

(b) For all $\varepsilon > 0$, there exists a division Δ of A such that

$$U(f,\Delta) - L(f,\Delta) < \varepsilon.$$

(c) There exists a sequence (Δ_k) of divisions of A such that

$$\lim_{k\to\infty} U(f,\Delta_k) - L(f,\Delta_k) = 0.$$

One sees that Theorem 1.6 is a generalized version of the Cauchy criterion for integrability.

Proof. We shall prove the cycle (a) $\implies$ (b) $\implies$ (c) $\implies$ (a). To prove (a) $\implies$ (b), we show that there exists a division Δ' of A such that

$$I \le U\left(f,\Delta'\right) < I + \frac{\varepsilon}{2}$$

It is clear that $I \le U(f,\Delta')$, since I is the infimum of the set of upper sums. Moreover, if there does not exist such Δ', then $I + \frac{\varepsilon}{2}$ is a lower bound for the set of upper sums, so we have a contradiction. We may similarly show that

$$I - \frac{\varepsilon}{2} < L\left(f,\Delta''\right) \le I$$

for some division Δ'' of A. Now, let Δ be a common refinement of Δ' and Δ''. Then,

$$I - \frac{\varepsilon}{2} < L\left(f,\Delta''\right) \le L(f,\Delta) \le U(f,\Delta) \le U\left(f,\Delta'\right) < I + \frac{\varepsilon}{2},$$

so we have

$$U(f,\Delta) - L(f,\Delta) < \left(I + \frac{\varepsilon}{2}\right) - \left(I - \frac{\varepsilon}{2}\right) = \varepsilon.$$

To prove (b) $\implies$ (c), define

$$\varepsilon = \frac{1}{k}$$

for all $k \in \mathbb{N}$. Then by (b), there exists Δ_k such that

$$U(f,\Delta_k) - L(f,\Delta_k) < \varepsilon = \frac{1}{k}$$

for all $k \in \mathbb{N}$. Since $U(f,\Delta_k) - L(f,\Delta_k)$ is lower bounded by 0, we have

$$\lim_{k\to\infty} U(f,\Delta_k) - L(f,\Delta_k) = 0$$

by the squeeze theorem. To prove (c) $\implies$ (a), let (Δ_k) be a sequence of divisions of A as described in (c). Then, by the definition of the upper and lower sums,

$$L(f,\Delta_k) \leq \underline{\int_A} f \leq \overline{\int_A} f \leq U(f,\Delta_k)$$

for all $k \in \mathbb{N}$. From the above inequalities one obtains

$$\overline{\int_A} f - \underline{\int_A} f \leq U(f,\Delta_k) - L(f,\Delta_k)$$

for all $k \in \mathbb{N}$. But

$$\lim_{k\to\infty} U(f,\Delta_k) - L(f,\Delta_k) = 0$$

and $\overline{\int_A} f - \underline{\int_A} f$ is lower bounded by 0, so by the squeeze theorem

$$\overline{\int_A} f - \underline{\int_A} f = 0,$$

which means f is integrable. ∎

Def'n. Witnessing Sequence of Divisions
Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. Suppose that f is integrable. A sequence of divisions (Δ_k) as described in (c) of Proposition 1.6 is called a ***witnessing sequence*** of divisions for the integrability of f.

(EX 1.17) here is a trivial example of applying ε-Δ criterion. Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. If there exists a partition Δ such that

$$U(f,\Delta) = L(f,\Delta),$$

then it seems intuitive to think that f is integrable. But why? The answer is that, we may define our witnessing sequence of integrability by

$$\Delta_k = \Delta$$

for all $k \in \mathbb{N}$. Then of course $\lim_{k\to\infty} U(f,\Delta_k) - L(f,\Delta_k) = 0$, which means f is integrable, and, in particular,

$$\int_A f = U(f,\Delta_k) = L(f,\Delta).$$

In order to make this example less trivial, consider the following proposition.

Proposition 1.7. *Let $A \in \mathcal{C}$ be nonempty and let $f : A \to \mathbb{R}$ be bounded. Suppose that f is piecewise constant, in a sense that there exists a division $\Delta = \{A_1, \ldots, A_p\}$ of A and some constants $c_1, \ldots, c_p \in \mathbb{R}$ such that*

$$x \in A_i \implies f(x) = c_i$$

for all $x \in A$ and $i \in \{1, \ldots, p\}$. Then it follows that f is integrable on A, with

$$\int_A f = \sum_{i=1}^{p} c_i \mu(A_i).$$

Proof. Observe that

$$\sup_{A_i}(f) = \inf A_i(f) = c_i$$

for all $i \in \{1, \ldots, p\}$. ∎

Linearity of the Integral

(1.18) Throughout this section, we fix

(a) a nonempty set X;

(b) a collection $\mathcal{C}$ of subsets of X such that $\varnothing \in \mathcal{C}$ and such that $\mathcal{C}$ is a π-system; and

(c) a decomposition-additive function $\mu : \mathcal{C} \to [0, \infty)$.

Moreover, for any nonempty $A \in \mathcal{C}$, we denote

$$F_b(A, \mathbb{R}) = \{ f : A \to \mathbb{R} : f \text{ is bounded} \},$$

the set of bounded functions, and

$$\mathrm{Int}_b(A, \mathbb{R}) = \{ f : A \to \mathbb{R} : f \text{ is integrable} \},$$

the set of integrable functions.

(1.19)
Remarks on $\mathrm{Int}_b(A, \mathbb{R})$

Here are some remarks regarding $\mathrm{Int}_b(A, \mathbb{R})$.

(a) We are interested in discovering the operations that can be done on $\mathrm{Int}_b(A, \mathbb{R})$. But before that, we first have to see that $\mathrm{Int}_b(A, \mathbb{R})$ is nonempty. Why so? At the very least, one knows that any constant function $f = c$ is integrable (i.e. $f \in \mathrm{Int}_b(A, \mathbb{R})$) and that

$$\int_A f = c\mu(A).$$

Notice that the justification for the above result comes from Proposition 1.7, since a constant function is piecewise constant.

(b) The subscript b is there to indicate that every $f \in \mathrm{Int}_b(A, \mathbb{R})$ is bounded.

(EX 1.20)
Relationship between $F_b(A, \mathbb{R})$ and $\mathrm{Int}_b(A, \mathbb{R})$

Verify the following:

(a) When A is finite, then $\mathrm{Int}_b(A, \mathbb{R})$ is as big as it can be. That is, $\mathrm{Int}_b(A, \mathbb{R}) = F_b(A, \mathbb{R})$.

(b) When A is an open bounded interval on $\mathbb{R}$, then $\mathrm{Int}_b(A, \mathbb{R})$ is as small as it can be. That is,

$$\mathrm{Int}_b(A, \mathbb{R}) = \{ f : A \to \mathbb{R} : f \text{ is constant} \}.$$

(1.21)
Vector Space Structure of $F_b(A, \mathbb{R})$

Now, we turn our attention to the main topic of this section, which is to look at the linear combinations of integrable functions. First observe that $F_b(A, \mathbb{R})$ is a vector space under the following operations:

(a) *addition*: For any $f, g \in F_b(A, \mathbb{R})$, define a new function $f + g : A \to \mathbb{R}$ by the mapping

$$x \mapsto f(x) + g(x)$$

for all $x \in A$. Then $f + g$ is also a bounded function, since, if $O_f, O_g \subseteq \mathbb{R}$ are open intervals containing $\{f : x \in A\}, \{g : x \in A\}$, respectively (which exist since f and g are bounded), then one sees

$$\{f + g : x \in A\} \subseteq (\min(O_f \cup O_g), \max(O_f \cup O_g)).$$

(b) *scalar multiplication*: For any $f \in F_b(A,\mathbb{R})$ and $\alpha \in \mathbb{R}$, define a new function $\alpha f : A \to \mathbb{R}$ by

$$x \mapsto \alpha f(x)$$

for all $x \in A$.

The verification that $F_b(A,\mathbb{R})$ satisfies all eight vector space axioms is a routine calculation. For convenience, we shall denote the *0 element* of $F_b(A,\mathbb{R})$ by $\underline{0} : A \to \mathbb{R}$. Now that $F_b(A,\mathbb{R})$ is a vector space, we may bring in notions from linear algebra (e.g. linear combination, linear independence, linear transformation, ...). In particular, the notion that we are going to use is the notion of subspace, as it turns out that $\mathrm{Int}_b(A,\mathbb{R})$ is a subspace of $F_b(A,R)$, as the next theorem shows (although it does not prove that $\underline{0} \in \mathrm{Int}_b(A,\mathbb{R})$, Proposition 1.7 provides this result).

Theorem 1.8. Linearity of Integral

Let $A \in \mathcal{C}$ be nonempty and let $f,g \in \mathrm{Int}_b(A,\mathbb{R})$. Then for any $\alpha \in \mathbb{R}$, we have $\alpha f + g \in \mathrm{Int}_b(A,\mathbb{R})$ with

$$\int_A \alpha f + g = \alpha \int_A f + \int_A g.$$

(1.22) Theorem 1.8 does more than what we claimed in (1.21). That is, not only it shows that a linear combination of integrable functions is again integrable (thereby showing that $\mathrm{Int}_b(A,\mathbb{R})$ is indeed a subspace of $F_b(A,\mathbb{R})$), but it also states that the integration operation is *linear*. That is, if we define a function $\varphi : \mathrm{Int}_b(A,\mathbb{R}) \to \mathbb{R}$ by the mapping

$$f \mapsto \int_A f,$$

then φ is a linear transformation. We shall prove Theorem 1.8 in few separate steps.

Lemma 1.8.1.

Let $A \in \mathcal{C}$ be nonempty and let $f,g \in F_b(A,\mathbb{R})$. Then

$$\overline{\int_A} f + g \leq \overline{\int_A} f + \overline{\int_A} g, \quad \underline{\int_A} f + g \geq \underline{\int_A} f + \underline{\int_A} g.$$

Proof. We shall only prove the first inequality, since the verification for the second inequality can be done in a completely analogous way. To do so, we claim that, for any division $\Delta = \{A_1,\ldots,A_p\}$ of A,

$$U(f+g,\Delta) \leq U(f,\Delta) + U(g,\Delta).$$

To verify this, observe that we have

$$\sup_{A_i}(f+g) \leq \sup_{A_i}(f) + \sup_{A_i}(g)$$

for all $i \in \{1,\ldots,p\}$, since

$$(f+g)(x) = f(x) + g(x) \leq \sup_{A_i}(f) + \sup_{A_i}(g),$$

so $\sup_{A_i}(f) + \sup_{A_i}(g)$ is an upper bound for $f+g$. Therefore, the least upper bound $\sup_{A_i}(f+g)$ satisfies

$$\sup_{A_i}(f+g) \leq \sup_{A_i}(f) + \sup_{A_i}(g).$$

Therefore,

$$\begin{aligned} U(f+g,\Delta) &= \sum_{i=1}^{p} \mu(A_i) \sup_{A_i}(f) \leq \sum_{i=1}^{p} \mu(A_i) \left(\sup_{A_i}(f) + \sup_{A_i}(g) \right) \\ &= \sum_{i=1}^{p} \mu(A_i) \sup_{A_i}(f) + \sum_{i=1}^{p} \mu(A_i) \sup_{A_i}(g) = U(f,\Delta) + U(g,\Delta), \end{aligned}$$

as claimed. Now the next claim is that, for every $k \in \mathbb{N}$, we can find divisions Δ'_k, Δ''_k of A such that

$$U\left(f, \Delta'_k\right) < \frac{1}{k} + \overline{\int_A} f, U\left(g, \Delta''\right) < \frac{1}{k} + \overline{\int_A} g.$$

The verification of the second inequality is can be done analogously, so we shall only prove the first one. Observe that tha inequality immediately follows from the definition of upper integral, since if there exists some $k \in \mathbb{N}$ such that, for any division Δ'_k

$$U\left(f, \Delta'_k\right) \geq \frac{1}{k} + \overline{\int_A} f,$$

then $\frac{1}{k} + \overline{\int_A} f$ is also a lower bound for the set of upper sums, which contradicts the fact that $\overline{\int_A} f$ is the greatest lower bound. Therefore the inequality follows. Now that we verified two claims, we pick a division Δ_k of A such that Δ_k refines Δ'_k and Δ''_k from the second claim. Then by using the first claim, we obtain

$$\begin{aligned}\overline{\int_A} f + g &\leq U\left(f+g, \Delta_k\right) \leq U\left(f, \Delta_k\right) + U\left(g, \Delta_k\right) \leq U\left(f, \Delta''\right) + U\left(g, \Delta''_k\right) \\ &< \left(\frac{1}{k} + \overline{\int_A} f\right) + \left(\frac{1}{k} + \overline{\int_A} g\right) = \frac{2}{k} + \overline{\int_A} f + \overline{\int_A} g.\end{aligned}$$

So we find that

$$\overline{\int_A} f + g < \frac{2}{k} + \overline{\int_A} f + \overline{\int_A} g,$$

so by letting $k \to \infty$,

$$\overline{\int_A} f + g = \lim_{k\to\infty} \overline{\int_A} f + g \leq \lim_{k\to\infty} \frac{2}{k} + \overline{\int_A} f + \overline{\int_A} g = \overline{\int_A} f + \overline{\int_A} g.$$

∎

Lemma 1.8.2. Integration Preserves Addition

Let $A \in \mathcal{C}$ be nonempty and let $f, g \in \mathrm{Int}_b(A, \mathbb{R})$. Then $f + g \in \mathrm{Int}_b(A, \mathbb{R})$ and

$$\int_A f + g = \int_A f + \int_A g. \qquad [1.1]$$

Proof. Observe the following: let $S = \overline{\int_A} f + g$,

$$S = \overline{\int_A} f + g \leq \overline{\int_A} f + \overline{\int_A} g = \int_A f + \int_A g = \underline{\int_A} f + \underline{\int_A} g \leq \underline{\int_A} f + g \leq \overline{\int_A} f + g = S.$$

Since we started and ended with S, everything in between must be equal, and in particular [1.1] holds. ∎

Lemma 1.8.3.

Let $A \in \mathcal{C}$ be nonempty, let $f \in F_b(A, \mathbb{R})$, and let $\alpha \in \mathbb{R}$. Then the following statements hold.

(a) If $\alpha > 0$, then $\overline{\int_A} \alpha f = \alpha \overline{\int_A} f$ and $\underline{\int_A} \alpha f = \alpha \underline{\int_A} f$.

(b) If $\alpha < 0$, then $\overline{\int_A} \alpha f = \alpha \underline{\int_A} f$ and $\underline{\int_A} \alpha f = \alpha \overline{\int_A} f$.

(c) If $\alpha = 0$, then $\overline{\int_A} \alpha f = \underline{\int_A} \alpha f = 0$.

Proof. (c) is trivial. We claim that, for any arbitrary subset $S \subseteq \mathbb{R}$ and nonzero $c \in \mathbb{R}$,

$$c\sup(S) = \begin{cases} \sup(cS) & \text{if } c > 0 \\ \inf(cS) & \text{otherwise} \end{cases}, c\inf(S) = \begin{cases} \inf(cS) & \text{if } c > 0 \\ \sup(cS) & \text{otherwise} \end{cases}.$$

We shall only verify the first equality when $c > 0$. Suppose, for the sake of contradiction, that $c\sup(S) \neq \sup(cS)$. If $\sup(cS) > c\sup(S)$, then $\frac{1}{c}\sup(cS) > \sup(cS)$, so it follows that, for any $s \in S$,

$$s < y < \frac{1}{c}\sup(cS)$$

for some $y \in \mathbb{R}$. This is a contradiction, since multiplying by c shows that

$$cs < cy < \sup(cS).$$

for any $s \in S$, which means $\sup(cS)$ is not the supremum of cS. Conversely, if we assume that $\sup(cS) < c\sup(S)$, then for any $s \in S$, there exists $y \in \mathbb{R}$ such that

$$cs < y < c\sup(S),$$

so by dividing c,

$$s < \frac{y}{c} < \sup(S)$$

for any $s \in S$, which is a contradicion. By using this result, one can verify both (a) and (b). ∎

Lemma 1.8.4. Integration Preserves Scalar Multiplication

Let $A \in \mathcal{C}$ be nonempty and let $f \in \operatorname{Int}_b(A, \mathbb{R})$. Then, for any $\alpha \in \mathbb{R}$,

$$\int_A \alpha f = \alpha \int_A f.$$

Proof. If $\alpha = 0$, then $\int_A \alpha f = \int_A \underline{0} = 0 = 0 \int_A f$. If $\alpha > 0$, then

$$\overline{\int_A} \alpha f = \alpha \overline{\int_A} f = \alpha \underline{\int_A} f = \underline{\int_A} \alpha f.$$

and if $\alpha < 0$, then

$$\overline{\int_A} \alpha f = \alpha \underline{\int_A} f = \alpha \overline{\int_A} f = \underline{\int_A} \alpha f.$$

So αf is integrable and the desired inequality holds. ∎

Proof of Theorem 1.8. Theorem 1.8 can be proved by combining Lemma 1.8.2 and 1.8.4. ∎

(EX 1.23) Glueing Together Several Integrable Functions

Let $A \in \mathcal{C}$ be nonempty and let $\Delta = \{A_1, \ldots, A_p\}$ be a division of A. Suppose that integrable functions

$$f_1 \in \operatorname{Int}_b(A_1, \mathbb{R}), \ldots, f_p \in \operatorname{Int}_b(A_p, \mathbb{R})$$

are given. Let $f : A \to \mathbb{R}$ be defined by *glueing together* $f_1, \ldots, f_p$. That is, for any $x \in A$,

$$x \in A_i \implies f(x) f_i(x).$$

Prove that $f \in \operatorname{Int}_b(A, \mathbb{R})$, and that

$$\int_A f = \sum_{i=1}^{p} \int_{A_i} f_i.$$

Operations on Integrable Functions

(1.24) We continue to use the same framework from previous chapters. That is:

(a) X is a nonempty set;

(b) $\mathcal{C} \subseteq (X)$ is a collection of subsets of X such that

  (i) $\varnothing \in \mathcal{C}$; and

  (ii) $\mathcal{C}$ is a π-system; and

(c) $\mu : \mathcal{C} \to [0, \infty)$ is a decomposition-additive function.

Additionally, we also fix a nonempty set $A \in \mathcal{C}$, and we consider the space of integrable functions

$$\text{Int}_b(A, \mathbb{R}) = \{ f : A \to \mathbb{R} : f \text{ is integrable} \},$$

which is a subspace of $F_b(A, \mathbb{R})$, the set of bounded functions. What we desire to find out is that if there are other operations (i.e. besides linear combinations) under which $\text{Int}_b(A, \mathbb{R})$ is closed. Similar to the previous section, we first take a look at the operations that naturally arises for functions in $F_b(A, \mathbb{R})$, and show that $\text{Int}_b(A, \mathbb{R})$ is closed under such operations.

(1.25) Operations on $F_b(A, \mathbb{R})$

Here are three operations under which $F_b(A, \mathbb{R})$ is closed:

(a) *product*: For any $f, g \in F_b(A, \mathbb{R})$, we define their *product* to be a function $fg : A \to \mathbb{R}$ by the mapping

$$x \mapsto f(x)\, g(x)$$

for all $x \in A$. Then $fg \in F_b(A, \mathbb{R})$. In other words, $F_b(A, \mathbb{R})$ is an *algebra* of functions.

Proof. Exercise. ∎

(b) *maximum* and *minimum*: For any $f, g \in F_b(A, \mathbb{R})$, we define the *maximum* of f and g, $f \vee g : A \to \mathbb{R}$, by

$$x \mapsto \max(f(x), g(x))$$

and the *minimum* $f \wedge g : A \to \mathbb{R}$ by

$$x \mapsto \min(f(x), g(x)).$$

Then $f \vee g, f \wedge g \in F_b(A, \mathbb{R})$. In other words, $F_b(A, \mathbb{R})$ is a *lattice* of functions.

Proof. Exercise. ∎

(c) *absolute value*: For any $f \in F_b(A, \mathbb{R})$, we define its *absolute value* $|f| : A \to |R$ by

$$x \mapsto |f(x)|.$$

Then it is immediate that $|f| \in F_b(A, \mathbb{R})$.

It turns out that $\text{Int}_b(A, \mathbb{R})$ is closed under the operations described above, as the following proposition shows.

Proposition 1.9. *Let $f,g \in \mathrm{Int}_b(A,\mathbb{R})$. Then*

(a) $fg \in \mathrm{Int}_b(A,\mathbb{R})$;

(b) $f \vee g, f \wedge g \in \mathrm{Int}_b(A,\mathbb{R})$; *and*

(c) $|f|, f^2 \in \mathrm{Int}_b(A,\mathbb{R})$.

In fact, one sees that (c) is a consequence of (a) and (b). However, as we are going to prove (a) and (b) using (c), we shall not utilize this fact.

(1.26) Lipschitz Function

It turns out that, in proving (c) of Proposition 1.9, the notion of Lipschitz function is handy.

Def'n. Lipschitz Function
Let $\alpha,\beta \in \mathbb{R}$ with $\alpha \leq \beta$ and let $f:[\alpha,\beta] \to \mathbb{R}$. Then we say f is ***c-Lipschitz*** if nonnegative $c \in \mathbb{R}$ is such that

$$|f(x) - f(y)| \leq c|x-y|$$

for all $x,y \in [\alpha,\beta]$. We often say that f is ***Lipschitz*** if f is c-Lipschitz for some nonnegative $c \in \mathbb{R}$. That is, we use this terminology when we do not want to insist on a particular value of c.

Proposition 1.10. *Let $f : A \to \mathbb{R}$ be a bounded function, and let $\alpha,\beta \in \mathbb{R}$ be suchthat*

$$\alpha \leq f \leq \beta$$

for all $x \in A$. Let $u:[\alpha,\beta] \to \mathbb{R}$ be a Lipschitz function, and we consider $g : A \to \mathbb{R}$ defined by

$$x \mapsto u(f(x)).$$

Then the following hold.

(a) g is bounded.

(b) Let $c \geq 0$ be such that u is c-Lipschitz. Then for any nonempty $B \subseteq A$, one has

$$\operatorname*{osc}_B(g) \leq c \operatorname*{osc}_B(f).$$

Proof. To prove (a), we fix a point $x_0 \in A$ and let nonnegative $c \in \mathbb{R}$ be such that u is c-Lipschitz. Then for any $x \in A$, one obtains

$$|g(x) - g(x_0)| = |u(f(x)) - u(f(x_0))| \leq c|f(x) - f(x_0)| \leq c(\beta - \alpha),$$

where the first inequality follows from c-Lipschitz property of u. Thus,

$$|g(x)| = |g(x_0) + (g(x) - g(x_0))| \leq |g(x_0)| + |g(x) - g(x_0)| \leq |g(x_0)| + c(\beta - \alpha),$$

which verifies that g is bounded. For (b), fix nonempty $B \subseteq A$. Observe that it is equivalent to prove that

$$|g(x) - g(y)| \leq c \operatorname*{osc}_B(f)$$

for all $x,y \in B$, since

$$\operatorname*{osc}_B(g) = \sup\{|g(x) - g(y)| : x,y \in B\}. \qquad [1.2]$$

But observe that

$$|g(x) - g(y)| \leq c|f(x) - f(y)| \leq c \operatorname*{osc}_B(f),$$

for all $x,y \in B$, since $\mathrm{osc}_B(f)$ is the supremum, similar to [1.1]. ∎

Proposition 1.11. *Let $f \in \operatorname{Int}_b(A,\mathbb{R})$ and let $\alpha,\beta \in \mathbb{R}$ be such that*

$$\alpha \leq f \leq \beta$$

for all $x \in A$. Suppose that $u : [\alpha,\beta]$ is a Lipschitz function. Then $g : A \to \mathbb{R}$ defined by

$$g = u(f)$$

is integrable, $g \in \operatorname{Int}_b(A,\mathbb{R})$.

Proof. By Proposition 1.10, we have that $g \in F_b(A,\mathbb{R})$, so it suffices to show that g satisfies $\varepsilon-\Delta$ criterion of integrability (Theorem 1.6). Now, let $(\Delta_k)_{k=1}^{\infty}$ be a witnessing sequence of divisions of A for f. That is, one has

$$\lim_{k\to\infty} U(f,\Delta_k) - L(f,\Delta_k) = 0.$$

We claim that (Δ_k) is also a witnessing sequence for g. To verify this, observe that (write $\Delta_k = \{A_{k1},\ldots,A_{kn_k}\}$ for each $k \in \mathbb{N}$)

$$\begin{aligned}
\lim_{k\to\infty} U(g,\Delta_k) - L(g,\Delta_k) &= \lim_{k\to\infty} \sum\nolimits_{i=1}^{n_k} \mu(A_{ki}) \left(\sup_{A_{ki}}(g) - \inf_{A_{ki}}(g) \right) \\
&= \lim_{k\to\infty} \sum\nolimits_{i=1}^{n_k} \mu(A_{ki}) \sup\{|g(x)-g(y)| : x,y \in A_{ki}\} \\
&= \lim_{k\to\infty} \sum\nolimits_{i=1}^{n_k} \mu(A_{ki}) \operatorname*{osc}_{A_{ki}}(g) \\
&\leq \lim_{k\to\infty} \sum\nolimits_{i=1}^{n_k} \mu(A_{ki})\, c \operatorname*{osc}_{A_{ki}}(f) = c \lim_{k\to\infty} \sum\nolimits_{i=1}^{n_k} \mu(A_{ki}) \left(\sup_{A_{ki}}(f) - \inf_{A_{ki}}(f) \right) = 0.
\end{aligned}$$

Thus by the squeeze theorem,

$$\lim_{k\to\infty} U(g,\Delta_k) - L(g,\Delta_k) = 0,$$

as required. ∎

Proof of Proposition 1.9. We verify $f^2 \in \operatorname{Int}_b(A,\mathbb{R})$ first. Since $f \in \operatorname{Int}_b(A,\mathbb{R}) \subseteq F_b(A,\mathbb{R})$, choose $\alpha,\beta \in \mathbb{R}$ such that

$$\alpha \leq f \leq \beta$$

for all $x \in \mathbb{R}$. Now define $u : [\alpha,\beta] \to \mathbb{R}$ by

$$x \mapsto x^2.$$

We claim that u is Lipschitz. To verify this, let

$$c = 2\max(|\alpha|,|\beta|).$$

Then for any $x,y \in [\alpha,\beta]$,

$$|f(x) - f(y)| = \left|x^2 - y^2\right| = |x-y|\,|x+y| \leq |x-y|\,(|x|+|y|) \leq c\,|x-y|,$$

as required. To show that $|f| \in \operatorname{Int}_b(A,\mathbb{R})$, define $u : [\alpha,\beta] \to \mathbb{R}$ by

$$x \to |x|.$$

To verify that u is Lipschitz, observe that

$$|f(x) - f(y)| = ||x| - |y|| \leq ||x-y|| = |x-y|$$

so u is 1-Lipschitz, and we are done with (c). Then (a) easily follows from the equation

$$fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right).$$

For (b), we use the identities

$$\begin{cases} f \vee g &= \frac{1}{2}(f+g+|f-g|) \\ f \wedge g &= \frac{1}{2}(f+g-|f-g|) \end{cases} \tag{1.3}$$

to obtain the desired result from (c). ∎

(1.27) Although the identities [1.2] are useful, where do they come from? Indeed, one can easily verify that (or at least make sense in mind that)

$$\begin{cases} f \vee g + f \wedge g &= f+g \\ f \vee g - f \wedge g &= |f-g| \end{cases}$$

solving which in terms of $f \vee g$ and $f \wedge g$ gives [1.2].

Cauchy Sequences and Bolzano-Weierstrass Theorem for Euclidean Spaces

(1.28) The main purpose of this section is to generalize the notion of Cauchy sequences and Bolzano-Weierstrass theorem into Euclidean spaces. But first, recall the following definitions and theorems.

> **Def'n. Limit** of a Sequence in $\mathbb{R}$
> Let $(a_n) : \mathbb{N} \to \mathbb{R}$ be a sequence. We say (a_n) is ***convergent*** if
>
> $$\exists L \in \mathbb{R} \forall \varepsilon > 0 \exists N \in \mathbb{N} [n \geq N \implies |L - a_n| < \varepsilon].$$
>
> If so, such L is unique and we say L is the ***limit*** of (a_n), denoted by
>
> $$\lim_{n\to\infty} a_n = L.$$

> **Def'n. Cauchy** Sequence in $\mathbb{R}$
> Let $(a_n) : \mathbb{N} \to \mathbb{R}$ be a sequence. We say (a_n) is ***Cauchy*** if
>
> $$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n, m \geq N [|a_n - a_m| < \varepsilon].$$

> **Def'n. Bounded** Sequence in $\mathbb{R}$
> Let $(a_n) : \mathbb{N} \to \mathbb{R}$ be a sequence. We say (a_n) is ***bounded*** if
>
> $$\exists M \geq 0 \forall n \in \mathbb{N} [M > |a_n|].$$

Here are two of the most important theorems in analysis on $\mathbb{R}$.

Theorem 1.12. Completeness of $\mathbb{R}$

Let $(a_n) : \mathbb{N} \to \mathbb{R}$ be a sequence. Then (a_n) is convergent if and only if (a_n) is Cauchy.

Theorem 1.13. Bolzano-Weierstrass Theorem

Let $(a_n) : \mathbb{N} \to \mathbb{R}$ be bounded. Then there exists a subsequene $\left(a_{n_p}\right)_{p=1}^{\infty}$ which is convergent. In other words, one can find indices $k(1), \ldots \in \mathbb{N}$ with

$$k(1) < k(2) < \cdots$$

such that $\left(a_{k(p)}\right)_{p=1}^{\infty}$ is convergent.

The above defintions and results can be adjusted for $\mathbb{R}^n$ as follows.

Def'n. Limit of a Sequence in $\mathbb{R}^n$

Let $(a_k)_{k=1}^{\infty} : \mathbb{N} \to \mathbb{R}^n$ be a sequence. We say (a_k) is ***convergent*** if

$$\exists a \in \mathbb{R}^n \forall \varepsilon > 0 \exists N \in \mathbb{N} [k \geq N \implies \|a_k - a\| < \varepsilon].$$

If such a exists, we say a is the ***limit*** of (a_k) and denote

$$\lim_{n\to\infty} a_k = a.$$

Def'n. Cauchy Sequence in $\mathbb{R}^n$

We say a sequence $(a_k)_{k=1}^{\infty}$ is ***Cauchy*** if

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \left[i, j \geq N \implies \left\|a_i - a_j\right\| < \varepsilon\right].$$

Def'n. Bounded Sequence in $\mathbb{R}^n$

We say a sequence $(a_k)_{k=1}^{\infty}$ is ***bounded*** if

$$\exists r \geq 0 \forall k \in \mathbb{N} [\|a_k\| < r].$$

Def'n. Component Sequences of a Sequence in $\mathbb{R}^n$

Let $(a_n)_{k=1}^{\infty}$ be a sequence. Then we may write $a_k = \left(a_k^{(1)}, \ldots, a_k^{(n)}\right) \in \mathbb{R}^n$ for all $k \in \mathbb{N}$. In such way, we obtain n sequences

$$\left(a_k^{(1)}\right)_{k=1}^{\infty}, \ldots, \left(a_k^{(n)}\right)_{k=1}^{\infty} : \mathbb{N} \to \mathbb{R},$$

which are called the ***component seqences*** of (a_k).

Proposition 1.14. Component Sequence

A sequence in $\mathbb{R}^n$ is

(a) convergent if and only if each component sequence is convergent;

(b) Cauchy if and only if each component sequence is Cauchy; and

(c) bounded if and only if each component sequence is bounded.

Proof. Let $(a_k)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}^n$ and let

$$\left(a_k^{(1)}\right)_{k=1}^{\infty}, \ldots, \left(a_k^{(n)}\right)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}$$

be component sequences. Since (b) and (c) can be proven similarly to the following proof of (a), we shall only verify (a). Suppose that there exists a divergent component sequecne, say $\left(a_k^{(i)}\right)_{k=1}^{\infty}$. Then

$$\forall a^{(i)} \in \mathbb{R} \exists \varepsilon > 0 \forall N \in \mathbb{N} \left[k \geq N \wedge \left|a_k^{(i)} - a^{(i)}\right| \geq \varepsilon\right],$$

which is equivalent to

$$\forall a^{(i)} \in \mathbb{R} \exists \varepsilon > 0 \forall k \in \mathbb{N} \left[\left|a_k^{(i)} - a^{(i)}\right| \geq \varepsilon\right],$$

Now, fix $a \in \mathbb{R}^n$. Observe that

$$\|a - a_k\| = \left(\sum_{i=1}^{n} \left(a_k^{(i)} - a^{(i)}\right)^2\right)^{\frac{1}{2}} \leq \max_{1 \leq i \leq n} \left|a_k^{(i)} - a^{(i)}\right| \geq \varepsilon,$$

so (a_k) is divergent. Conversely, suppose that each component sequence converges to $a^{(i)}$. Then

$$\forall \frac{\varepsilon}{n} > 0 \exists N \in \mathbb{N} \left[k \geq N \implies \left|a_k^{(i)} - a^{(i)}\right| < \varepsilon\right],$$

so

$$\|a - a_k\| = \left(\sum_{i=1}^{n} \left(a_k^{(i)} - a^{(i)}\right)^2\right)^{\frac{1}{2}} \leq \sum_{i=1}^{n} \left|a_k^{(i)} - a^{(i)}\right| < \sum_{i=1}^{n} \frac{\varepsilon}{n} = \varepsilon.$$

Thus (a_k) is convergent, as required. ∎

Theorem 1.15. Completeness of $\mathbb{R}^n$

Let $(a_k)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}^n$. Then (a_k) is convergent if and only if (a_k) is Cauchy.

Proof. The result follows from (a) and (b) of Proposition 1.14 and the completeness of $\mathbb{R}$. ∎

Theorem 1.16. Bolzano-Weierstrass Theorem in $\mathbb{R}^n$

Let $(a_k)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}^n$ be bounded. Then (a_k) has a convergent subsequence.

Proof. We proceed inductively on n. We have the result for $n = 1$. Now suppose that the result holds for some $n = m \in \mathbb{N}$ and fix a sequecne

$$(a_k)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}^{n+1}.$$

Write

$$a_k = \left(a_k^{(1)}, \ldots, a_k^{(n+1)}\right)$$

for each $k \in \mathbb{N}$ to obtain component sequences

$$\left(a_k^{(1)}\right)_{k=1}^{\infty}, \ldots, \left(a_k^{(n+1)}\right)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}.$$

By taking the first n component sequences, we obtain another sequence

$$(a_k')_{k=1}^{\infty} = \left(\left(a_k^{(1)}, \ldots, a_k^{(n)}\right)\right)_{k=1}^{\infty}: \mathbb{N} \to \mathbb{R}^n.$$

By the induction hypothesis, $\left(a'_k\right)$ has a convergent subsequence, say $\left(a'_{k_s}\right)_{s=1}^{\infty}$. By using the indices $k_1, k_2, \ldots \in \mathbb{N}$, we obtain a subsequence $\left(a_{k_s}^{(n+1)}\right)_{k=1}^{\infty}$ of $\left(a_k^{(n+1)}\right)$, which is also bounded. By the Bolzano-Weierstrass theorem, $\left(a_{k_s}^{(n+1)}\right)$ has a convergent subseqence, say $\left(a_{k_{s_p}}^{(n+1)}\right)_{p=1}^{\infty}$. Then by taking indices

$$k(1) = k_{s_1}, k(2) = k_{s_2}, \ldots, k(p) = k_{s_p}, \ldots,$$

we have convergent subsequences $\left(a'_{k(p)}\right)_{p=1}^{\infty}$ of $\left(a'_k\right)$ and $\left(a_{k(p)}^{(n+1)}\right)_{p=1}^{\infty}$ of $\left(a^{(n+1)}\right)$. It follows from (a) that

$$\left(a_{k(p)}\right)_{p=1}^{\infty} = \left(\left(a_{k(p)}^{(1)}, \ldots, a_{k(p)}^{(n+1)}\right)\right)_{p=1}^{\infty}$$

is a convergent subsequence of (a_k), as required. ∎

Integration in $\mathbb{R}^n$

(1.29) Here is the framework for this section, until Proposition 1.19. We fix

(a) a nonempty set X;

(b) a collection $\mathcal{C}$ of subsets of X such that $\varnothing \in \mathcal{C}$ and that $\mathcal{C}$ is a π-system;

(c) a function $\mu : \mathcal{C} \to [0,)$ which is decomposition-additive; and

(d) a nonempty set $A \in \mathcal{C}$, for which we consider the space of integrable functions $\mathrm{Int}_b(A, \mathbb{R})$.

From the previous section, we have seen that $\mathrm{Int}_b(A, \mathbb{R})$ is closed under multiplication, maximum, and minimum, but we have not discussed any explicit formulas for computing the integral of these functions. In connection to that, we shall look at inequalities that can be proved concerning the values of the integral of $fg, f \vee g, f \wedge g$. The fundamental property behind any of these inequalities is the following *nonnegative* property.

Proposition 1.17. Nonnegativity of Integral

Let $f \in \mathrm{Int}_b(A, \mathbb{R})$ be such that $f \geq 0$ for all $x \in A$. Then

$$\int_A f \geq 0.$$

Proof. From the definition of the lower integral as a supremum, we know that

$$\int_A f = \underline{\int_A} f \geq L(f, \Delta)$$

for any division Δ of A. So it suffices to find a division Δ such that $L(f, \Delta) \geq 0$. Indeed, $\Delta = \{A\}$ works for this case, since

$$L(f, \{A\}) = \mu(A) \inf_A (f) = \mu(A) \inf(f) \geq 0,$$

since $\mu(A)$ and $f \geq 0$. ∎

Proposition 1.18. *Let $f,g \in \mathrm{Int}_b(A,\mathbb{R})$. Then*

(a) if $f < g$, then $\int_A f \le \int_A g$;

(b) $\int_A f \vee g \ge \max\left(\int_A f, \int_A g\right)$ and $\int_A f \wedge g \le \min\left(\int_A f, \int_A g\right)$; and

(c) $\left|\int_A f\right| \le \int_A |f|$.

Proof. Let $f,g \in \mathrm{Int}_b(A,\mathbb{R})$.

(a) Let $h = g - f$. Then $h \ge 0$, so $\int_A h \ge 0$ by Proposition 1.17. Then by the linearity of integral, one has
$$\int_A g - \int_A f = \int_A g - f = \int_A h \ge 0,$$
and the result follows.

(b) The verification of $\int_A f \vee g \ge \max\left(\int_A f, \int_A g\right)$ follows form (a) and the fact that $\int_A f \vee g \ge \int_A f, \int_A g$. $\int_A f \wedge g \ge \min\left(\int_A f, \int_A g\right)$ can be verified similarly.

(c) We apply (a) to the inequalities $-|f| \le f \le |f|$ to obtain
$$-\int_A |f| \le \int_A f \le \int_A |f|,$$
from which the result follows.

∎

(1.30) The inquality in (c) of Proposition 1.18 is a particular case of the Cauchy-Schwarz inequality for integrals. In order to put this into perspective, let us introduce the notion of semi-inner product of integrable functions.

Def'n. Semi-inner Product of Integrable Functions
We define the ***semi-inner product*** $\langle \cdot, \cdot \rangle : \mathrm{Int}_b(A,\mathbb{R}) \times \mathrm{Int}_b(A,\mathbb{R}) \to \mathbb{R}$ by
$$\langle f, g \rangle = \int_A fg$$
for all $f,g \in \mathrm{Int}_b(A,\mathbb{R})$.

Observe that the closure under multiplication of $\mathrm{Int}_b(A,\mathbb{R})$ allows the above definition. We can verify that the provided definition indeed defines a semi-inner product on $\mathrm{Int}_b(A,\mathbb{R})$ as follows.

Proof. Fix $f,g,h \in \mathrm{Int}_b(A,\mathbb{R})$.

(a) *bilinearity*: Observe that
$$\langle \alpha f + \beta g, \rangle = \int_A (\alpha f + \beta g) h = \int_A \alpha f h + \beta g h = \int_A \alpha f h + \int_A \beta g h = \langle \alpha f, h \rangle + \langle \beta g, h \rangle,$$
and the verification of $\langle f, \alpha g + \beta h \rangle = \langle f, \alpha g \rangle + \langle f, \beta h \rangle$ follows from the commutativity of multiplication of functions.

(b) *symmetry*: Observe that
$$\langle f, g \rangle = \int_A fg = \int_A gf = \langle g, f \rangle.$$

(c) *positive semidefiniteness*: Observe that

$$\langle f, f\rangle = \int_A f^2 \geq 0$$

by Proposition 1.17.

∎

Observe that the positive semidefiniteness instead of positive definiteness is what we mean by *semi*-inner product.

Proposition 1.19. Cauchy-Schwarz Inequality for Integrals

Let $f, g \in \mathrm{Int}_b(A, \mathbb{R})$. *Then*

$$\left|\int_A fg\right| \leq \sqrt{\int_A f^2}\sqrt{\int_A g^2}.$$

Proof. Define $\varphi : \mathbb{R} \to \mathbb{R}$ by

$$\varphi(t) = \langle f + tg, f + tg\rangle.$$

Then by the positive semidefiniteness, $\varphi > 0$ for all $t \in \mathbb{R}$. Moreover, by using the bilinearity and symmetry,

$$\varphi(t) = \langle f, f\rangle + 2t\langle f, g\rangle + t^2\langle g, g\rangle$$

is a quadratic function, where $\langle f, f\rangle, \langle g, g\rangle \geq 0$ by positive semidefiniteness. We now consider the following cases:

(a) Suppose that $\langle g, g\rangle = 0$ (i.e. φ is linear). Then we see that $\langle f, g\rangle = 0$ as well, since a linear function

$$\varphi(t) = bt + c$$

satisfies $\varphi \geq 0$ for all $t \in \mathbb{R}$ if and only if $b = 0$ and $c \geq 0$. So it follows that

$$\left|\int_A fg\right| = 0 \leq 0 = \sqrt{\langle f, f\rangle}\sqrt{\langle g, g\rangle}.$$

(b) Suppose that $\langle g, g\rangle \neq 0$. Then we may write φ as

$$\varphi = \langle g, g\rangle\left(t^2 + \frac{2\langle f, g\rangle}{\langle g, g\rangle}t + \frac{\langle f, f\rangle}{\langle g, g\rangle}\right) = \langle g, g\rangle\left(t + \frac{\langle f, g\rangle}{\langle g, g\rangle}\right)^2 + \left(\langle f, f\rangle - \frac{\langle f, g\rangle^2}{\langle g, g\rangle}\right),$$

so when $t = -\frac{\langle f, g\rangle}{\langle g, g\rangle}$,

$$\langle f, f\rangle\langle g, g\rangle \geq \langle f, g\rangle^2,$$

or

$$|\langle f, g\rangle| \leq \langle f, f\rangle\langle g, g\rangle.$$

∎

(1.31) So far, we have been discussing integration is a general setting, where we fix arbitrary

(a) nonempty set X;

(b) collection $\mathcal{C}$ of subsets of X such that $\varnothing \in \mathcal{C}$ and such that $\mathcal{C}$ is a π-system; and

(c) decomposition-additivie function $\mu : \mathcal{C} \to [0, \infty)$.

As we move into more specific setting of Euclidean spaces, we fix $X = \mathbb{R}^n$ for some $n \in \mathbb{N}$, $\mathcal{C} = \mathcal{P}_n$, and $\mu : \mathrm{vol}_n : \mathcal{P}_n \to [0, \infty)$. From the preceding sections, we also know about when a function

$$f : P \to \mathbb{R}$$

is integrable, when $P \in \mathcal{P}_n$ is given. Moreover, we know some properties of $\mathrm{Int}_b(P, \mathbb{R})$, that it is closed under certain operations and some equalities and inequalities. However, what we now desire to do is expand our framework so that the notion of integrability can be defined for any bounded function on a nonempty bounded subset $A \subseteq \mathbb{R}$. To achieve this goal, we introduce the following notion.

Def'n. Extension of a Function on a Bounded Set
Let $A \subseteq B \subseteq \mathbb{R}^n$ be bounded subsets and let $f : A \to \mathbb{R}$. Then we define the (zero) ***extension*** of f to B, denoted by $\tilde{f}$, to be

$$\tilde{f} = \begin{cases} f & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} : B \to \mathbb{R}.$$

The idea is this. Any bounded subset $A \subseteq \mathbb{R}^n$ has a superset $B \supseteq A$ which is a half-open rectangle in $\mathbb{R}^n$, on which we already know how to integrate.

Proposition 1.20. *Let nonempty $P, Q \in \mathcal{P}_n$ with $P \subseteq Q$, let $f \in \mathbb{F}_b(P, \mathbb{R})$, and let $\tilde{f} \in F_b(Q, \mathbb{R})$ be the extension of f to Q. Then*

$$\overline{\int_P} f = \overline{\int_Q} \tilde{f}, \underline{\int_P} f = \underline{\int_Q} \tilde{f}.$$

Proof. Assume, without loss of generality, that $P \neq Q$. Since $\mathcal{P}_n$ is a semi-ring, we can find pairwise disjoint half-open rectangles $E_1, \ldots, E_k \in \mathcal{P}_n$ such that

$$\bigcup_{i=1}^{k} E_i = Q \setminus P.$$

We may also assume that each E_i is nonempty, as any empty E_i can be removed safely without affecting the above union. We can also find pairwise disjoint half-open rectangles $P_1, \ldots, P_r \in \mathcal{P}_n$ such that

$$\Delta = \{P_1, \ldots, P_r\}$$

is a division of P, which naturally allows us to obtain

$$\tilde{\Delta} = \{P_1, \ldots, P_r, E_1, \ldots, E_k\},$$

a division of Q. Then from the definition of f, we see that

$$U\left(\tilde{f}, \tilde{\Delta}\right) = \sum\nolimits_{i=1}^{r} \mu(P_i) \sup_{P_i}\left(\tilde{f}\right) + \sum\nolimits_{j=1}^{k} \mu(E_j) \sup_{E_j}\left(\tilde{f}\right) = \sum\nolimits_{i=1}^{r} \mu(P_i) \sup_{P_i}(f) = U(f, \Delta)$$

and that $L\left(\tilde{f}, \tilde{\Delta}\right) = L(f, \Delta)$ in a similar way. Conversely, given a division $\tilde{\Gamma}$ of Q, we can find a refinement

$$\tilde{\Gamma}' = \{P_1, \ldots, P_r, E_1, \ldots, E_k\}$$

such that $\{P_1, \ldots, P_r\} = \Gamma'$ is a division of P, and one also finds that

$$U\left(f, \Gamma'\right) = U\left(\tilde{f}, \tilde{\Gamma}'\right) \leq U\left(f, \tilde{\Gamma}\right)$$

and

$$L\left(f, \Gamma'\right) = L\left(\tilde{f}, \tilde{\Gamma}'\right) \geq L\left(\tilde{f}, \tilde{\Gamma}\right).$$

The result easily follows from the definition of upper and lower integrals as infimum and supremum. ∎

Proposition 1.21. *Let $A \subseteq \mathbb{R}^n$ be bounded and let $P_1, P_2 \subseteq \mathcal{P}_n$ be supersets of A. Let $f \in F_b(A, \mathbb{R})$ and let*

$$\tilde{f}_i = \begin{cases} f & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} : P_i \to \mathbb{R}$$

for each $i \in \{1,2\}$. Then

$$\overline{\int_{P_1}} \tilde{f}_1 = \overline{\int_{P_2}} \tilde{f}_2, \underline{\int_{P_1}} \tilde{f}_1 = \underline{\int_{P_2}} \tilde{f}_2.$$

Proof. Observe that $A \subseteq Q = P_1 \cap P_2 \subseteq \mathcal{P}_n$. Define the extension of f to Q

$$\hat{f} = \begin{cases} f & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} : Q \to \mathbb{R},$$

then we see that each $\tilde{f}_i$ is the extension of $\hat{f}$ to P_i and the result easily follows from Proposition 1.20. ∎

(1.32) The result from Proposition 1.21 allows the following definitions.

Def'n. Upper Integral, Lower Integral of a Function on a Bounded Set
Let $A \in \mathbb{R}^n$ be nonempty and bounded and let $f \in F_b(A, \mathbb{R})$. Given a superset $Q \in \mathcal{P}_n$ with $A \subseteq Q$, we define the ***upper integral*** of f by

$$\overline{\int_A} f = \overline{\int_Q} f.$$

Similarly, we define the ***lower integral*** of f by

$$\underline{\int_A} f = \underline{\int_Q} f.$$

Def'n. Integrable Function on a Bounded Set
Let $A \in \mathbb{R}^n$ be nonempty and bounded and let $f \in F_b(A, \mathbb{R})$. We say f is ***integrable*** if

$$\underline{\int_A} f = \overline{\int_A} f.$$

If so, the ***integral*** of f is defined as

$$\int_A f = \underline{\int_A} f = \overline{\int_A} f.$$

(1.33) Let us check some possible issues with the proceding definitions.

(a) We verify that each bounded set is indeed enclosed by a half-open rectangle. Suppose that $A \subseteq \mathbb{R}^n$ is bounded. Then by definition, there exists $x \in \mathbb{R}^n$ and ≥ 0 such that

$$A \subseteq B(x; r).$$

Then by inspection, we find that

$$B(x; r) \subseteq \prod_{i=1}^n [x_i - r, x_i + r).$$

(b) The provided definition of upper and lower integrals may not be well-defined if they depends on the particular choice of the enclosing half-open rectangle. The role of Proposition 1.21 is precisely to prevent this kind of trouble.

(c) What if we begin with some $A \subseteq \mathcal{P}_n$? Then we already assigned the values of $\overline{\int_A} f$ and $\underline{\int_A} f$ from preceding sections. So, what if we run into some different values? Fortunately, we can choose our enclosing half-open rectangle to be A, and Proposition 1.21 then guarantees that the values are going to be the same even if we choose some proper superset $Q \in \mathcal{P}_n$ of A.

Continuity

(1.34) What we desire to do in this section is to generalize the notion of continuity and integrability theorems on $\mathbb{R}$ to Euclidean spaces. Recall that, in $\mathbb{R}$, $\varepsilon - \delta$ definition of continuity and sequential continuity are equivalent, and we see the same phenomenon in Euclidean spaces.

Def'n. Continuous, Sequentially Continuous Function

Let $A \subseteq \mathbb{R}^n$, let $a \in A$, and let $f : A \to \mathbb{R}$.

(a) We say f is ***continuous*** at a if

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in A \left[\|x-a\| < \delta \implies |f(x) - f(a)|\right].$$

We say f is ***continuous*** on $B \subseteq A$ if f is continuous at every $a \in B$.

(b) We say f is ***sequentially continuous*** at a if, for any sequence $(x_k)_{k=1}^{\infty}$ on A with $\lim_{k\to\infty} x_k = a$, we have

$$\lim_{k\to\infty} f(x_k) = f\left(\lim_{k\to\infty} x_k\right) = f(a).$$

Proposition 1.22. Sequential Characterization of Continuity

Let $A \subseteq \mathbb{R}^n$, let $a \in A$, and let $f : A \to \mathbb{R}$. Then f is continuous at a if and only if f is sequentially continuous at a.

Proof. Supppose that f is continuous at a and fix $(x_k)_{k=1}^{\infty}$ on A such that $\lim_{k\to\infty} x_k = a$. Then for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|x-a\| < \delta \implies |f(x) - f(a)| < \varepsilon \qquad [1.4]$$

for all $x \in A$. Since (x_k) converges to a, one can find $N \in \mathbb{N}$ such that

$$n \geq N \implies \|x_k - a\| < \delta. \qquad [1.5]$$

Combining [1.3] and [1.4] together gives

$$n \geq N \implies \|x_k - a\| < \delta \implies |f(x) - f(a)|$$

so $\lim_{k\to\infty} f(x_k) = f(a)$. Conversely, suppose that f is not continuous at a, or, equivalently,

$$\exists \varepsilon > 0 \forall \delta > 0 \exists x \in A \left[\|x-a\| < \delta \wedge |f(x) - f(a)| < \varepsilon\right].$$

That is, given such $\varepsilon > 0$, we can find $x_k \in A$ with

$$\|x_k - a\| < \frac{1}{k} \wedge |f(x_k) - f(a)| < \varepsilon$$

for all $k \in \mathbb{N}$. Since $\lim_{k\to\infty} x_k = a$, it follows that f is not sequentially continuous. ∎

Def'n. Uniformly Continuous Function
Let $A \in \mathbb{R}^n$ and let $f : A \to \mathbb{R}$. We say that f is ***uniformly continuous*** on $B \subseteq A$ if

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in A\left[\|x-y\| < \delta \implies |f(x) - f(y)| < \varepsilon\right].$$

(1.35) Fix $A \in \mathbb{R}^n$ and recall that we call $f : A \to \mathbb{R}$ is *Lipschitz* on A if there exists $c > 0$ such that

$$\forall x, y \in A\left[|f(x) - f(y)| < c\|x-y\|\right].$$

It is immediate from the above definitions that

$$f \text{ is Lipschitz} \implies f \text{ is uniformly continuous} \implies f \text{ is continuous}.$$

One can show that these implications are indeed one-way by providing some examples.

(EX 1.36) Uniformly Continuous but not Lipschitz

Let

$$A = \left\{(s,t) \in \mathbb{R}^2 : s^2 + t^2 \leq 1\right\},$$

the closed unit disc centered at the origin, and let

$$g((s,t)) = \sqrt{1 - s^2 - t^2},$$

for all $(s,t) \in A$. Then g is uniformly continuous but not Lipschitz.

Proof. Exercise. ∎

(EX 1.37) Continuous but not Uniformly Continuous

Consider a half-open rectangle $P = (0,1]^2 \subseteq \mathcal{P}_2$ and let $f : P \to \mathbb{R}$ be defined by

$$(s,t) \mapsto \sin\left(\frac{s}{t}\right),$$

for all $(s,t) \in P$. Then f is continuous but not uniformly continuous.

Proof. Exercise. ∎

(1.38) Mesh of a Division

For the remaining part of the section, we fix $n \in \mathbb{N}$ (i.e. the dimension we are working with) and we consider the semi-ring $\mathcal{P}_n$ of half-open rectangles in $\mathbb{R}^n$. Here is an additional notation that we are going to use: given a division $\Delta = \{P_1, \ldots, P_k\}$ of $P \in \mathcal{P}_n$, we define the ***mesh*** of Δ, denoted as $\|\Delta\|$ by

$$\|\Delta\| = \max\{\operatorname{diam}(P_1), \ldots, \operatorname{diam}(P_k)\}.$$

One useful thing to know about mesh is that, given a division Δ_0 of P and $\varepsilon > 0$, one can always find a refinement Δ of Δ_0 with $\|\Delta\| < \varepsilon$. We use this notion to prove the following theorem relating the notion of uniform continuity to integrability.

Theorem 1.23. Uniform Continuity Implies Integrability

Let $P \in \mathcal{P}_n$ and let $f \in F_b(P,\mathbb{R})$. If f is uniformly continuous, then f is integrable on P.

Proof. We are going to use ε-Δ criterion of integrability, so fix $\varepsilon > 0$, and we desire to find a division Δ of P such that

$$U(f,\Delta) - L(f,\Delta) < \varepsilon.$$

For convenience, denote $V = \mathrm{vol}_n(P) > 0$. We invoke the hypothesis that f is uniformly continuous on P, but in connection to $\varepsilon' = \frac{\varepsilon}{2V}$. That is, there exists $\delta > 0$ such that

$$\forall x, y \in P \left[\|x - y\| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{2} \right].$$

Then by (1.38), there exists a division Δ of P with $\|\Delta\| < \delta$, and write

$$\Delta = \{P_1, \ldots, P_r\}.$$

We now claim the following:

(a) For any P_i, $\mathrm{osc}_{P_i}(f) \leq \frac{\varepsilon}{2V}$. To verify this, fix $i \in \{1, \ldots, r\}$ and observe that

$$\forall x, y \in P_i \left[\|x - y\| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{2V} \right],$$

since $P_i \subseteq P$. Therefore, by the fact that

$$\operatorname*{osc}_{P_i}(f) = \sup\{|f(x) - f(y)| : x, y \in P_i\},$$

we have $\mathrm{osc}_{P_i}(f) \leq \frac{\varepsilon}{2V}$.

(b) Δ satisfies $U(f,\Delta) - L(f,\Delta) < \varepsilon$. To see this, observe that

$$U(f,\Delta) - L(f,\Delta) = \sum\nolimits_{i=1}^{r} \mathrm{vol}_n(P_i) \operatorname*{osc}_{P_i}(f) \leq \sum\nolimits_{i=1}^{r} \mathrm{vol}_n(P_i) \frac{\varepsilon}{2V} = \frac{\varepsilon}{2V} \mathrm{vol}_n(P) = \frac{\varepsilon}{2} < \varepsilon.$$

Observe that conbining (a) and (b) provides the desired result. ∎

(1.39) Now that we have proved integrability theorem for a uniformly continuous function, let as look back at (EX 1.36) and (EX 1.37). For g, it is uniformly continuous, so it is integrable by Theorem 1.23. On the other hand, f is not uniformly continuous, so Theorem 1.23 cannot be applied. But fortunately, we observe that if we *remove* the domain of f little bit (say, only allow points $(s,t) \in \mathrm{domain}(f)$ with $t \geq \frac{1}{100}$), then one finds that f is uniformly continuous (and even Lipschitz) on $(0,1] \times \left(\frac{1}{100}, 1\right]$. In fact, we may make this removal of the domain of f arbitrarily small to make f uniformly continuous, and one can show that (Theorem 1.24) such function is indeed integrable. This motivates the following definition.

Def'n. Uniformly Continuous Modulo Parts of Small Volume Function

Let $P \in \mathcal{P}_n$ and let $f \in F_b(P,\mathbb{R})$. We say f is ***uniformly continuous modulo parts of small volume*** if, for all $\varepsilon > 0$, there exists $E_1, \ldots, E_k \in \mathcal{P}_n$ such that

(a) $E_1, \ldots, E_k \subseteq P$;

(b) $E_i \cap E_j = \varnothing$ whenever $i \neq j$;

(c) $\sum_{i=1}^{r} \mathrm{vol}_n(E_i) < \varepsilon$; and

(d) f is uniformly continuous on $P \setminus \bigcup_{i=1}^{k} E_i$.

Theorem 1.24. Uniform Continuity Modulo Parts of Small Volume Implies Integrability

Let $P \in \mathcal{P}_n$ and let $f \in F_b(P,\mathbb{R})$. If f is uniformly continuous modulo parts of small volume, them f is integrable.

Proof. We first verify the claim that, if $f \in F_b(A,\mathbb{R})$ for some $A \in \mathcal{P}_n$ is such that f is integrable on $A_1, A_2 \subseteq A$ with $A_1 \cap A_2 = \varnothing$ and $A_1 \cup A_2 = A$, then f is integrable on A. To see this, fix $\varepsilon, \varepsilon_1, \varepsilon_2 > 0$ such that $\varepsilon = \varepsilon_1 + \varepsilon_2$ and observe one can find a division Δ_1 of A_1 such that

$$U(f,\Delta_1) - L(f,\Delta_1) < \varepsilon_1 .$$

Similarly, one can find a division Δ_2 of A_2 such that

$$U(f,\Delta_2) - L(f,\Delta_2) < \varepsilon_2 .$$

Let $\Delta = \Delta_1 \cup \Delta_2$. Then Δ is a division of A and we have

$$U(f,\Delta) - L(f,\Delta) = U(f,\Delta_1) + U(f,\Delta_2) - L(f,\Delta_1) - L(f,\Delta_2) < \varepsilon_1 + \varepsilon_2 = \varepsilon,$$

so f is integrable, as claimed. Now, to prove the theorem, suppose that f is uniformly continuous modulo parts of small volume. Fix $\varepsilon > 0$ and let $E_1, \ldots, E_k \in \mathcal{P}_n$ be as described in the definition. Then f is integrable on $P \setminus \bigcup_{i=1}^k E_i$. To show that f is integrable on $\bigcup_{i=1}^k E_i$, observe that $\Delta = \{E_1, \ldots, E_k\}$ is a division of $\bigcup_{i=1}^k E_i$ such that

$$U(f,\Delta) - L(f,\Delta) = \sum_{i=1}^k \mathrm{vol}_n(E_i) \operatorname*{osc}_{E_i}(f) \leq \operatorname{osc}(f) \sum_{i=1}^k \mathrm{vol}_n(E_i) < \varepsilon .$$

Thus f is integrable on $\bigcup_{i=1}^k E_i$, and the result follows from the claim. ∎

This page intentionally left blank.

2.
Topology and Analysis

Topology of Euclidean Spaces

(2.1) We now begin to discuss the *topology* on Euclidean spaces (let us fix $n \in \mathbb{N}$, the dimension we are working with, for convenience). We shall persue this direction until we discuss uniformly continuous functions defined on *compact* subsets of $\mathbb{R}^n$. The reason for this is as follows. At the end of Chapter 1, we have arrived to a property (*uniform continuity modulo parts of small volume*) that guarantees the integrability of a function $f : A \to \mathbb{R}$ defined on a half-open rectangle $A \in \mathcal{P}_n$ of $\mathbb{R}^n$. Although this notion works, it is not a property of a function which one may naturally think about, so we introduce uniform continuity on compact sets to further clarify this. And while doing so, it is natural to introduce in detail some fundamental concepts for the study of $\mathbb{R}^n$.

Recall. Open Ball, Closed Ball of $\mathbb{R}^n$.
Let $a \in \mathbb{R}^n$ and $r \geq 0$.

(a) We define the ***open ball*** centered at a of radius r by

$$B(a;r) = \{x \in \mathbb{R}^n : \|a - x\| < r\}.$$

(b) We define the ***closed ball*** centered at a of radius r by

$$\overline{B}(a;r) = \{x \in \mathbb{R}^n : \|a - x\| \leq r\}.$$

Def'n. Interior, Closure of a Subset of $\mathbb{R}^n$
Let $A \subseteq \mathbb{R}^n$.

(a) We say a point $a \in \mathbb{R}^n$ is an ***interior point*** of A if

$$\exists r > 0 \left[B(a;r) \subseteq A\right].$$

We call the set of all interior points the ***interior*** of A, denoted by $\operatorname{int}(A)$.

(b) We say a point $a \in \mathbb{R}^n$ is ***adherent*** to A if

$$\forall r > 0 \left[B(b;r) \cap A \neq \varnothing\right].$$

We call the set of all points adherent to A the ***closure*** of A, denoted by $\operatorname{cl}(A)$.

(2.2) Boundary of Subset

From the above definitions, it is immediate that

$$\operatorname{int}(A) \subseteq A \subseteq \operatorname{cl}(A)$$

for any subset $A \subseteq \mathbb{R}^n$. To elaborate the second inclusion a bit (the firs one is very clear, as we only consider the points in A when defining $\operatorname{int}(A)$), observe that, for any $b \in A$ and $r > 0$, one has

$$b \in B(b;r) \implies B(b;r) \cap A \neq \varnothing. \qquad [2.1]$$

Inclusions in [2.1] can be read as saying that, when we define $\operatorname{int}(A)$, we lose some poins in A, while when we take $\operatorname{cl}(A)$, some poins are added to A. The information on overall difference between these gains and losses can be captured by the set difference $\operatorname{cl}(A) \setminus \operatorname{int}(A)$. This is called the ***boundary*** of A, denoted by

$$\operatorname{bd}(A) = \operatorname{cl}(A) \setminus \operatorname{int}(A).$$

Any $b \in \operatorname{bd}(A)$ is also called a ***boundary point*** of A.

(2.3) Here are some remarks about the operations of taking interior and closure of a subset $A \subseteq \mathbb{R}^n$.

(a) The operations are well-behaved with respect to inclusions. That is, given $M \subseteq N \subseteq \mathbb{R}^n$, one has

(i) $\operatorname{int}(M) \subseteq \operatorname{int}(N)$; and

(ii) $\operatorname{cl}(M) \subseteq \operatorname{cl}(N)$.

Proof. For (i), fix $x \in \operatorname{int}(M)$. Then there exists $r > 0$ such that

$$B(x;r) \in M \subseteq N.$$

For (ii), fix $x \in \operatorname{cl}(M)$. Then for any $r > 0$,

$$\varnothing \neq B(x;r) \cap M \subseteq B(x;r) \cap N,$$

as desired. ∎

(b) The above inclusions have the following consequences. Given $A, B \in \mathbb{R}^n$,

(i) $\operatorname{int}(A \cup B) \supseteq \operatorname{int}(A) \cup \operatorname{int}(B)$;

(ii) $\operatorname{cl}(A \cup B) \supseteq \operatorname{cl}(A) \cup \operatorname{cl}(B)$;

(iii) $\operatorname{int}(A \cap B) \subseteq \operatorname{int}(A) \cap \operatorname{int}(B)$; and

(iv) $\operatorname{cl}(A \cap B) \subseteq \operatorname{cl}(A) \cap \operatorname{cl}(B)$.

We only verify (ii), as other inclusions can be verified in a similar manner.

Proof. By (a), we have $\operatorname{cl}(A), \operatorname{cl}(B) \subseteq \operatorname{cl}(A \cup B)$, as $A, B \subseteq A \cup B$. This means $\operatorname{cl}(A) \cup \operatorname{cl}(B) \subseteq \operatorname{cl}(A \cup B)$. ∎

In fact, (ii) and (iii) always hold with equality, whereas (i) and (iv) are sometimes strict. Again, we only verify about (ii).

Proof. Suppose that $x \in \mathbb{R}^n$ is such that $x \notin \operatorname{cl}(A), \operatorname{cl}(B)$. Then there exists $r > 0$ such that

$$B(x;r) \cap A = B(x;r) \cap B = \varnothing.$$

But this means

$$B(x;r) \cap (A \cup B) = (B(x;r) \cap A) \cup (B(x;r) \cap B) = \varnothing.$$ ∎

(EX 2.4)
Interior, Closure, Boundary of a Half-open Rectangle

For convenience, consider working in the 2-dimensional Euclidean space, $\mathbb{R}^2$. Let

$$A = (0,1]^2 \in \mathcal{P}_2,$$

a half-open rectangle in $\mathbb{R}^2$. What are the sets $\operatorname{int}(A), \operatorname{cl}(A), \operatorname{bd}(A)$? In fact, one can show that

$$\begin{cases} \operatorname{int}(A) & = (0,1)^2 \\ \operatorname{cl}(A) & = [0,1]^2 \\ \operatorname{bd}(A) & = [0,1]^2 \setminus (0,1)^2 \end{cases}.$$

Notice how intuitive the result is. For instance, $\operatorname{bd}(A)$ is the union of four line segments, the four sides of the rectangle A.

(EX 2.5) A Less Intuitive Example

Consider the following set:

$$S = \left\{(v_1, v_2) \in \mathbb{R}^2 : v_1, v_2 \in (0,1) \cap \mathbb{Q}\right\}.$$

Find $\operatorname{int}(S), \operatorname{cl}(S), \operatorname{bd}(S)$.

Answer. Fix $v = (v_1, v_2) \in S$ and an irrational $r > 0$ with $v_1 + \frac{r}{2} \in (0,1)$. Observe that this is as good as choosing any arbitrary $r > 0$, as if $r = \varepsilon$ for some arbitrary small $\varepsilon > 0$ fails to satisfy $B(v;r) \subseteq S$, then any $r > \varepsilon$ would also fail to satisfy the condition. Then $\left(v_1 + \frac{r}{2}, v_2\right) \in B(v;r)$ but $v_1 + \frac{r}{2} \notin \mathbb{Q}$, so

$$B(v;r)\ \ S$$

for all $r > 0$, which means $\operatorname{int}(S) = \varnothing$. On the other hand, let $u = (u_1, u_2) \in [0,1]^2$. Then, given any $r > 0$, there exists a rational $q_1 \in (u_1 - r, u_1 + r) \cap (0,1)$ by the density of $\mathbb{Q}$ in $\mathbb{R}$, and, similarly, a rational $q_2 \in (u_2 - r, u_2 + r)$ exists. So it follows that

$$[0,1]^2 \subseteq \operatorname{cl}(S).$$

But from (EX 2.4), we know that $\operatorname{cl}\left([0,1]^2\right) = [0,1]^2$. Since $S \subseteq [0,1]^2$, by (ii) of (a) of (2.3), we have

$$\operatorname{cl}(S) \subseteq [0,1]^2 \subseteq \operatorname{cl}(S),$$

which means $\operatorname{cl}(S) = [0,1]^2$. We also have that $\operatorname{bd}(S) = [0,1]^2$ also, as $\operatorname{int}(S) = \varnothing$. ∎

This result is much less intuitive than (EX 2.4) (e.g. we ahve the whole closed rectangle as boundary).

(2.6) We begin to develop some tools via which we can do some proofs or specific calculations concerning interior and closure. First thing to do is to prove the possiblility of reducing one of these notions to another via some suitable *duality* equations.

Proposition 2.1. Duality between Interior and Closure

For every $A \subseteq \mathbb{R}^n$, we have

(a) $\operatorname{int}(\mathbb{R}^n \setminus A) = \mathbb{R}^n \setminus \operatorname{cl}(A)$*; and*

(b) $\operatorname{cl}(\mathbb{R}^n \setminus A) = \mathbb{R}^n \setminus \operatorname{int}(A)$.

Proof. Since (a) and (b) can be proven in an analogous way, we shall only verify (a). Observe that

$$\begin{aligned} x \in \operatorname{int}(\mathbb{R}^n \setminus A) &\iff (\exists r > 0\, [B(x;r) \subseteq \mathbb{R}^n \setminus A]) \iff (\exists r > 0\, [B(x;r) \nsubseteq A]) \\ &\iff x \notin \operatorname{cl}(A) \iff x \in \mathbb{R}^n \setminus \operatorname{cl}(A). \end{aligned}$$

∎

We can also characterize closure and boundary in terms of sequences.

Proposition 2.2. Sequential Characterization of Closure

Let $A \subseteq \mathbb{R}^n$ and let $b \in \mathbb{R}^n$. Then $b \in \operatorname{cl}(A)$ if and only if there exists a sequence $(x_k)_{k=1}^{\infty}$ on A that converges to b.

Proof. Let $b \in \operatorname{cl}(A)$. Then for each $k \in \mathbb{N}$, one can choose $x_k \in B\left(b; \frac{1}{k}\right) \cap A$, which provides a sequence $(x_k)_{k=1}^{\infty}$ on A which converges to b. Conversely, let $(x_k)_{k=1}^{\infty}$ be a sequence on A such that $\lim_{k\to\infty} x_k = b$. Then for any $r > 0$, there exists $k \in \mathbb{N}$ such that $x_k \in B(b;r) \cap A$, so it follows that $b \in \operatorname{cl}(A)$. ∎

Corollary 2.2.1. Characterization of Boundary

Let $A \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$. Then the following are equivalent.

(a) $a \in \operatorname{bd}(A)$.

(b) $a \in \operatorname{cl}(A) \cap \operatorname{cl}(\mathbb{R}^n \setminus A)$.

(c) There exist sequences $(x_i)_{i=1}^{\infty}$ on A and $(y_j)_{j=1}^{\infty}$ on $\mathbb{R}^n \setminus A$ such that

$$\lim_{i\to\infty} x_i = \lim_{j\to\infty} y_j = a.$$

(d) $a \in \operatorname{bd}(\mathbb{R}^n \setminus A)$.

Proof. We show that (b), (c), (d) are equivalent to (a). For (a) $\iff$ (b), observe that

$$\operatorname{bd}(A) = \operatorname{cl}(A) \setminus \operatorname{int}(A) = \operatorname{cl}(A) \cap (\mathbb{R}^n \setminus \operatorname{int}(A)) = \operatorname{cl}(A) \cap \operatorname{cl}(\mathbb{R}^n \setminus A),$$

where the last equality holds by Proposition 2.1. Then (a) $\iff$ (c) follows immediately from (a) $\iff$ (b) and Proposition 2.2. (a) $\iff$ (d) is by the fact that $\mathbb{R}^n \setminus (\mathbb{R}^n \setminus A) = A$. ∎

Def'n. Open, Closed Subset of $\mathbb{R}^n$

Let $A \subseteq \mathbb{R}^n$. We say A is

(a) ***open*** if $A = \operatorname{int}(A)$; and

(b) ***closed*** if $A = \operatorname{cl}(A)$.

Proposition 2.3. Alternative Definition of Closedness

Let $A \subseteq \mathbb{R}^n$. Then A is closed if and only if $\mathbb{R}^n \setminus A$ is open.

Proof. Observe that

$$\begin{aligned} A \text{ is closed} &\iff A = \operatorname{cl}(A) \iff \operatorname{bd}(A) \subseteq A \iff \operatorname{bd}(\mathbb{R}^n \setminus A) \subseteq A \\ &\iff \operatorname{bd}(\mathbb{R}^n \setminus A) \cap (\mathbb{R}^n \setminus A) = \varnothing \iff \mathbb{R}^n \setminus A = \operatorname{int}(\mathbb{R}^n \setminus A) \iff \mathbb{R}^n \setminus A \text{ is open}. \end{aligned}$$ ∎

Open Sets and Closed Sets

(2.7) We desire to obtain some topological descriptions of open sets and closed sets.

(a) Recall that, given a subset $A \subseteq \mathbb{R}^n$, we say A is open if $A = \operatorname{int}(A)$. But, by definition, $\operatorname{int}(A) \subseteq A$, so whenever we say A is open, the only information we add is that $A \subseteq \operatorname{int}(A)$. Therefore, we see that A is an open set if and only if every point in A is an interior point.

(b) On the other hand, there are several alternative descriptions of closedness of $A \subseteq \mathbb{R}^n$.

Proposition 2.4. Characterization of Closed Sets

Let $A \subseteq \mathbb{R}$. Then the following are equivalent.

(a) A is closed.

(b) $\mathbb{R}^n \setminus A$ is open.

(c) For every convergent sequence $(x_k)_{k=1}^{\infty}$ on A, $\lim_{k\to\infty} x_k \in A$.

Proof. We verify (a) $\iff$ (b) and (a) $\iff$ (c). For (a) $\iff$ (b), observe that

$$A \text{ is closed} \iff A = \operatorname{cl}(A) = \operatorname{cl}(\mathbb{R}^n \setminus (\mathbb{R}^n \setminus A)) = \mathbb{R}^n \setminus \operatorname{int}(\mathbb{R}^n \setminus A)$$

For (a) $\iff$ (c), observe that

$$A \text{ is closed} \iff A = \operatorname{cl}(A) \iff \text{For any convergent } (x_k)_{k=1}^{\infty} \text{ on } A \left[\lim_{x\to\infty} x_k \in \operatorname{cl}(A) = A\right].$$ ∎

Since (a), (b), (c) are equivalent, one can choose (b) or (c) as the definition of closed sets, and prove that it is equivalent to (a). From the beginning, we have been referencing $B(x;r)$ as *open* ball and $\overline{B}(x;r)$ as *closed* ball. Are they actually open and closed, respectively?

Proposition 2.5. Open Balls and Closed Balls

For any $x \in \mathbb{R}^n$ and for any $r > 0$,

(a) $B(x;r)$ is open; and

(b) $\overline{B}(x;r)$ is closed.

Proof. Fix $x \in \mathbb{R}^n$ and fix $r > 0$.

(a) Fix $z \in B(x;r)$ and let $\varepsilon = r - \|x-z\|$, which is positive by definition. We claim that $B(z;\varepsilon) \subseteq B(x;r)$. To see this, observe that, for any $v \in B(\varepsilon;z)$,

$$\|v-x\| \leq \|v-z\| + \|z-x\| < \varepsilon + \|z-x\| = r - \|x-z\| + \|z-x\| = r.$$

(b) Fix $z \in \mathbb{R}^n \setminus \overline{B}(x;r)$ and let $\varepsilon = \|x-z\| - r$, which is positive by definition. We claim that $B(z;\varepsilon) \subseteq \mathbb{R}^n \setminus \overline{B}(x;r)$. To verify this, observe that, for any $v \in B(z;\varepsilon)$,

$$\|x-v\| = \|(x-z)-(v-z)\| \geq \|x-z\| - \|v-z\| > \|x-z\| - \varepsilon = \|x-z\| - (\|x-z\| - r) = r.$$ ∎

(2.8) Topology of $\mathbb{R}^n$

Consider the following collection of subsets of $\mathbb{R}^n$,

$$\mathcal{T}_n = \{G \subseteq \mathbb{R}^n : G \text{ is open}\}.$$

We call this the ***topology*** of $\mathbb{R}^n$. Here are some basic properties of $\mathcal{T}_n$.

(a) $\mathbb{R}^n \in \mathcal{T}_n$.

Proof. For any $x \in \mathbb{R}^n$, it is clear that $B(x,r) \subseteq \mathbb{R}^n$ for any $r > 0$. ∎

(b) $\varnothing \in \mathcal{T}_n$.

Proof. Notice that $\operatorname{cl}(\mathbb{R}^n) = \mathbb{R}^n$, as any convergent $(x_k)_{k=1}^{\infty}$ on $\mathbb{R}^n$ converges to a point in $\mathbb{R}^n$. Since $\varnothing = \mathbb{R}^n \setminus \mathbb{R}^n$, it follows that $\varnothing$ is open. ∎

(c) For any collection $\{G_i\}_{i\in I} \subseteq \mathcal{T}_n$ (where I is an index set),

$$\bigcup_{i\in I} G_i \in \mathcal{T}_n\,.$$

Proof. If $\bigcup_{i\in I} G_i = \varnothing$, then the result holds by (b). Otherwise, fix $x \in \bigcup_{i\in I} G_i$. Then there exists $j \in I$ such that $x \in G_j$. But G_j is open, so there exists $r > 0$ such that $B(x;r) \subseteq G_j \subseteq \bigcup_{i\in I} G_i$. ∎

(d) For any finite collection $\{G_k\}_{k=1}^m \subseteq \mathcal{T}_n$,

$$\bigcap_{k=1}^m G_k \in \mathcal{T}_n\,.$$

Proof. If $\bigcap_{k=1}^m G_k = \varnothing$, then the result holds by (b). Otherwise, fix $x \in \bigcap_{k=1}^m G_k$. Then $x \in G_k$ for each $k \in \{1,\ldots,m\}$, so there exists $r_1,\ldots,r_m > 0$ such that $B(x;r_1) \subseteq G_1,\ldots,B(x;r_m) \subseteq G_m$. Take $r = \min_{1\le k\le m}(r_k)$. Then $B(x;r) \subseteq G_k$ for all $k \in \{1,\ldots,m\}$, so it follows that $B(x;r) \subseteq \bigcap_{k=1}^m G_k$, as desired. ∎

Observe that, together with the De Morgan's law, the above result shows that if

$$\mathcal{F}_n = \{F \subseteq \mathbb{R}^n : F \text{ is closed}\}\,,$$

then T_n has the following properties.

(a) $\mathbb{R}^n \in \mathcal{F}_n$.

(b) $\varnothing \in \mathcal{F}_n$.

(c) For any finite collection $\{F_k\}_{k=1}^m \subseteq \mathcal{F}_n$,

$$\bigcup_{k=1}^m F_k \in \mathcal{F}_n\,.$$

(d) For any collection $\{F_i\}_{i\in I}$ (where I is an index set),

$$\bigcap_{i\in I} F_i \in \mathcal{F}_n\,.$$

We now proceed to obtain descriptions about the interior and closure of a subset of $\mathbb{R}^n$.

Proposition 2.6. *Let $A \subseteq \mathbb{R}^n$ and let $G \subseteq A$. If G is open, then $G \subseteq \operatorname{int}(A)$.*

Proof. Recall that, whenever $G \subseteq A$, $\operatorname{int}(G) \subseteq \operatorname{int}(A)$. But $G = \operatorname{int}(G)$, so $G \subseteq \operatorname{int}(A)$. ∎

Proposition 2.7. *For any $A \subseteq \mathbb{R}^n$, $\operatorname{int}(A)$ is an open set.*

Proof. Let $A \subseteq \mathbb{R}^n$ and fix $x \in \operatorname{int}(A)$. Then there exists $r > 0$ such that $B(x;r) \subseteq A$. Since $B(x;r)$ is open, $B(x;r) \subseteq \operatorname{int}(A)$ by Proposition 2.6. By the same argument, $B(x;r) \subseteq \operatorname{int}(\operatorname{int}(A))$. Thus $\operatorname{int}(A)$ is an open set. ∎

(2.9) Proposition 2.6 and 2.7 show that, given any $A \subseteq \mathbb{R}^n$, $\operatorname{int}(A)$ is the largest open subset contained in A. Indeed, one can show the following result.

Proposition 2.8. Characterization of Interior

Let $A \subseteq \mathbb{R}^n$. Then there exists unique $U \subseteq A$ such that

(a) U is open; and

(b) for any open $G \subseteq A$, $G \subseteq U$.

Proof. We have seen that $U = \operatorname{int}(A)$ satisfies the above properties. To show uniqueness, suppose that $U' \subseteq A$ also satisfies (a), (b). Then $U' \subseteq \operatorname{int}(A)$ by Proposition 2.6. On the other hand, $\operatorname{int}(A)$ is an open subset of A, so $\operatorname{int}(A) \subseteq U'$. So it follows that $U = \operatorname{int}(A) = U'$. ∎

(2.10) We now proceed to show the counterpart of Proposition 2.8 in connection to closures.

Proposition 2.9. *Let $A, F \subseteq \mathbb{R}^n$ be such that $A \subseteq F$. If F is closed, then $\operatorname{cl}(A) \subseteq F$.*

Proof. Since $A \subseteq F$, $\operatorname{cl}(A) \subseteq \operatorname{cl}(F)$. But $F = \operatorname{cl}(F)$, so $\operatorname{cl}(A) \subseteq F$. ∎

Proposition 2.10. *Let $A \subseteq \mathbb{R}^n$. Then there exists unique $U \subseteq \mathbb{R}^n$ such that*

(a) U is closed; and

(b) for any closed $F \subseteq \mathbb{R}^n$ with $F \supseteq A$, $F \supseteq U$.

Proof. Observe that $\operatorname{cl}(A)$ is closed, as

$$\mathbb{R}^n \setminus \operatorname{cl}(A) = \mathbb{R}^n \setminus (\mathbb{R}^n \setminus \operatorname{int}(\mathbb{R}^n \setminus A)) = \operatorname{int}(\mathbb{R}^n \setminus A)$$

is open. So $U = \operatorname{cl}(A)$ satisfies (a), (b) by Proposition 2.9. The uniqueness part is analogous to the presented proof of Proposition 2.8. ∎

By Proposition 2.10, we obtain the following alternative description of $\operatorname{cl}(A)$: $\operatorname{cl}(A)$ is the smallest closed subset of $\mathbb{R}^n$ containing A.

(2.11) Recall that, for any $A \subseteq \mathbb{R}^n$,

$$\operatorname{bd}(A) = \operatorname{cl}(A) \setminus \operatorname{int}(A),$$

which cna be written as

$$\operatorname{bd}(A) = \operatorname{cl}(A) \setminus \operatorname{int}(A) = \operatorname{cl}(A) \cap (\mathbb{R}^n \setminus \operatorname{int}(A)).$$

But $\mathbb{R}^n \operatorname{int}(A)$ is a closed set. That is, $\operatorname{bd}(A)$ is also a closed set.

Compact Sets

(2.12) We introduce the notion of compactness in this section. It turns out that it is great to have a compact set as the domain of a continuous function: it allows continuous functions to have properties that we would not normally assume. We begin by introducing, among many possible definitions of compactness, the shortest description of what compact subsets of $\mathbb{R}^n$ are. Moreover, we introduce a rather distinct notion called sequential compactness, and we are going to show that the two notions are equivalent.

Def'n. Compact, Sequentially Compact Subsets of $\mathbb{R}^n$
Let $A \subseteq \mathbb{R}^n$.

(a) We say A is ***compact*** if it is closed and bounded.

(b) We say A is ***sequentially compact*** if for every sequence $(x_k)_{k=1}^{\infty}$ on A, there exists a convergent subsequence $\left(x_{k(p)}\right)_{p=1}^{\infty}$ such that $\lim_{p\to\infty} x_{k(p)} \in A$.

Proposition 2.11. Sequential Characterization of Compactness

Let $A \subseteq \mathbb{R}^N$. Then A is compact if and only if A is sequentially compact.

Proof. For the forward direction, suppose that A is compact and fix a sequence $(x_k)_{k=1}^{\infty}$ on A. Since (x_k) is defined on a bounded set, it is a bounded sequence. So by the Bolzano-Weierstrass theorem, there exists a convergent sequence $\left(x_{k(p)}\right)_{p=1}^{\infty}$ of (x_k), which clearly converges in A, as A is closed, so A is sequentially compact. Conversely, suppose that A is not compact. We split the remaining proof intwo two parts.

(a) Suppose that A is not bounded and fix $x \in A$. Then for each $k \in \mathbb{N}$, we can choose $x_k \in A$ such that $x_k \notin B(x;k)$. This defines $(x_k)_{k=1}^{\infty}$, a sequence in A that is not bounded. So any subsequence $\left(x_{k(p)}\right)_{p=1}^{\infty}$ of (x_k) is not bounded as well, and hence (x_k) is a sequence on A which does not have a convergent sequence. Thus A is not sequentially compact.

(b) Suppose that A is not closed. Then there exists $x \in \operatorname{cl}(A) \setminus A$. But every $z \in \operatorname{cl}(A)$, there exists a sequence $(z_k)_{k=1}^{\infty}$ on A such that $\lim_{k\to\infty} z_k = z$. So there exists $(x_k)_{k=1}^{\infty}$ on A such that $\lim_{k\to\infty} x_k = x \notin A$. But for such (x_k), every subsequence $\left(x_{k(p)}\right)_{p=1}^{\infty}$ converges to $x \notin A$, so A is not sequentially compact. ∎

(2.13) One of the pleasing properties of compact sets is that, if $A \subseteq \mathbb{R}^n$ is compact and if $f : A \to \mathbb{R}$ is continuous, then f is uniformly continuous. Since we already know that a uniformly continuous function is continuous, it means that the notion of uniform continuity coincides with the notion of continuity on compact sets.

Proposition 2.12.

Let $A \subseteq^n$ be compact and let $f : A \to \mathbb{R}$. If f is continuous, then f is uniformly continuous.

Proof. Let $f : A \to \mathbb{R}^n$ be continuous and suppose that f is not uniformly continuous, for the sake of contradiction. Then there exists $\varepsilon > 0$ such that

$$\forall \delta > 0 \exists x, y \in A \left[\|x-y\| < \delta \wedge |f(x) - f(y)| \geq \varepsilon\right].$$

So fix such $\varepsilon > 0$ and for each $k \in \mathbb{N}$, let $x_k, y_k \in A$ be such that

$$\|x_k - y_k\| < \frac{1}{k} \wedge |f(x_k) - f(y_k)| \geq \varepsilon. \tag{2.2}$$

This gives two sequences $(x_k)_{k=1}^{\infty}, (y_k)_{k=1}^{\infty}$ on A. By Proposition 2.11, there exists a convergent subsequence $\left(x_{k(p)}\right)_{p=1}^{\infty}$ of (x_k). Now, we claim the following.

(a) $\left(y_{k(p)}\right)_{p=1}^{\infty}$ is convergent and, in particular, $\lim_{p\to\infty} y_{k(p)} = \lim_{p\to\infty} x_{k(p)}$

Proof. For convenience, denote $L = \lim_{p\to\infty} x_{k(p)}$. Then for each $p \in \mathbb{N}$,

$$0 \leq \left\|y_{k(p)} - L\right\| \leq \left\|y_{k(p)} - x_{k(p)}\right\| + \left\|x_{k(p)} - L\right\| < \frac{1}{k(p)} + \left\|x_{k(p)} - L\right\|,$$

so

$$0 = \lim_{p\to\infty} 0 \le \lim_{p\to\infty} \left\|y_{k(p)} - L\right\| \le \lim_{p\to\infty} \frac{1}{k(p)} + \left\|x_{k(p)} - L\right\| = 0.$$

Thus by the squeeze theorem, $\lim_{p\to\infty} \left\|y_{k(p)} - L\right\| = 0$, or, $\lim_{p\to\infty} x_{k(p)} = L$.

(b) $\lim_{p\to\infty} f\left(x_{k(p)}\right) = \lim_{p\to\infty} f\left(y_{k(p)}\right)$.

Proof. This is a direct consequence of the sequential characterization of continuity, as

$$\lim_{p\to\infty} f\left(x_{k(p)}\right) = f\left(\lim_{p\to\infty} x_{k(p)}\right) = f(L) = f\left(\lim_{p\to\infty} y_{k(p)}\right) = \lim_{p\to\infty} f\left(y_{k(p)}\right)$$

by (a).

Observe that the claim (b) contradicts the construction of (x_k) and (y_k): [2.2] implies

$$\left|f\left(x_{k(p)}\right) - f\left(y_{k(p)}\right)\right| \ge \varepsilon$$

for all $p \in \mathbb{N}$, but this clearly means $\lim_{p\to\infty} f\left(x_{k(p)}\right) \ne \lim_{p\to\infty} f\left(y_{k(p)}\right)$. Thus f is uniformly continuous, as required. ∎

(2.14) So far we have been working with real-valued functions only, and it seems like a natural question to ask that if there are analogous definitions and results (i.e. continuity, uniform continuity, ...) for functions which take values in $\mathbb{R}^m$ for some $m > 1$. One useful perspective concerning functions taking values in $\mathbb{R}^n$ is to treat them as an m-tuple of m real-valued functions. That is, given a function $f : A \to \mathbb{R}^m$ for some $A \subseteq \mathbb{R}^n$, we may define the components $f_1, \ldots, f_m : A \to \mathbb{R}^m$ such that

$$f(x) = (f_1(x), \ldots, f_m(x))$$

for all $x \in A$. This idea allows the following definitions.

Def'n. Continuous, Uniformly Continuous Functions whose Codomain Is $\mathbb{R}^m$

Let $f : A \to \mathbb{R}^m$ for some $A \subseteq \mathbb{R}^n$.

(a) We say f is ***continuous*** if each component function $f_i : A \to \mathbb{R}$ of f is continuous.

(b) We say f is ***uniformly continuous*** if each component function $f_i : A \to \mathbb{R}$ of f is uniformly continuous.

Observe that we have not used any $\varepsilon - \delta$ argument in the above definitions. For instance, one would expect to say $f : A \to \mathbb{R}^m$ is ***continuous*** at $a \in A$ if

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in A \left[\|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon\right]$$

and ***uniformly continuous*** if

$$\varepsilon > 0 \exists \delta > 0 \forall x, y \in A \left[\|x - y\| < \delta \implies \|f(x) - f(y)\|\right].$$

In fact, one can show that both ways of defining continuity (and uniform continuity) presented above are equivalent. Now that we have generalized notions of continuity and uniform continuity, let as prove a generalized version of Proposition 2.12 for functions which take values in $\mathbb{R}^m$.

Proposition 2.13. *Let $A \subseteq \mathbb{R}^n$ be compact and let $f : A \to \mathbb{R}^m$. If f is continuous, then f is uniformly continuous.*

Proof. Observe that, by Proposition 2.12,

$$\begin{aligned} f \text{ is continuous} &\implies \text{each component } f_i \text{ of } f \text{ is continuous} \\ &\implies \text{each component } f_i \text{ of } f \text{ is uniformly continuous} \\ &\implies f \text{ is uniformly continuous}. \end{aligned}$$

∎

Continuity on Compact Sets

(2.15) Continuing our previous work, we introduce three equivalent notions of continuous $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ and $n, m \in \mathbb{N}$. So fix dimensions $n, m \in \mathbb{N}$.

Proposition 2.14. *Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Fix $a \in A$. Then the following are equivalent.*

(a) Each component function $f_i : A \to \mathbb{R}^m$ is continuous at a.

(b) f is continuous in the sense of ε-δ. That is,

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in A \left[\|x - a\| \le \delta \implies \|f(x) - f(a)\| \le \varepsilon\right].$$

(c) f is sequentially continuous. That is, given any sequence $(x_k)_{k=1}^{\infty}$ on A which converges to a,

$$\lim_{k \to \infty} f(x_k) = f(a).$$

Proof. We proceed by the cycle (a) $\implies$ (b) $\implies$ (c) $\implies$ (a).

- (a) $\implies$ (b) Suppose that the component functions $f_1, \ldots, f_m : A \to \mathbb{R}$ are continuous and fix $\varepsilon > 0$. Then there exists $\delta_1, \ldots, \delta_m > 0$ such that, for all $x \in A$,

  $$\|x - a\| < \delta_i \implies \|f_i(x) - f_i(a)\| < \frac{\varepsilon}{m}.$$

  Take $\delta = \min_{1 \le i \le m}(\delta_i) > 0$. Then for any $x \in A$ with $\|x - a\| < \delta$, we have $|f_i(x) - f_i(a)| < \frac{\varepsilon}{m}$ for all $i \in \{1, \ldots, m\}$. But we have that

  $$\|f(x) - f(a)\| = \sqrt{\sum_{i=1}^{m} (f_i(x) - f_i(a))^2} \le \sum_{i=1}^{m} |f_i(x) - f_i(a)| < \sum_{i=1}^{m} \frac{\varepsilon}{m} = \varepsilon,$$

  so for any $x \in A$ with $\|x - a\| < \delta$ satisfies $\|f(x) - f(a)\| < \varepsilon$, as required. ∎

- (b) $\implies$ (c) Suppose that (b) holds and fix a convergent sequence $(x_k)_{k=1}^{\infty}$ on A such that $\lim_{k\to\infty} x_k = a$. Since (b) holds, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

  $$\forall x \in A \left[\|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon\right].$$

  But since $\lim_{k\to\infty} x_k = a$, for such δ, there exists $N \in \mathbb{N}$ such that

  $$k \ge N \implies \|x_k - a\| < \delta.$$

Therefore, this $N \in \mathbb{N}$ satisfies that

$$k \geq N \implies \|x_k - a\| < \delta \implies \|f(x_k) - f(a)\| < \varepsilon,$$

as required. ∎

- (c) $\implies$ (a) Recall that a sequence $(y_j)_{j=1}^{\infty}$ on $\mathbb{R}^m$ converges to $y \in \mathbb{R}^m$ if and only if each ith component sequence of (y_j) converges to the ith component of y. Therefore, given a sequence $(x_k)_{k=1}^{\infty}$ on A that converges to a, we have that $\lim_{k\to\infty} f(x_k) = f(a)$ by assumption, and so each component sequence $(f_i(x_k))_{k=1}^{\infty}$ converges to $f_i(a)$. But this exactly means that each f_i is sequentially continuous, which exactly means that it is continuous. ∎

Notice that (b) and (c) of Proposition 2.14 are essentially the m-dimensional analogues of continuity (in the sense of $\varepsilon - \delta$) and sequential continuity of real valued functions. We see this pattern repeats for uniform continuity and Lipschitzness.

Def'n. Uniformly Continuous Function

Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. We say f is ***uniformly continuous*** if

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in A \left[\|x - y\| < \delta \implies \|f(x) - f(y)\| < \varepsilon\right].$$

Proposition 2.15. *Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Then the following are equivalent.*

(a) f is uniformly continuous.

(b) Each component function $f_i : A \to \mathbb{R}$ are uniformly continuous.

Proof. Suppose that each component function $f_i : A \to \mathbb{R}$ is uniformly continuous and fix $\varepsilon > 0$. Then there exist $\delta_1, \ldots, \delta_m > 0$ such that

$$\forall x, y \in A \left[\|x - y\| < \delta_i \implies |f_i(x) - f_i(y)| < \varepsilon\right].$$

for all $i \in \{1, \ldots, m\}$. Take $\delta = \min_{1 \leq i \leq m}(\delta_i)$. Then such δ satisfies

$$\forall x, y \in A \left[\|x - y\| < \delta \implies |f_i(x) - f_i(y)| < \varepsilon\right].$$

It follows that, given $x, y \in A$ such that $\|x - y\| < \delta$,

$$\|f(x) - f(y)\| = \sqrt{\sum_{i=1}^{m} (f_i(x) - f_i(y))^2} \leq \sum_{i=1}^{m} |f_i(x) - f_i(y)| < \sum_{i=1}^{m} \frac{\varepsilon}{m} = \varepsilon.$$

Conversely, suppose that f is uniformly continuous. Then given $\varepsilon > 0$, there exists $\delta > 0$ such that, for any $x, y \in A$ with $\|x - y\| < \delta$, we have $\|f(x) - f(y)\| < \varepsilon$. But for such δ,

$$\|x - y\| < \delta \implies |f_i(x) - f_i(y)| = \sqrt{(f_i(x) - f_i(y))^2} \leq \sqrt{\sum_{i=1}^{m} (f_i(x) - f_i(y))^2} = \|f(x) - f(y)\| < \varepsilon,$$

as desired. ∎

Def'n. Lipschitz Function

Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. We say f is ***Lipschitz*** if there exists $c > 0$ such that

$$\forall x, y \in A \left[\|f(x) - f(y)\| < c\,\|x - y\|\right].$$

Proposition 2.16. *Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Then the following are equivalent.*

(a) f is Lipschitz.

(b) Each component function $f_i : A \to \mathbb{R}^m$ is Lipschitz.

Proof. Let $c > 0$ be such that

$$\forall x, y \in A \left[\|f(x) - f(y)\| < c\,\|x - y\|\right].$$

Then it is immediate that each component function is c-Lipschitz. Conversely, suppose that there exist $c_1, \ldots, c_m > 0$ such that

$$\forall x, y \in A \left[|f_i(x) - f_i(y)| < c_i\,\|x - y\|\right]$$

for all $i \in \{1, \ldots, m\}$. Take $c = \sum_{i=1}^m c_i$. Then, for any $x, y \in A$,

$$\|f(x) - f(y)\| = \sqrt{\sum_{i=1}^m (f_i(x) - f_i(y))^2} \leq \sum_{i=1}^m |f_i(x) - f_i(y)| < \sum_{i=1}^m c_i\,\|x - y\| = c\,\|x - y\|,$$

so f is c-Lipschitz. ∎

(2.16) It is an immediate consequence of Proposition 2.14, 2.15, 2.16 that, given $f : A \to \mathbb{R}^m$ for any $A \subseteq \mathbb{R}^n$, we have implications

$$f \text{ is Lipschitz} \implies f \text{ is uniformly continuous} \implies f \text{ is continuous}.$$

We now turn our attention to continuous functions with compact domains.

Proposition 2.17. *Let $A \subseteq \mathbb{R}^n$ be compact and let $f : A \to \mathbb{R}^m$ be continuous. Then*

$$\operatorname{image}(f) = f(A) = \{f(x) : x \in A\} \subseteq \mathbb{R}^m$$

is compact.

Proof. We prove that $\operatorname{image}(f)$ is sequentially compact. So fix a sequence $(y_k)_{k=1}^\infty$. Notice that each $y_k \in \operatorname{image}(f)$, so there exists $x_k \in A$ such that $f(x_k) = y_k$, and we obtain a sequence $(x_k)_{k=1}^\infty$ on A. But A is a compact set, so there exists a convergent subsequence $\left(x_{k(p)}\right)_{p=1}^\infty$ of (x_k) with $\lim_{p\to\infty} x_{k(p)} \in A$. Thus, for such indices $k(1) < k(2) < \cdots$, we have

$$\lim_{p\to\infty} y_{k(p)} = \lim_{p\to\infty} f\left(x_{k(p)}\right) = f\left(\lim_{p\to\infty} x_{k(p)}\right) \in \operatorname{image}(f),$$

as desired. ∎

(2.17) The main result we desire to prove is the extreme value theorem. So let us introduce the following notions.

Def'n. Global Maximum, Global Minimum of a Function
Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$.

(a) We say $a \in A$ is a point of ***global maximum*** if

$$\forall x \in A\,[f(x) \leq f(a)].$$

(b) We say $a \in A$ is a point of ***global minimum*** if

$$\forall x \in A\,[f(x) \geq f(a)].$$

Note that a function function may not have a global maximum or a global minimum. On the other hand, if they exist, they need not be unique. But at least we can get rid of the existence problem when the domain of a function is compact.

Theorem 2.18. Extreme Value Theorem

Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$. If f is continuous and A is compact, then there exist a point of global maximum and a point of global minimum for f.

Lemma 2.18.1.

Let $K \subseteq \mathbb{R}$ be compact. Then there exist $\alpha, \beta \in K$ such that $\alpha \leq t \leq \beta$ for all $t \in K$.

Proof. We are going to prove the existence of β only, as the verification of the existence of α is essentially the same. Since K is compact, $\sup(K)$ is well-defined. Now, define a sequence $(x_k)_{k=1}^{\infty}$ by observing the fact that, there exists $x_k \in K$ such that

$$\sup(K) - \frac{1}{k} < x_k < \sup(K)$$

for all $k \in \mathbb{N}$ (if not, then $\sup(K) - \frac{1}{k}$ is another upper bound for K, which violates the minimality of $\sup(K)$). Then by the squeeze theorem,

$$\lim_{k\to\infty} \sup(K) - \frac{1}{k} = \sup(K) \leq \lim_{k\to\infty} x_k \leq \sup(K),$$

which means $\lim_{k\to\infty} x_k = \sup(K)$. But K is closed, so any convergent sequence on K converges to a point in K, and, in particular, $\sup(k) = \lim_{k\to\infty} x_k \in K$. Thus $\beta = \sup(K) \in K$ is the point that we are looking for. ∎

Proof of Theorem 2.18. Since f is continuous and A is compact, $\text{image}(f) \subseteq \mathbb{R}$ is compact. So by Lemma 2.18.1, there exist $M, m \in \text{image}(f)$ such that

$$m \leq t \leq M$$

for all $t \in \text{image}(f)$. For such M, m, there exist $x_M, x_m \in A$ such that $f(x_M) = M$ and $f(x_m) = m$, as $M, m \in \text{image}(f)$, and x_M, x_m are points of global maximum and global minimum, respectively. ∎

Integrability of Bounded Continuous Functions

(2.18) Previously, given a function $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}^n$, we have seen that f is integrable if f is uniformly continuous (modulo parts of small volume) - which is not so satisfying result. In this section, we discover a more satisfying result, that if f is continuous and bounded, then f is integrable. After this, we ask the following question: if f has *some*, but *not so many*, discontinuities, then is f still integrable? To deal with question, we introduce the notion of null sets. Throughout this section, let us fix $n \in \mathbb{N}$, the dimension we are working with.

Theorem 2.19. Bounded Continuous Function Is Integrable

Let

$$P = \prod_{i=1}^{n} [a_i, b_i) \in \mathcal{P}_n,$$

where $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ *with* $a_1 < b_1 \cdots a_n < b_n$ *be a half-open rectangle and let* $f : P \to \mathbb{R}$. *If* f *is bounded and continuous, then* f *is uniformly continuous modulo parts of small volume. In particular,* f *is integrable.*

Lemma 2.19.1.

Let

$$P = \prod_{i=1}^{n} [a_i, b_i) \in \mathcal{P}_n.$$

(a) For every $\varepsilon > 0$, *there exists* $S \in \mathcal{P}_n$ *such that* $\operatorname{cl}(S) \subseteq P$ *and such that* $\operatorname{vol}_n(S) > \operatorname{vol}_n(P) - \varepsilon$.

(b) For every $\varepsilon > 0$, *there exists* $Q \in \mathcal{P}_n$ *such that* $\operatorname{cl}(Q) \subseteq P$ *and such that* $\operatorname{vol}_n(Q) < \operatorname{vol}_n(P) + \varepsilon$.

Proof. We are going to prove (a) only and (b) is left as an exercise. Fix $\varepsilon > 0$.

(a) Notice that $\operatorname{vol}_n(P) = \prod_{i=1}^{n} l_i$, where $l_i = b_i - a_i$ for each $i \in \{1, \ldots, n\}$. The idea is to pick $\sigma_1, \ldots, \sigma_n > 0$ such that $S = \prod_{i=1}^{n} [a_i, b_i - \sigma_i)$, where $\sigma_i < l_i$ for each $i \in \{1, \ldots, n\}$. Recall that the closure of the Cartesian product of intervals is the Cartesian product of closures of each interval, so

$$\operatorname{cl}(S) = \operatorname{cl}\left(\prod_{i=1}^{n} [a_i, b_i - \sigma_i)\right) = \prod_{i=1}^{n} \operatorname{cl}([a_i, b_i - \sigma_i)) = \prod_{i=1}^{n} [a_i, b_i - \sigma_i] \subseteq P,$$

where the last inclusion holds by the fact that $[a_i, b_i - \sigma_i] \subseteq [a_i, b_i)$ for all $i \in \{1, \ldots, n\}$. In order to discuss $\operatorname{vol}_n(S)$, let us pick σ_i as follows: for each $i \in \{1, \ldots, n\}$, let

$$\sigma_i = \frac{l_i}{k}$$

and let S_k denote the corresponding half-open rectangle, for all $k \in \mathbb{N}$, $k \geq 2$. Then

$$\operatorname{vol}_n(S_k) = \prod_{i=1}^{n} (l_i - \sigma_i) = \prod_{i=1}^{n} l_i - \frac{l_i}{k} = \left(\frac{k-1}{k}\right)^n \prod_{i=1}^{n} l_i,$$

so

$$\operatorname{vol}_n(P) - \operatorname{vol}_n(S_k) = \left(1 - \left(\frac{k-1}{k}\right)^n\right) \operatorname{vol}_n(P).$$

Thus,

$$\lim_{k \to \infty} \operatorname{vol}_n(P) - \operatorname{vol}_n(S_k) = \lim_{k \to \infty} \left(1 - \left(\frac{k-1}{k}\right)^n\right) = 0,$$

so given any $\varepsilon > 0$, one can find $k \in \mathbb{N}$ such that $S = S_k$ satisfies

$$\operatorname{vol}_n(P) - \operatorname{vol}_n(S) = \operatorname{vol}_n(P) - \operatorname{vol}_n(S_k) < \varepsilon,$$

as desired. ∎

(b) Exercise. ∎

Proof of Theorem 2.19. Fix $\varepsilon > 0$. By Lemma 2.19.1, let $S \in \mathcal{P}_n$ be such that $\operatorname{cl}(S) \subseteq P$ and that $\operatorname{vol}_n(S) > \operatorname{vol}_n(P) - \varepsilon$. Then $P \setminus S$ is a set difference of half-open rectangles, so there exists pairwise disjoint $E_1, \ldots, E_k \in \mathcal{P}_n$ such that $\bigcup_{i=1}^k E_i = P \setminus S$. Moreover, $\operatorname{cl}(S)$ is bounded and closed, so $\operatorname{cl}(S)$ is compact. Therefore, f is uniformly continuous on $\operatorname{cl}(S)$, and, in particular, on S. Thus f is uniformly continuous modulo parts of small volume. ∎

(2.19) As mentioned in (2.18), we are going to discuss about null sets in $\mathbb{R}^n$ to strengthen the result shown in Theorem 2.19, by allowing functions to have null-set discontinuities. But before going at them, it would be more convenient to point out few basic results about the half-open rectangles, to allow the notion of volume for a broader collection of subsets of $\mathbb{R}^n$. This larger collection turn out to be a ring of subsets of $\mathbb{R}^n$ which contains $\mathcal{P}_n$. That is, if we denote

$$\hat{\mathcal{P}}_n = \left\{ A \subseteq \mathbb{R}^n : \exists P_1, \ldots, P_k \in \mathcal{P}_n \left[\bigcup_{i=1}^k P_i = A \right] \right\},$$

then it is immediate that $\hat{\mathcal{P}}_n \supseteq \mathcal{P}_n$. Moreover, we have the following consequence of this definition.

Proposition 2.20. $\hat{\mathcal{P}}_n$ Is a Ring of Subsets

$\hat{\mathcal{P}}_n \subseteq \mathcal{P}(\mathbb{R}^n)$ is a ring of subsets of $\mathbb{R}^n$.

Proof. We verify four things.

(a) Since $\varnothing \in \mathcal{P}_n$, $\varnothing \in \hat{\mathcal{P}}n$. ∎

(b) Let $P, Q \in \hat{\mathcal{P}}_n$. Then there exists $P_1, \ldots, P_k, Q_1, \ldots, Q_r \in \mathcal{P}_n$ such that $P = \bigcup_{i=1}^k P_i$ and $Q = \bigcup_{j=1}^r Q_j$. So

$$P \setminus Q = \left(\bigcup_{i=1}^k P_i \right) \setminus \left(\bigcup_{j=1}^r Q_j \right) = \bigcup_{i=1}^k \left(P_i \setminus \bigcup_{j=1}^r Q_j \right) = \bigcup_{i=1}^k \left(\bigcup_{j=1}^r P_i \setminus Q_j \right) = \bigcup_{i=1}^k \bigcup_{j=1}^r P_i \setminus Q_j,$$

where each $P_i \setminus Q_j$ is a union of members of $\mathcal{P}_n$, since $\mathcal{P}_n$ is a semi-ring. So $P \setminus Q \in \hat{\mathcal{P}}_n$. ∎

(c) Let $P = \bigcup_{i=1}^k P_i, Q = \bigcup_{j=1}^r Q_j \in \hat{\mathcal{P}}_n$. Then

$$P \cap Q = \left(\bigcup_{i=1}^k P_i \right) \cap \left(\bigcup_{j=1}^r Q_j \right) = \bigcup_{i=1}^k \bigcup_{j=1}^r P_i \cap Q_j,$$

where each $P_i \cap Q_j \in \mathcal{P}_n$, since $\mathcal{P}_n$ is a semi-ring. So $P \cap Q \in \hat{\mathcal{P}}_n$. ∎

(d) Let $P, Q \in \hat{\mathcal{P}}_n$. Then P, Q are unions of members of $\mathcal{P}$, so $P \cup Q$ is a union of members of $\mathcal{P}$. Thus $P \cup Q \in \hat{\mathcal{P}}_n$. ∎

Notice that (a)-(c) are the properties of a semi-ring of sets, where (d) is the only property a semi-ring needs to become a ring. Thus $\hat{\mathcal{P}}_n$ is a ring of subsets of $\mathbb{R}^n$. ∎

(2.20) Now, consider a different collection of subsets of $\mathbb{R}^n$,

$$\hat{\mathcal{R}}_n = \left\{ A \subseteq \mathbb{R}^n : \exists Q_1, \ldots, Q_r \in \mathcal{P}_n \left[i \neq j \implies P_i \cap P_j = \varnothing \wedge \bigcup_{i=1}^k P_i = A \right] \right\}.$$

That is, we take subsets of A which can be written as a disjoint union of half-open rectangles of $\mathbb{R}^n$. Then it is immediate from the definition that $\hat{\mathcal{R}}_n \subseteq \hat{\mathcal{P}}_n$. Is the inclusion in the reversed direction also true? The answer is yes, as shown by the following proof.

Proof. Fix $A = \bigcup_{i=1}^k P_i \in \hat{\mathcal{P}}_n$. We define $Q_{1_1} = P_1$ and, for each $i \in \{2, \dots, k\}$, we proceed as follows. Take a set difference

$$P_i \setminus \bigcup_{j=1}^{i-1} P_j.$$

Since $\mathcal{P}_n$ is a semi-ring, there must exist $Q_{i_1}, \dots, Q_{i_{l_i}} \in \mathcal{P}_n$ such that

$$P_j \setminus \bigcup_{j=1}^{i-1} P_j = \bigcup_{\alpha=1}^{l_i} Q_{i_\alpha}.$$

Thus we obtain $Q_{1_1}, Q_{2_1}, \dots, Q_{2_{l_2}}, \dots, Q_{k_1}, \dots, Q_{k_{l_k}} \in \mathcal{P}_n$ such that

$$\bigcup_{i=1}^k \bigcup_{\alpha=1}^{l_i} Q_{i_\alpha} = \bigcup_{i=1}^k P_i = A,$$

so $A \in \hat{\mathcal{R}}_n$. ∎

That is, $\hat{\mathcal{R}}_n$ is an equivalent way of defining $\hat{\mathcal{P}}_n$. We now look at more or less canonical way of extending $\mathrm{vol}_n : \mathcal{P}_n \to [0, \infty)$ to $\hat{\mathrm{vol}}_n : \mathcal{P}_n \to [0, \infty)$. That is, we desire $\hat{\mathrm{vol}}_n$ to have properties that vol_n has, while maintaining the property $\hat{\mathrm{vol}}_n(P) = \mathrm{vol}_n(P)$ for all $P \in \mathcal{P}_n$.

Proposition 2.21. *There exists unique $\hat{\mathrm{vol}}_n : \hat{\mathcal{P}}_n \to [0, \infty)$ such that $\hat{\mathrm{vol}}_n(P) = \mathrm{vol}_n(P)$ for all $P \in \mathcal{P}_n$.*

Lemma 2.21.1. *Let $A \in \hat{\mathcal{P}}_n$ and suppose that there exist collection $\{P_i\}_{i=1}^k, \{Q_j\}_{j=1}^r \subseteq \mathcal{P}$ of pairwise disjoint half-open rectangles such that*

$$A = \bigcup_{i=1}^k P_i = \bigcup_{j=1}^r Q_j.$$

Then

$$\sum_{i=1}^k \mathrm{vol}_n(P_i) = \sum_{j=1}^r \mathrm{vol}_n(Q_j).$$

Proof. Observe that

$$C = \{P_i \cap Q_j : i \in \{1, \dots, k\} \wedge j \in \{1, \dots, r\}\}$$

is a collection of disjoint half-open rectangles satisfying $\bigcup C = A$. But it is clear that $P_i = \bigcup_{j=1}^r P_i \cap Q_j$ for all $i \in \{1, \dots, k\}$ and $Q_j = \bigcup_{i=1}^k P_i \cap Q_j$ for all $j \in \{1, \dots, r\}$. Thus

$$\sum_{i=1}^k \mathrm{vol}_n(P) = \sum_{i=1}^k \sum_{j=1}^r \mathrm{vol}_n(P_i \cap Q_j) = \sum_{j=1}^r \mathrm{vol}_n(Q_j)$$

by decomposition additivity. ∎

Proof of Proposition 2.21. Define $\hat{\mathrm{vol}}_n : \hat{\mathcal{P}}_n \to [0, \infty)$ by

$$\hat{\mathrm{vol}}_n(A) = \sum_{i=1}^k \mathrm{vol}_n(P_i)$$

given that $\{P_i\}_{i=1}^k$ is a disjoint collection of subsets such that $\bigcup_{i=1}^k P_i = A$. Then by Lemma 2.21.1, this $\hat{\text{vol}}_n$ is well-defined. Moreover, it is clear from the definition that $\hat{\text{vol}}_n(P) = \text{vol}_n(P)$ for all $P \in \mathcal{P}_n$. The uniqueness part is left as an exercise. ∎

Proposition 2.22. Properties of $\hat{\text{vol}}_n$

(a) Let $A, C \in \hat{\mathcal{P}}_n$. Then $\hat{\text{vol}}_n(A) \leq \hat{\text{vol}}_n(C)$.

(b) For any $A, B \in \hat{\mathcal{P}}_n$,

$$\hat{\text{vol}}_n(A \cup B) \leq \hat{\text{vol}}_n(A) + \hat{\text{vol}}_n(B).$$

Moreover, this holds with an equality if A, B are disjoint.

Proof. We first verify the second part of (b). Suppose that A, B are disjoint and let $\{P_i\}_{i=1}^k, \{Q_j\}_{j=1}^r \subseteq \mathcal{P}_n$ be collections of pairwise disjoint half-open rectangles such that $A = \bigcup_{i=1}^k P_i, B = \bigcup_{j=1}^r Q_j$. Then $P_1, \ldots, P_k, Q_1, \ldots, Q_r$ are pairwise disjoint, so

$$\hat{\text{vol}}_n(A \cup B) = \sum_{i=1}^k \text{vol}_n(P_i) + \sum_{j=1}^r \text{vol}_n(Q_j) = \hat{\text{vol}}_n(A) + \hat{\text{vol}}_n(B).$$

(a) Notice that $C = (C \setminus A) \cup A$. So

$$\hat{\text{vol}}_n(C) = \hat{\text{vol}}_n((C \setminus A) \cup A) = \hat{\text{vol}}_n(C \setminus A) + \hat{\text{vol}}_n(A) \geq \hat{\text{vol}}_n(A)$$

by the previous result. ∎

(b) Let $A, B \in \hat{\mathcal{P}}_n$. Then

$$\hat{\text{vol}}_n(A \cup B) = \hat{\text{vol}}_n(A \cup (B \setminus A)) = \hat{\text{vol}}_n(A) + \hat{\text{vol}}_n(B \setminus A) \geq \hat{\text{vol}}_n(A) + \hat{\text{vol}}_n(B)$$

by (a). ∎

(2.21) It is an immediate result of (b) of Proposition 2.22 that, given $A_1, \ldots, A_k \in \hat{\mathcal{P}}_n$,

$$\hat{\text{vol}}_n\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k \hat{\text{vol}}_n(A_i),$$

which holds with an equality if $A_1, \ldots, A_k$ are pairwise disjoint.

Def'n. Null Set
We say $N \subseteq \mathbb{R}^n$ is a ***null set*** if for every $\varepsilon > 0$, there exist countably many $P_1, \ldots, P_k \in \mathcal{P}_n$ such that $N \subseteq \bigcup_{i=1}^k P_i$ and that $\sum_{i=1}^k \text{vol}_n(P_i) < \varepsilon$.

(2.22) Here are some basic properties of null sets.

(a) Every null set is bounded.

(b) Let $M, N \subseteq \mathbb{R}^n$ be such that $M \subseteq N$. If N is a null set, then M is a null set.

(c) Let $N_1, \ldots, N_k \in \mathbb{R}^n$ and let $U = \bigcup_{i=1}^k N_i$. If each N_i is a null set, then U is a null set.

Moreover, one can also show that every finite subset of $\mathbb{R}^N$ is also a null set.

(2.23) Our definition of null sets allows one to have *cover* $\{P_1,\ldots,P_n\}$ without inforcing $P_1,\ldots,P_k$ to be pairwise disjoint. But what happens if one decides to enforce this condition? It is clear that, a null set in terms pairwise disjoint $P_1,\ldots,P_k$ is a null set in terms of any $P_1,\ldots,P_k$, but the reverse direction is not so trivial. Nevertheless, we verify the reverse direction as follows. In fact, we show that the following three statements are equivalent. Fix $N \subseteq \mathbb{R}^n$.

(a) N is a null set.

(b) For any $\varepsilon > 0$, there exists $A \in \hat{\mathcal{P}}_n$ such that $A \supseteq N$ and that $\hat{\mathrm{vol}}_n(A) < \varepsilon$.

(c) For any $\varepsilon > 0$, there exist pairwise disjoint $P_1,\ldots,P_k$ such that $\bigcup_{i=1}^k P_i \supseteq N$ and $\sum_{i=1}^k \mathrm{vol}_n(P_i) < \varepsilon$.

Proof. We proceed to verify the chain (a) $\implies$ (b) $\implies$ (c) $\implies$ (a).

- (a) $\implies$ (b) Let $P_1,\ldots,P_k \in \hat{\mathcal{P}}_n$ be such that $N \subseteq \bigcup_{i=1}^k P_i$ and $\sum_{i=1}^k \mathrm{vol}_n(P_i) < \varepsilon$. Then $\bigcup_{i=1}^k P_i \subseteq \hat{\mathcal{P}}_n$ by (2.20). So $A = \bigcup_{i=1}^k P_i$ is such that $A \subseteq \hat{\mathcal{P}}_n$ and $\hat{\mathrm{vol}}_n(A) < \varepsilon$. ▮
- (b) $\implies$ (c) Fix such $A \in \hat{\mathcal{P}}_n$ and write $A = \bigcup_{i=1}^k P_i$ for some pairwise disjoint $P_1,\ldots,P_k \in \mathcal{P}_n$. Then it is clear that $N \subseteq \bigcup_{i=1}^k P_i = A$ and that
$$\sum_{i=1}^k \mathrm{vol}_n(P_i) = \hat{\mathrm{vol}}_n(A) < \varepsilon\,.$$ ▮
- (c) $\implies$ (a) This is immediate as mentioned before the beginning of the proof. ▮

We can also show other properties of null sets in terms of topological nature.

Proposition 2.23. Topological Properties of Null Sets

Let $N \subseteq \mathbb{R}^n$ be a null set. Then

(a) $\mathrm{int}(N) = \varnothing$*; and*

(b) $\mathrm{cl}(N)$ *is a null set.*

Proof.

(a) Suppose, for the sake of contradiction, that $\mathrm{int}(N) \neq \varnothing$. Then we can choose an interior point $a \in N$, which means there exists $r > 0$ such that $B(a;r) \subseteq N$. But for such ball $B(a;r)$, we can choose small $l > 0$ (e.g. $l = \frac{r}{\sqrt{n}}$) such that
$$C = \prod_{i=1}^n [a_i, a_i + l) \subseteq B(a;r) \subseteq N.$$
But this means that, for any $A \in \hat{\mathcal{P}}_n$ such that $A \supseteq N$, $A \supseteq C$ and so
$$\mathrm{vol}_n(A) \geq \mathrm{vol}_n(C) = l^n,$$
so by choosing $\varepsilon < l^n$, we have no $A \in \hat{\mathcal{P}}_n$ such that $A \supseteq N$ and $\hat{\mathrm{vol}}_n(A) < \varepsilon$, which is a contradiction. Thus $\mathrm{int}(N) = \varnothing$. ▮

(b) Fix $\varepsilon > 0$. Then there exist $P_1,\ldots,P_k \in \mathcal{P}_n$ such that $N \subseteq \bigcup_{i=1}^k P_i$ and that $\sum_{i=1}^k \mathrm{vol}_n(P_i) < \frac{\varepsilon}{2}$. Now, for every $i \in \{1,\ldots,k\}$, Lemma 2.19.1 provides $Q_i \in \mathcal{P}_n$ such that $\mathrm{cl}(P_i) \subseteq Q_i$ and that $\mathrm{vol}_n(Q_i) \leq 2\,\mathrm{vol}_n(P_i)$. Then
$$\sum_{i=1}^k \mathrm{vol}_n(Q_i) \leq 2\sum_{j=1}^k \mathrm{vol}_n(P_i) < \varepsilon\,.$$

On the other hand, observe that

$$\mathrm{cl}(N) = \mathrm{cl}\left(\bigcup_{i=1}^{k} P_i\right) = \bigcup_{i=1}^{k} \mathrm{cl}(P_i) \subseteq \bigcup_{i=1}^{k} Q_i.$$

Thus N is a null set, as required. ∎

This is a side note. A subset $S \subseteq \mathbb{R}^n$ is said to be ***nowhere dense*** if $\mathrm{int}(\mathrm{cl}(S)) = \varnothing$. So Proposition 2.23 guarantees that a null set is always nowhere dense. It turns out that, given a dimension n, any *lower dimensional* subsets of $\mathbb{R}^n$ is a null set (e.g. line segment in $\mathbb{R}^2$). To make this idea precise, consider the following proposition.

Proposition 2.24. *Let $m, n \in \mathbb{N}$ with $m < n$, $Y \subseteq \mathbb{R}^m$ be bounded, and let $h : Y \to \mathbb{R}^n$ be Lipschitz. Then* $\mathrm{image}(h)$ *is a null set.*

(EX 2.24) For now, let us accept Proposition 2.24 as a fact and see what examples can be derived from it.

(a) A continuous function $f : [0,1] \to \mathbb{R}^n$ is called a ***path*** in $\mathbb{R}^n$. If f is Lipschitz, we call f a ***Lipschitz path***. It is a special case of Proposition 2.24 that the image of any Lipschitz path is a null set.

(b) The sphere $S = \{x \in \mathbb{R}^n : \|x\| = 1\}$ is a null set.

(c) Let $P \subseteq \mathbb{R}^3$ be a closed parallelipiped

$$P = \prod_{i=1}^{3} [a_i, b_i],$$

where $a_1 < b_1, a_2 < b_2, a_3 < b_3$ in $\mathbb{R}$. The boundary $\mathrm{bd}(P)$ can be written as the union of 6 *faces* of dimension 2, so weach face is a null set. It follows from (c) of (2.22) that $\mathrm{bd}(P)$ is also a null set. This is good - we shall show that having such small (i.e. null set) boundary is a pleasing property.

Def'n. Continuous Modulo a Null Set
Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$. We say f is ***continuous modulo a null set*** if there exists a null set $N \subseteq \mathbb{R}^n$ such that f is continuous on $A \setminus N$.

(2.25) A continuous function is sure to be continuous modulo a null set: take $N = \varnothing$. Now, our goal is to prove that, given any bounded $f : P \to \mathbb{R}$ on a half-open rectangle $P \in \mathcal{P}_n$,

$$f \text{ is continuous modulo a null set} \implies f \text{ is uniformly continuous modulo parts of small volume},$$

which gives us that f is integrable. Notice that this is a stronger version of Proposition 2.19.

Proposition 2.25. *Let $P \in \mathcal{P}_n$ and let $P \to \mathbb{R}$ be bounded and continuous modulo a null set. Then f is uniformly continuous modulo parts of small volume, and, in particular, f is integrable.*

Lemma 2.25.1. *Let $A \in \hat{\mathcal{P}}_n$. Then for all $\varepsilon > 0$, there exists $B \in \hat{\mathcal{P}}_n$ such that* $\mathrm{cl}(B) \subseteq A$ *and that* $\hat{\mathrm{vol}}_n(B) > \hat{\mathrm{vol}}_n(A) - \varepsilon$.

Proof. Consider writing A as a disjoint union $\bigcup_{i=1}^{k} P_i$ for some $P_1, \ldots, P_k \in \mathcal{P}_n$. We know that

$$\hat{\mathrm{vol}}_n(A) = \sum_{i=1}^{k} \mathrm{vol}_n(P_i)$$

by definition. For every $i \in \{1, \ldots, k\}$, we invoke Lemma 2.19.1 for P_i and in connection to $\frac{\varepsilon}{k}$. This provides $S_i \in \mathcal{P}_n$ such that $\mathrm{cl}(S_i) \subseteq P_i$ and that $\mathrm{vol}_n(S_i) > \hat{\mathrm{vol}}_N(P_i) - \frac{\varepsilon}{k}$. Now, define

$$B = \bigcup_{i=1}^{k} S_i$$

and observe that

$$\mathrm{cl}(B) = \mathrm{cl}\left(\bigcup_{i=1}^{k} S_i\right) = \bigcup_{i=1}^{k} \mathrm{cl}(S_i) \subseteq \bigcup_{i=1}^{k} P_i = A.$$

Moreover, notice that $S_1, \ldots, S_k$ are pairwise disjoint, since for any $i, j \in \{1, \ldots, k\}$ with $i \neq j$, $S_i \cap S_j \subseteq P_i \cap P_j = \varnothing$. This allows us to write

$$\hat{\mathrm{vol}}_n(B) = \sum_{i=1}^{k} \mathrm{vol}_n(S_i) > \sum_{i=1}^{k} \left(\mathrm{vol}_n(P_i) - \frac{\varepsilon}{k}\right) = \hat{\mathrm{vol}}_n(A) - \varepsilon,$$

as desired. ∎

Proof of Proposition 2.25. For convenience, let us rephrase uniform continuity modulo parts of small volume. That is, a function $g : P \to \mathbb{R}$ is uniformly continuous modulo parts of small volue if, for any $\varepsilon > 0$, there exists $E \in \hat{\mathcal{P}}_n$ such that $\hat{\mathrm{vol}}_n(E) < \varepsilon$ and that g is uniformly continuous on $P \setminus E$. This is an indeed equivalent definition, with slight changes of notation. Now fix $\varepsilon > 0$ for the remaining four steps of proof.

(a) Since f is continuous modulo a null set, let $N \subseteq P$ be a null set such that f is continuous on $P \setminus N$.

(b) Since N is a null set, let $A' \in \hat{\mathcal{P}}_n$ be such that $A' \supseteq N$ and $\hat{\mathrm{vol}}_n(A') < \frac{\varepsilon}{2}$ and let $A = P \cup A'$. Then $A \supseteq N$ and $\hat{\mathrm{vol}}_n(A) \leq \hat{\mathrm{vol}}_n(A') < \frac{\varepsilon}{2}$.

(c) Lemma 2.25.1 provides that there exists $B \in \hat{\mathcal{P}}_n$ such that $\mathrm{cl}(B) \subseteq P \setminus A$ and $\hat{\mathrm{vol}}_n(B) > \hat{\mathrm{vol}}_n(P \setminus A) - \frac{\varepsilon}{2}$.

(d) Take $E = P \setminus B$. We claim that this E is such that $\hat{\mathrm{vol}}_n(E) < \varepsilon$ and that f is uniformly continuous on $P \setminus E = B$.

   (i) $\hat{\mathrm{vol}}_n(E) < \varepsilon$.

   Proof. Notice that

$$\begin{aligned}\hat{\mathrm{vol}}_n(E) = \hat{\mathrm{vol}}_n(P \setminus B) = \hat{\mathrm{vol}}_n(P) - \hat{\mathrm{vol}}_n(B) &< \hat{\mathrm{vol}}_n(P) - \left(\hat{\mathrm{vol}}_n(P \setminus A) - \frac{\varepsilon}{2}\right) \\ &= \hat{\mathrm{vol}}_n(P) - \left(\hat{\mathrm{vol}}_n(P) - \hat{\mathrm{vol}}_n(A) - \frac{\varepsilon}{2}\right) = \hat{\mathrm{vol}}_n(A) + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$ ∎

   (ii) f is uniformly continuous on $P \setminus E$.

   Proof. f is continuous oon $P \setminus N$. Moreover, $\mathrm{cl}(B) \subseteq P \setminus A \subseteq P \setminus N$, so f is continuous on $\mathrm{cl}(B)$. But $\mathrm{cl}(B)$ is compact, so f is uniformly continuous on $\mathrm{cl}(B)$. Thus f is uniformly continuous on $P \setminus E = B \subseteq \mathrm{cl}(B)$. ∎

Jordan Measurability

(2.26) In the preceding section, we addressed the issue of integrating functions with few discontinuities, as long as the discontinuous points form a null set. We now move onto a different issue: functions are more likely to have domains that are not half-open rectangles. Here again, if the domain of a function is *suitable*, then we expect the function to be integrable. This idea leads to the notion of *Jordan measurable* sets and *Jordan content*, the extension of vol_n function to the Jordan measurable subsets of $\mathbb{R}^n$. The main purpose of all these listed works is to upgrade our integrability theorem: any continuous (with discontinuities up to forming a null set), bounded function on a Jordan measurable set is integrable.

Def'n. Jordan Measurable Set
We say $A \subseteq \mathbb{R}^n$ is ***Jordan measurable*** if A is bounded and $\text{cl}(A)$ is a null set.

We shall use the notation

$$\mathcal{J}_n = \{A \subseteq \mathbb{R}^n : A \text{ is Jordan measurable }\}.$$

(EX 2.27)

(a) Any open or closed ball is Jordan measurable.

(b) Any open, half-open, or closed rectangle is Jordan measurable.

(c) Any null set is Jordan measurable.

(d) Recall the set from (EX 2.5):

$$S = \left\{x \in (0,1)^2 : x \in \mathbb{Q}\right\}.$$

We have found that $\text{int}(S) = \varnothing$ and $\text{cl}(S) = [0,1]^2$, so $\text{bd}(S) = \text{cl}(S) \setminus \text{int}(S) = [0,1]^2$. Therefore, $\text{bd}(S)$ is not a null set, so S is not Jordan measurable.

An important property of $\mathcal{J}_n$ is that it is a ring of subsets.

Proposition 2.26. $\mathcal{J}_n$ Is a Ring

$\mathcal{J}_n$ is a ring subsets. Moreover, $\hat{\mathcal{P}}_n \subseteq \mathcal{J}_n$.

Lemma 2.26.1.

For any $A, B \subseteq \mathbb{R}^n$,

(a) $\text{bd}(A \cup B) \subseteq \text{bd}(A) \cup \text{bd}(B)$*;*

(b) $\text{bd}(A \cap B) \subseteq \text{bd}(A) \cup \text{bd}(B)$*; and*

(c) $\text{bd}(A \setminus B) \subseteq \text{bd}(A) \cup \text{bd}(B)$.

Proof. Fix $A, B \subseteq \mathbb{R}^n$.

(a) Observe that

$$\text{bd}(A \cup B) = \text{cl}(A \cup B) \setminus \text{int}(A \cup B) = (\text{cl}(A) \cup \text{cl}(B)) \setminus \text{int}(A \cup B).$$

But $\text{int}(A \cup B) \supseteq \text{int}(A) \cup \text{int}(B)$, so

$$\begin{aligned} \text{bd}(A \cup B) &= (\text{cl}(A) \cup \text{cl}(B)) \setminus \text{int}(A \cup B) \subseteq (\text{cl}(A) \cup \text{cl}(B)) \setminus (\text{int}(A) \cup \text{int}(B)) \\ &= (\text{cl}(A) \setminus \text{int}(A)) \cup (\text{cl}(B) \setminus \text{int}(B)) = \text{bd}(A) \cup \text{bd}(B). \end{aligned}$$

∎

(b) Observe that, by De Morgan's law and (a),

$$\begin{aligned}\operatorname{bd}(A\cap B) &= \operatorname{bd}(\mathbb{R}^n\setminus(A\cap B)) = \operatorname{bd}(\mathbb{R}^n\setminus A)\cup\operatorname{bd}(\mathbb{R}^n\setminus B)\\ &\subseteq \operatorname{bd}(\mathbb{R}^n\setminus A)\cup\operatorname{bd}(\mathbb{R}^n\setminus B) = \operatorname{bd}(A)\cup\operatorname{bd}(B).\end{aligned}$$

(c) Observe that, by (b),

$$\operatorname{bd}(A\setminus B) = \operatorname{bd}(A\cap(\mathbb{R}^n\setminus B)) \subseteq \operatorname{bd}(A)\cup\operatorname{bd}(\mathbb{R}^n\setminus B) = \operatorname{bd}(A)\cup\operatorname{bd}(B).$$

Proof of Proposition 2.26. Observe that $\varnothing\in\mathcal{J}_n$. Moreover, Lemma 2.26.1 provides the other properties required for $\mathcal{J}_n$ to be a ring, since, for any $A,B\in\mathcal{J}_n$ (i.e. bounded $A,B\subseteq\mathbb{R}^n$ with null sets $\operatorname{cl}(A),\operatorname{cl}(B)$),

(a) $\operatorname{cl}(A\cup B)\subseteq\operatorname{cl}(A)\cup\operatorname{cl}(B)$ is a null set;

(b) $\operatorname{cl}(A\cap B)\subseteq\operatorname{cl}(A)\cup\operatorname{cl}(B)$ is a null set; and

(c) $\operatorname{cl}(A\setminus B)\subseteq$ is a null set,

where induction arguments on (a), (b) shows that $\mathcal{J}_n$ is closed under finite union and intersections. Moreover, for any $A\in\hat{\mathcal{P}}_n$, there exist $P_1,\dots,P_k\in\mathcal{P}_n$ such that $A=\bigcup_{i=1}^k P_i$, and by (b) of (EX 2.27) and (a), $A\in\mathcal{J}_n$, so $\hat{\mathcal{P}}_n\subseteq\mathcal{J}_n$.

(2.28) We are now ready to prove the most powerful version of integrability theorem. For the following that, recall the following defintion of integrability. For a bounded function $f:A\to\mathbb{R}$ on a bounded subset $A\subseteq\mathbb{R}^n$, we have considered an extension $\tilde{f}:P\to\mathbb{R}$ for some $P\in\mathcal{P}_n$ such that $A\subseteq P$, by

$$\tilde{f}(x)=\begin{cases} f(x) & \text{if } x\in A\\ 0 & \text{otherwise}\end{cases}$$

for all $x\in P$, and we say f is *integrable* whenever $\tilde{f}$ is integrable. Moreover, if f is integrable, we define

$$\int_A f = \int_P \tilde{f}.$$

Now we are all set.

Theorem 2.27. Integrability Theorem

Let $A\subseteq\mathbb{R}^n$ be Jordan measurable and let $f:A\to\mathbb{R}$. If f is bounded and continuous modulo a null set, then f is integrable on A.

Proof. Since A is Jordan measurable, A is bounded, and so let $P\in\mathcal{P}_n$ be such that $P\supseteq\operatorname{cl}(A)$. Let $\tilde{f}:P\to\mathbb{R}$ be defined by

$$\tilde{f}(x)=\begin{cases} f(x) & \text{if } x\in A\\ 0 & \text{otherwise}\end{cases}$$

for all $x\in P$. (tl;dr)

Def'n. Jordan Content of a Jordan Measurable Set

Let $A \subseteq \mathcal{J}_n$. Then we define the ***Jordan content*** of A, denoted by $\tilde{\text{vol}}_n(A)$ as follows.

(a) If $A = \varnothing$, then $\tilde{\text{vol}}_n(A) = 0$.

(b) If $A \neq \varnothing$, then consider $\mathbb{1} : A \to \mathbb{R}$ such that $\mathbb{1} = 1$ for all $x \in A$. Then, we define

$$\tilde{\text{vol}}_n(A) = \int_A \mathbb{1}.$$

Notice that the justification of the use of integral in the above defintion is provided by Theorem 2.27.

Proposition 2.28. *For any disjoint $A, B \in \mathcal{J}_n$,*

$$\tilde{\text{vol}}_n(A \cup B) = \tilde{\text{vol}}_n(A) + \tilde{\text{vol}}_n(B).$$

Proof. Let $\mathbb{1}_A, \mathbb{1}_B, \mathbb{1}_{A\cup B} : A \cup B \to \{0,1\}$ be indicator functions. That is, for all $X \in \{A, B, A \cup B\}$ and for all $x \in X$,

$$\mathbb{1}_X(x) = \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases}.$$

Then, by the fact that $A \cap B = \varnothing$, we have

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B.$$

Thus, by the linearity of integral,

$$\tilde{\text{vol}}_n(A \cup B) = \int_{A\cup B} \mathbb{1}_{A\cup B} = \int_{A\cup B} \mathbb{1}_A + \mathbb{1}_B = \int_{A\cup B} \mathbb{1}_A + \int_{A\cup B} \mathbb{1}_B = \int_A \mathbb{1}_A + \int_B \mathbb{1}_B = \tilde{\text{vol}}_n(A) + \tilde{\text{vol}}_n(B).$$ ∎

Corollary 2.28.1. Properties of Jordan Content

The map $\tilde{\text{vol}}_n : \mathcal{J}_n \to \mathbb{R}$ has the following properties.

(a) subadditivity: For any finite collection $\{A_i\}_{i=1}^k \subseteq \mathcal{J}_n$, one has

$$\tilde{\text{vol}}_n\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k \tilde{\text{vol}}_n(A_i),$$

which holds with an equality when $A_1, \ldots, A_k$ are pairwise disjoint.

(b) Let $A, C \in \mathcal{J}_n$ be such that $A \subseteq C$. Then $\tilde{\text{vol}}_n(A) \leq \tilde{\text{vol}}_n(C)$.

Proof. All properties above are immediate from the fact that $\mathcal{J}_n$ is a ring of subsets. For more details, see the presented proof of Proposition 2.22. ∎

Corollary 2.28.2. *The map $\tilde{\text{vol}} : \mathcal{J}_n \to [0,\infty)$ extends the map $\hat{\text{vol}}_n \to [0,\infty)$. That is, for any $A \in \hat{\mathcal{P}}_n$,*

$$\tilde{\text{vol}}_n(A) = \hat{\text{vol}}_n(A).$$

Proof. Observe that, for any half-open rectangle $P \in \mathcal{P}_n$,

$$\tilde{\text{vol}}_n(A) = \int_P \mathbb{1}_P = \text{vol}_n(P).$$

Let $P_1, \ldots, P_k \in \mathcal{P}_n$ be disjoint and such that $\bigcup_{i=1}^{k} P_i = A$. Then

$$\tilde{\mathrm{vol}}_n(A) = \sum_{i=1}^{k} \tilde{\mathrm{vol}}_n(P_i) = \sum_{i=1}^{k} \mathrm{vol}_n(P_i) = \hat{\mathrm{vol}}_n(A).$$

Computing Integrals by Slicing

(2.29) In the preceding section, we have arrived to the integrability theorem (Theorem 2.27) which we require for integrating multivariate functions. Here is a summary of the results we have so far.

(a) We desire to have Jordan measurable sets as the domains of the functions of our interest.

(b) On Jordan measurable sets, functions that are continuous modulo a null set are integrable.

Although these results are quite satisfying, we have yet another thing to consider.

(c) How do we calculate the integrals? In particular, how can we compute the Jordan content of the sets described in (a)?

This section provides a useful method for (c), by *slicing* the domain under consideration. That is, we *reduce the dimension of the integral* until we obtain an 1-dimensional integral, which can be evaluated easily. The example we are going to use to illustrate this idea is the unit ball - we do not specify if it is open or closed, as they differ by a null set. We shall see that null sets play no role in the computations of integrals and Jordan contents.

(2.30) Recall that the notion of Jordan content is introduced by the map $\tilde{\mathrm{vol}}_n : \mathcal{J}_n \to [0, \infty)$, which assigns the corresponding Jordan content to each $A \in \mathcal{J}_n$. We can also easily prove that, if $N \in \mathcal{J}_n$ is a null set, then $\tilde{\mathrm{vol}}_n(N) = 0$, giving us the inclusion

$$\{N \subseteq \mathcal{J}_n : N \text{ is a null set }\} \subseteq \left\{A \subseteq \mathcal{J}_n : \tilde{\mathrm{vol}}_n(A) = 0\right\}. \qquad [2.3]$$

Proof. Suppose that $N \in \mathcal{J}_n$ is a null set and suppose, for the sake of contradiction, that $\tilde{\mathrm{vol}}_n(N) = \varepsilon$ for some $\varepsilon > 0$. Since N is a null set, there exists $A \in \hat{\mathcal{P}}_n$ such that $A \supseteq N$ and $\hat{\mathrm{vol}}_n(A) < \varepsilon$. This leads to inequalities

$$\varepsilon = \tilde{\mathrm{vol}}_n(N) \leq \tilde{\mathrm{vol}}_n(A) = \hat{\mathrm{vol}}_n(A) < \varepsilon,$$

which is a contradiction. Thus $\tilde{\mathrm{vol}}_n(N) = 0$.

In fact, although we are not going to prove here, [2.3] holds with an equality. Now let us formalize the idea that null sets do not matter in calculating Jordan contents.

Proposition 2.29. *Let $A, B \subseteq \mathbb{R}^n$ be such that*

(a) $A \setminus B, B \setminus A$ are null sets; and

(b) A is Jordan measurable.

Then B is also Jordan measurable, and, in particular

$$\tilde{\mathrm{vol}}_n(B) = \tilde{\mathrm{vol}}_n(A).$$

Proof. Consider the union $A \cup B$. We can write this union as a disjoint union

$$A \cup B = A \cup (B \setminus A).$$

But A is Jordan measurable $B \setminus A$ is a null set, so $A \cup B = A \cup (B \setminus A)$ is Jordan measurable and

$$\tilde{\text{vol}}_n (A \cup B) = \tilde{\text{vol}}_n (A) + \tilde{\text{vol}}_n (B \setminus A) = \tilde{\text{vol}}_n (A).$$

It follows that $B = (A \cup B) \setminus (A \setminus B)$ is also Jordan measurable, and

$$\tilde{\text{vol}}_n (B) = \tilde{\text{vol}}_n (B) + \tilde{\text{vol}}_n (A \setminus B) = \tilde{\text{vol}}_n (A \cup B) = \tilde{\text{vol}}_n (A),$$

as desired. ∎

We also have an analogous result for the integrals.

Proposition 2.30. *Let $A, B \in \mathcal{J}_n$ be such that $A \setminus B, B \setminus A$ are null sets, and let $f : A \to \mathbb{R}, g : B \to \mathbb{R}$ be bounded. If we further assume f to be integrable and f, g coincide on $A \cap B$ (i.e. $f(x) = g(x)$ for all $x \in A \cap B$), then g is also integrable. In particular,*

$$\int_A f = \int_B g.$$

Lemma 2.30.1. *Let $N \in \mathcal{J}_n$ be a null set. Then for any bounded $f : N \to \mathbb{R}$,*

$$\int_N f = 0.$$

Proof. See appendix. ∎

Proof of Proposition 2.30. See appendix. ∎

(2.31) Integration of Integrals by Slicing

Let us introduce the framework we are going to use throughout the remaining part of this section. Fix dimensions $n, p, q \in \mathbb{N}$ such that $n = p + q$ and

$$S = \prod_{i=1}^{n} (a_i, b_i] \in \mathcal{P}_n,$$

then we can write $S = P \times Q$, where

$$\begin{cases} P &= \prod_{i=1}^{p} (a_i, b_i] \in \mathcal{P}_p \\ Q &= \prod_{i=p+1}^{n} (a_i, b_i] \in \mathcal{P}_q \end{cases}.$$

That is, given any $x \in S$, we can write x as

$$x = (v, w)$$

where $v \in P, w \in Q$. We are ready to discuss the method of slicing, which is justified by the following theorem. In fact, we shall soon see that this theorem we are about to present is a special case of the Fubini-Tonelli theorem (Theorem 2.33).

Proposition 2.31. Principle of Cavalier

Consider (2.31) and let $M \subseteq S$. For every $w \in Q$, let us consider the slice of M defined as

$$M_w = \{v \in P : (v,w) \in M\} \subseteq P.$$

Further suppose that

(a) M is Jordan measurable; and

(b) for every $w \in Q$, M_w is Jordan measurable.

Then by defining $F : Q \to \mathbb{R}$ by

$$w \mapsto \tilde{\text{vol}}_n (M_w)$$

for all $w \in Q$, F is integrable on Q and in particular

$$\int_Q F \, \mathrm{d}w = \tilde{\text{vol}}_n (M).$$

In order to how Theorem 2.31 works, consider applying it to calculating the volume of unit balls in various dimensions.

(EX 2.32) Volume of the Unit Open Ball

For convenience, let us denote

$$B_n = B(0;1) = \{x \in \mathbb{R}^n : \|x\| < 1\}$$

for all $n \in \mathbb{N}$, the unit open ball in $\mathbb{R}^n$. We know that B_n is Jordan measurable, so it makes sense to discuss the Jordan content, which we denote as

$$V_n = \tilde{\text{vol}}_n (B_n).$$

For instance, we have $V_1 = 2, V_2 = \pi, V_3 = \frac{4\pi}{3}$. But what about V_4 and V_5? By using Theorem 2.31, we shall obtain a recursive formula for calculating V_n for all $n \in \mathbb{N}$. It is useful to record the following observation.

Proposition 2.32.

For any $n \in \mathbb{N}$ and $r > 0$, the open ball $B(0;r) \subseteq \mathbb{R}^n$ is Jordan measurable and

$$\tilde{\text{vol}}_n (B(0;r)) = r^m V_m.$$

Proof. See appendix. ∎

We now claim that, for every $n \geq 2$, one has

$$\frac{V_n}{V_{n-1}} = \int_{-1}^{1} \left(1 - t^2\right)^{\frac{n-1}{2}} \mathrm{d}t.$$

We can verify this by using Proposition 2.32.

Proof. We enclose $M = B_n$ by $S = (-1,1]^n$, which can be written as $S = P \times Q$, where $P = (-1,1]^{n-1}$ and $Q = (-1,1]$. Then for every $w \in Q$, the slice of M is

$$\begin{aligned} M_w &= \{v = (v_1, \ldots, v_{n-1}) \in P : (v_1, \ldots, v_{n-1}, w) \in M = B_n\} \\ &= \{v = (v_1, \ldots, v_{n-1}) \in P : \|(v,w)\| \leq 1\} = \left\{v \in \mathbb{R}^{n-1} : \|v\| < \sqrt{1 - w^2}\right\}. \end{aligned}$$

Then by Proposition 2.32, one sees that M_w is Jordan measurable and

$$\tilde{\text{vol}}_{n-1} (M_w) = \left(1 - w^2\right)^{\frac{n-1}{2}} V_{n-1}$$

for all $w \in (-1,1]$. Furthermore, notice that the conditions (a), (b) of Theorem 2.31 are satisfied, so one has

$$V_n = \int_Q \tilde{\text{vol}}_n(M_w)\,\mathrm{dw} = \int_{-1}^{1} \left(1-w^2\right)^{\frac{n-1}{2}} V_{n-1}\,\mathrm{dw},$$

rearranging which gives the desired equality. ∎

Theorem 2.33. Fubini-Tonelli Theorem

Consider (2.31) again and let $f : S \to \mathbb{R}$ be bounded. For every $w \in Q$, define the slice $f_w : P \to \mathbb{R}$ by

$$f_w(v) = f((v,w))$$

for all $v \in P$. If

(a) f is integrable on S; and

(b) f_w is integrable on P for every $w \in Q$,

then, by defining $F : Q \to \mathbb{R}$ by $F(w) = \int_P f_w(v)\,\mathrm{dv}$, F is integrable on Q. In particular,

$$\int_Q F(w)\,\mathrm{dw} = \int_S f(x)\,\mathrm{dx}.$$

Proof. See appendix. ∎

(2.33) One of the consequence of Theorem 2.33 is that,

$$\int_S f(v,w)\,\mathrm{d}(v,w) = \int_S f(x)\,\mathrm{dx} = \int_Q F(w)\,\mathrm{dw} = \int_Q \left(\int_S f_w(v)\,\mathrm{dv}\right)\mathrm{dw} = \int_Q \int_P f(v,w)\,\mathrm{dv}\,\mathrm{dw}. \quad [2.4]$$

It is important to keep in mind we can write [2.4] without forcing v to be the first p components of x. As a consequence, we have

$$\int_Q \int_P f(v,w)\,\mathrm{dv}\,\mathrm{dw} = \int_P \int_Q f(v,w)\,\mathrm{dw}\,\mathrm{dv}.$$

This can be particularly useful when evaluating certain integrals, as the next example shows.

(EX 2.34) Consider $S = (0,1]^2$ and $f : S \to \mathbb{R}$ by

$$f(s,t) = \begin{cases} e^{-s^2} & \text{if } t \le s \\ 0 & \text{if } t > s \end{cases}$$

It is an easy task to show that f is continuous modulo a null set (an example of such null set would be $\{(s,t) \in S : s = t\}$) so f is integrable. So we may integrate f by slicing. However, if we begin our iterated integral with respect to s,

$$\int_0^1 \int_t^1 e^{-s^2}\,\mathrm{ds}\,\mathrm{dt},$$

we see an immediate problem: $\int e^{-s^2}\,\mathrm{ds}$ cannot be expressed in terms of elementary functions. On the other hand, if we begin with t, we obtain

$$\int_0^1 \int_0^s e^{-s^2}\,\mathrm{dt}\,\mathrm{ds} = \int_0^1 se^{-s^2}\,\mathrm{ds} = -\frac{1}{2}e^{-s^2}\Big|_0^1 = \frac{e-1}{2e}.$$

(2.35) Here we record how Theorem 2.31 is a particular case of Theorem 2.33, as mentioned in (2.31). In the given setting, let $f : S \to \mathbb{R}$ be the indicator function

$$f(x) = \begin{cases} 1 & \text{if } x \in M \\ 0 & \text{otherwise} \end{cases}$$

for all $x \in S$. Now, we verify that two conditions described in Theorem 2.31 impiles the two conditions in Theorem 2.33.

This page intentionally left blank.

3. Differentiation

Directional Derivatives

(3.1) We shall proceed to discuss about functions with more regularity properties. In fact, we are going to discuss about *differentiability* of functions in particular. The first trick we are going to use is to utilize the 1-dimensional differentiability – which we already know – in a very natural way. That is, we are going to choose a *direction* and see if the function is differentiable in 1-dimensional sense. Fix $n \in \mathbb{N}$, a dimension, throughout this chapter.

Def'n. Directional Derivative of a Function
Let $A \subseteq \mathbb{R}^n$, let $a \in \operatorname{int}(A)$, and let $f : A \to \mathbb{R}^n$. Let us also consider a *direction* vector $v \in \mathbb{R}^n$.

(a) It is convenient to have the following notation. Let $\varphi : (-c, c) \to \mathbb{R}$ be an *1-dimensional reduction* of f in direction v around the point a. That is, we define φ by

$$\varphi(t) = f(a + tw)$$

for all $t \in (-c, c)$. Note that we can define such for some $c > 0$ by the fact that a is an interior point of A.

(b) Suppose that we have defined φ as described in (a). If φ is differentiable at 0, then $\varphi'(0)$ is called the ***directional derivative*** of f at a in direction of v, denoted by $\partial_v f(a)$.

(3.2) Notice that the choice of 1-dimensional reduction φ in the above definition is not unique. For, we have (infinitely) many choices of working $c > 0$. So we have to check if, given $a \in \operatorname{int}(A)$ and $v \in \mathbb{R}^n$, the above definition well-defines the directional derivative $\partial_v f(a)$. But we know that this is the case, since, given any function $g : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$ that is differentiable at some point $b \in I$, (of course, such b is an interior point of I), for any $c > 0$ such that $(b-c, b+c) \subseteq I$, the restriction $\tilde{g} : (b-c, b+c) \to \mathbb{R}$ by $\tilde{g}(x) = g(x)$ for all $x \in (b-c, b+c)$ is also differentiable at b, and precisely we have

$$\tilde{g}'(b) = g'(b).$$

Thus we see that $\varphi'(0)$ is the same regardless of the choice of c, so $\partial_v f(a)$ is well-defined under the conditions specified in the above defintion. Furthermore, this observation allows one to write

$$\partial_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} \tag{3.1}$$

whenever the above limit exists. That is, f has the directional derivative at a in direction of v whenever the limit in [3.1] exists, and the value is precisely the limit.

(3.3) One observes that we have not put any restrictions on the directional vector v in the above definition. In particular, one can use $v = 0$ (which is a quite trivial case). That is, given $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}^n$, $\partial_0 f(a)$ exists at every $a \in \operatorname{int}(A)$, and we precisely know its value: $\partial_0 f(a) = 0$.

Proof. Notice that

$$\lim_{t \to 0} \frac{f(a + t0) - f(a)}{t} = \lim_{t \to 0} \frac{0}{t} = \lim_{t \to 0} 0 = 0.$$

∎

The reason for not putting restriction on v is as follows: with $f : A \to \mathbb{R}$ and $a \in \operatorname{int}(A)$ being fixed, we shall show that $L_a : \mathbb{R}^n \to \mathbb{R}$ defined by

$$L_a(v) = \partial_v f(a)$$

is a linear transformation. But quite often, people only focus on directional derivatives with unit directional vectors (i.e. vectors of norm 1). This is because of the homogeniety described in the following proposition, along with the fact that any $w \in \mathbb{R}^n$ can be written as $w = \alpha v$ for some $v \in \mathbb{R}^n$ with $\|v\| = 1$ and $\alpha \in \mathbb{R}$.

Proposition 3.1. Homogeniety of Directional Derivatives

Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $a \in \operatorname{int}(A)$.

(a) *$\partial_0 f(a)$ exists and $\partial_0 f(a) = 0$.*

(b) *Let $v \in \mathbb{R}^n$ be nonzero and let $\alpha \in \mathbb{R}$. If $\partial_v f(a)$ exists, then $\partial_{\alpha v} f(a)$ exists. In particular, one has*

$$\partial_{\alpha v} f(a) = \alpha \partial_v f(a).$$

Proof.

(a) See (3.3).

(b) Suppose thatt $\partial_v f(a)$ exists. We may assume $\alpha \neq 0$, since this case is done by (a). Then,

$$\begin{aligned}\lim_{t\to 0} \frac{f(a + t(\alpha v)) - f(a)}{t} &= \lim_{t\to 0} \frac{f(a + \alpha t v) - f(a)}{t} = \lim_{t\to 0} \frac{f(a + \alpha t v) - f(a)}{\alpha t} \alpha \\ &= \alpha \lim_{t\to 0} \frac{f(a + \alpha t v) - f(a)}{\alpha t} = \alpha \lim_{t\to 0} \frac{f(a + t v) - f(a)}{t} = \alpha \partial_v f(a).\end{aligned}$$

Thus $\partial_{\alpha v} f(a) = \alpha \partial_v f(a)$, as desired. ∎

(3.4) The next result that we desire to show is the mean value theorem, in the direction of a directional vector $v \in \mathbb{R}^n$. For this, we are going to utilize the following notion.

> **Recall. Line Segment** in $\mathbb{R}^n$
> For any $x, y \in \mathbb{R}$, we call the set
> $$\operatorname{Co}(x, y) = \{(1-t)x + ty : t \in [0,1]\} \subseteq \mathbb{R}^n$$
> the ***line segment*** in $\mathbb{R}^n$ with endpoints x, y.

Theorem 3.2. Mean Value Theorem in the Direction of v

Let $a \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$, and let $v \in \mathbb{R}^n$ with $\|v\| = 1$. Suppose that $\partial_v f(a)$ exists for all $a \in A$, and further suppose that $x, y \in A$ are distinct points with the following properties.

(a) *$\operatorname{Co}(x, y) \subseteq A$.*

(b) *The direction from x to y is given by v. That is, there exists $\alpha > 0$ such that $y - x = \alpha v$.*

Then there exists $z \in \operatorname{Co}(x, y)$ such that

$$\frac{f(y) - f(x)}{\|y - x\|} = \partial_v f(z).$$

Proof. We are going to utilize the 1-dimensional mean value theorem. Consider a helper function $u : [0,1] \to \mathbb{R}$ defined by

$$u(t) = f((1-t)x + ty)$$

for all $t \in [0,1]$. Direct inspection shows that $u(0) = f(x), u(1) = f(y)$, which gives us the idea that we can try to go with

$$f(y) - f(x) = \frac{u(1) - u(0)}{1 - 0}$$

and then use the 1-dimensional mean value theorem. To build on this idea, we consider the following claims.

(a) u is differentiable with

$$u'(t)\, \alpha \partial_v f((1-t)x - ty)$$

for all $t \in [0,1]$ (of course, this means one-sided differentiability at $0,1$).

Proof. Fix $t_0 \in [0,1]$. It is convenient to denote

$$a = (1-t_0)x + t_0 y \in A$$

and it is direct from the definition that $u(t_0) = f(a)$. On the other hand, if we pick an increment $\tau \in \mathbb{R} \setminus \{0\}$ with $|\tau|$ small enough to ensure $t_0 + \tau \in [0,1]$, we have

$$\begin{aligned} u(t_0 + \tau) &= f((1 - t_0 - \tau)x + (t_0 + \tau)y) \\ &= f((1-t_0)x + t_0 y + \tau(y - x)) = f(a + \tau(\alpha v)) = f(a + \alpha\tau v). \end{aligned}$$

Hence, for τ above, we have the Newton quotient

$$\frac{u(t_0 + \tau) - u(t_0)}{\tau} = \frac{f(a + \alpha\tau v) - f(a)}{\tau}. \tag{3.2}$$

To show that u is differentiable at t_0, one has to see if the limit of the Newton quotient in [3.2] as $\tau \to 0$ exists. But clearly,

$$\lim_{\tau \to 0} \frac{f(a + \alpha\tau v) - f(a)}{\tau} = \alpha \lim_{\tau \to 0} \frac{f(a + \tau v) - f(a)}{\tau} = \alpha \partial_i f(a),$$

where the second equality holds by the assumption that $\partial_v(a)$ exists. Thus u is differentiable at τ_0 with

$$u'(t_0) = \alpha \partial_i f(a).$$

∎

(b) There exists $c \in (0,1)$ such that $u(1) - u(0) = u'(c)$.

Proof. Recall that the sufficient condition for a functoin to ensure this mean value property is that u is continuous on $[0,1]$ and differentiable on $(0,1)$. But we have that u is differentiable on $[0,1]$ by (a), so the conditions are met. ∎

(c) Let $c \in (0,1)$ be as described in (b) and put

$$z = (1-c)x + cy \in \mathrm{Co}(x,y).$$

Then we have

$$\frac{f(y) - f(x)}{\|y - x\|} = \partial_i f(z).$$

Proof. We now can use the displayed equation before (a). That is,

$$f(y) - f(x) = \frac{u(1) - u(0)}{1 - 0} = u(1) - u(0) = u'(c) = \partial_v f((1-c)x + cy) = \alpha \partial_i f(z).$$

So

$$\partial_v f(z) = \frac{f(y) - f(x)}{\alpha} = \frac{f(y) - f(x)}{\|y - x\|},$$

based on the fact that $\|v\| = 1$, which implies $\alpha = \|y - x\|$. ∎

(3.5) We have some subtile points regarding Theorem 3.2, which we should make a note of.

(a) Although we have shown that u is differentiable on $[0,1]$, at endpoints $0,1$ we mean the existence of one-sided derivatives. But we have successfully ddealt with these cases by the choice of the increment τ: notice that the presented proof of Theorem 3.2 chooses τ to be always positive at 0 and always negative at 1.

(b) Note that, unlike the 1-dimensional mean value theorem, the stated equation

$$\frac{f(y) - f(x)}{\|y - x\|} = \delta_v f(z)$$

is not symmetric in x, y. This is because the directional vector v is fixed, in a way that $\alpha v = y - x$.

(3.6) Partial Derivatives

Since we have defined directional derivatives first, it would be natural to introduce partial derivatives as some special directional derivatives. That is, consider the following special vectors,

$$e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1) \in \mathbb{R}^n,$$

that form the standard ordered basis for $\mathbb{R}^n$. We are going to define the partial derivative of a function at a point to be the directional vectors in direction of $e_1, \ldots, e_n$, respectively, provided that they exist.

Def'n. Partial Derivative, Gradient of a Function

Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $a \in \operatorname{int}(A)$.

(a) For each $i \in \{1, \ldots, n\}$, we define the ***ith partial derivative*** of f at a, denoted by $\partial_i f(a)$, to be

$$\partial_i f(a) = \partial_{e_i} f(a),$$

provided that $\partial_{e_i} f(a)$ exists.

(b) If $\partial_i f(a)$ exists for all $i \in \{1, \ldots, n\}$, then we define the ***gradient*** of f at a, denoted by $\nabla f(a)$ to be

$$\nabla f(a) = (\partial_1 f(a), \ldots, \partial_n f(a)).$$

We have an alternative, computational-friendly way of defining the partial derivatives. That is, for each $i \in \{1, \ldots, n\}$, analogous to how we defined the 1-dimensional reduction of f at a in direction of a direction veector, we let $u : (a_i - c, a_i + c) \to \mathbb{R}$ to be defined by

$$u(t) = f((a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n))$$

for all $t \in (a_i - c, a_i + c)$ for some small $c > 0$ such that $(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n) \in A$ for all $t \in (a_i - c, a_i + c)$. Then we find that $\partial_i f(a)$ exists if and only if u is differentiable at a, and when this is the case,

$$\partial_i f(a) = u'(a). \qquad [3.3]$$

The next example demonstrates why [3.3] provides convenience in computation of partial derivatives, along with some motivation about why would we care about the linearity – although we have not proven this fact yet – of directional vectors, as mentioned in (3.3).

(EX 3.7) We are going to work with the 2-dimensional Euclidean space here. Let

$$A=\left\{(s,t)\in\mathbb{R}^2 : s,t\geq 0\right\}\subseteq\mathbb{R}^2,$$

which is open, and let $f : A\to\mathbb{R}$ by

$$(s,t)\mapsto s^t$$

for all $(s,t)\in A$. Let $v=(5,2)$ and let $a=(2,3)$.

(a) Find $\partial_1 f(a)$.

Answer. Let $u : (1,3)\to\mathbb{R}$ be defined by

$$u(s)=f((s,3))=s^3$$

for all $s\in(1,3)$. Clearly $(s,3)\in A$ for all $s\in(1,3)$. Moreover, u is differentiable at 2, with

$$u'(2)=3(2)^2=12.$$

Thus $\partial_1 f(a)=12$. ∎

(b) Find $\partial_2 f(a)$.

Answer. Let $u : (2,4)\to\mathbb{R}$ be defined by

$$u(t)=f((2,t))=2^t$$

for all $t\in(2,4)$. Clearly $(2,t)\in A$ for all $t\in(2,4)$. Moreover, u is differentiable at 3, with

$$u'(3)=\ln(2)2^3=8\ln(2).$$

Thus $\partial_2 f(a)=8\ln(2)$. ∎

(c) Find $\partial_v f(a)$ directly using the definition.

Answer. Let $\varphi : \left(-\frac{1}{5},\frac{1}{5}\right)\to\mathbb{R}$ be defined by

$$\varphi(t)=f(a+tv)$$

for all $t\in\left(-\frac{1}{5},\frac{1}{5}\right)$. Clearly $a+tv\in A$ for all $t\in\left(-\frac{1}{5},\frac{1}{5}\right)$. Now, to show that φ is differentiable at 0, we have

$$\begin{aligned}\lim_{t\to 0}\frac{\varphi(t)-\varphi(0)}{t}&=\lim_{t\to 0}\frac{f(a+tv)-f(a)}{t}=l\lim_{t\to 0}\frac{(2+5t)^{3+2t}-2^3}{t}=\lim_{t\to 0}\frac{e^{(3+2t)\ln(2+5t)}-2^t}{t}\\&=\lim_{t\to 0}\left(2\ln(2+5t)+\frac{15+10t}{2+5t}\right)(2+5t)^{3+2t}=16\ln(2)+60.\end{aligned}$$

Thus $\partial_v f(a)$ exists, and $\partial_v f(a)=16\ln(2)+60$. ∎

(d) Find $\partial_v f(a)$, by assuming

$$\partial_v f(a) = \partial_{5e_1+2e_2} f(a) = 5\partial_{e_1} f(a) + 2\partial_{e_2} f(a) = 5\partial_1 f(a) = 2\partial_2 f(a).$$

Answer. We have computed $\partial_1 f(a) = 12, \partial_2 f(a) = 8\ln(2)$ in (a), (b). Thus

$$\partial_v f(a) = 5\partial_1 f(a) + 2\partial_2 f(a) = 60 + 16\ln(2).$$

∎

Notice that we have the same answer for (c), (d), but the computation for (d) is comparably shorter, even when (a), (b) are considered together.

Continuously Differentiable Functions

(3.8) Now that we have discussed about derivatives, it would be natural to introduce what it means for a multivariate fucntion to be differentiable. But unlike the univariate case, the plain notion of multivariate differentiability is rather uncomfortable, and it is more practical to use the related – and stronger – notion of *continuously differentiable* function. We shall see in the next section how this notion relates to the plain differentiability.

Def'n. Continuously Differentiable (C^1-) Function
Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$. Suppose that, for all $a \in A$, the partial derivatives $\partial_1 f(a), \ldots, \partial_n f(a)$ exist, which gives rise to functions $\partial_1 f, \ldots, \partial_n f : A \to \mathbb{R}$ by the mappings

$$a \mapsto \partial_1 f(a), \ldots, a \mapsto \partial_n f(a).$$

If $\partial_1 f, \ldots, \partial_n f$ are continuous on A, then we say f is a ***continuously differentiable*** (or C^1-) function.

The main result of this section is the following theorem.

Theorem 3.3. *Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}$ be of class C^1. Then given any $a \in A$,*

$$\lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, \nabla f(a) \rangle}{\|x - a\|} = 0. \qquad [3.4]$$

(3.9) Before we get into the proof of Theorem 3.3, let us first analyze what it actually means. In fact, [3.4] is our first usage of limits of multivariate functions. But how could we disccus a large partion of multivariate calculus without using limits of functions? The answer is that, we successfully get rid of (and will continue to do so) limits of functions by using limits of sequences, which can easily substitute limits of functions in most cases. When stated in terms of sequences, [3.4] becomes the following.

(a) For any sequence $(x_k)_{k=1}^{\infty}$ on A such that $x_k \neq a$ for all $k \in \mathbb{N}$ and such that $\lim_{k\to\infty} x_k = a$, one has that

$$\lim_{k \to \infty} \frac{f(x_k) - f(a) - \langle x_k - a, \nabla f(a) \rangle}{\|x_k - a\|} = 0.$$

In order to not have to separate the case $x = a$ from $x \neq a$, we can give an $\varepsilon - \delta$ condition equivalent to [3.4] and (a) as follows.

(b) For every $\varepsilon > 0$, there exists $\delta > 0$ such that $B(a;\delta) \subseteq A$ and that

$$x \in B(a;\delta) \implies |f(x) - f(a) - \langle x-a, \nabla f(a)\rangle| \leq \varepsilon \|x-a\|.$$

Since descriptions so far are rather abstract, let us consider the following setting. Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$, and let $a \in A$. Suppose that we are given $\varepsilon > 0$, and let us aggre that we write $s \approx t$ whenever $s,t \in \mathbb{R}$ are such that $|s-t| < \varepsilon$. Then rearranging (c) gives us the following reulst: there eixsts $\delta > 0$ such that $B(a;\delta) \subseteq A$ and that

$$x \in B(a;\delta) \implies f(x) \approx f(a) + \langle x-a, \nabla f(a)\rangle.$$

Now, write $a = (a_1,\dots,a_n) \in \mathbb{R}^n$, $f(a) = b \in \mathbb{R}$, $\nabla f(a) = (v_1,\dots,v_n) \in \mathbb{R}^n$. Then for any $x = (x_1,\dots,x_n) \in \mathbb{R}^n$,

$$f(a) + \langle x-a, \nabla f(a)\rangle = b + \sum_{i=1}^n (x_i - a_i) v_i = \sum_{i=1}^n x_i v_i + \left(b - \sum_{i=1}^n a_i v_i\right).$$

Therefore, we arrive to the conclusion that

$$f(x) \approx \sum_{i=1}^n x_i v_i + \left(b - \sum_{i=1}^n a_i v_i\right),$$

so [3.4] is an equation which tells us how to do a *linear approximation* of the C^1-function f near the point a.

(3.10)

(a) The C^1-condition is usually verifiable one. That is, given that the partial derivatives exist, it is not much work to prove or disprove that they are continuous by using the tools of multivariate calculus.

(b) In the defintion of C^1-function, we did not insist on the existence of directional derivatives other than partial derivatives. But here is a pleasing fact: at the end of this section, we shall see that C^1-functions indeed have directional derivatives in all directions.

The proof of Theorem 3.3 would come as a nice application of the mean value theorem. We first prove the following lemma.

Lemma 3.3.1. *Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$, and let $a \in A$. Pick $\delta > 0$ such that $B(a;\delta) \subseteq A$. Then, for every $x \in B(a;\delta)$, there exists $z_1,\dots,z_n \in B(a;\delta)$ such that*

$$f(x) - f(a) = \langle x-a, z\rangle,$$

where $z = (\partial_1 f(z_1),\dots,\partial_n f(z_n))$.

Proof. Write $a = (a_1,\dots,a_n)$ and fix $x = (x_1,\dots,x_n) \in B(a;\delta)$. The idea for this proof is that we make a transition from a to x in n steps, via gradual substitution of components. More precisely, we define points $w_0,\dots.w_n \in \mathbb{R}^n$ – which would help us in defining $z_1,\dots,z_n$ and in applying the mean value theorem – as follows:

$$\begin{cases} w_0 &= (a_1,\dots,a_n) = a \\ w_1 &= (x_1,\dots,a_n) \\ &\vdots \\ w_i &= (x_1,\dots,x_i,a_{i+1},\dots,a_n) \\ &\vdots \\ w_n &= (x_1,\dots,x_n) = x \end{cases}$$

Then it is immediate that

$$0 = \|w_i - a\| \leq \|w_1 - a\| \leq \cdots \|w_n - a\| = \|x - a\| < \delta,$$

so $w_0, \ldots, w_n \in B(a;\delta)$. Now, for every $i \in \{1, \ldots, n\}$, the line segment $\operatorname{Co}(w_{i-1}, w_i)$ is contained in $B(a;\delta)$, as both endpoints are in $B(a;\delta)$. Therefore, by Proposition 3.2, there exists $z_i \in \operatorname{Co}(w_{i-1}, w_i)$ such that

$$f(w_i) - f(w_{i-1}) = (x_i - a_i)\,\partial_i f(z_i).$$

Thus we have

$$f(x) - f(a) = f(w_i) - f(w_0) = \sum\nolimits_{i=1}^{n} f(w_i) - f(w_{i-1}) = \sum\nolimits_{i=1}^{n} (x_i - a_i)\,\partial_i f(z_i) = \langle x - a, z\rangle,$$

as required. ∎

Proof of Theorem 3.3. We are going to verify (b) of (3.9). So fix $\varepsilon > 0$ and we find the required $\delta > 0$ as follows.

(a) Let $r > 0$ be such that $B(a;r) \subseteq A$. Such r exists since A is open.

(b) Since each $\partial_i \to \mathbb{R}$ are continuous, we ahve $\delta_i > 0$ such that

$$\forall z \in A\left[\|z - a\| < \delta \implies |\partial_i f(z) - \partial_i f(a)| < \frac{\varepsilon}{n}\right].$$

(c) Let $\delta = \min(r, \delta_1, \ldots, \delta_n)$. Then it is immediate from the definition that $\delta > 0$ and that $B(a;\delta) \subseteq A$. We now claim that this δ is what we desire to find. That is,

$$x \in B(a;\delta) \implies |f(x) - f(a) - \langle x - a, \nabla f(a)\rangle| \leq \varepsilon\,\|x - a\|.$$

Proof. The result is clear when $x = a$, so assume $x \neq a$. By Lemma 3.3.1, there exists $z_1, \ldots, z_n \in B(a;\delta)$ such that

$$f(x) - f(a) = \langle x - a, (\partial_1(z_1), \ldots, \partial_n(z_n))\rangle.$$

But we have

$$\begin{aligned}
&\langle x - a, (\partial_1 f(z_1), \ldots, \partial_n f(z_n))\rangle - \langle x - a, \nabla f(a)\rangle \\
&= \sum\nolimits_{i=1}^{n} (x_i - a_i)\,\partial_i f(z_i) - \sum\nolimits_{i=1}^{n} (x_i - a_i)\,\partial_i f(a) \\
&= \sum\nolimits_{i=1}^{n} (x_i - a_i)\,(\partial_i f(z_i) - \partial_i f(a)) \\
&< \sum\nolimits_{i=1}^{n} (x_i - a_i)\,\frac{\varepsilon}{n} = \frac{\varepsilon}{n}\sum\nolimits_{i=1}^{n} (x_i - a_i) && \text{by (b)} \\
&\leq \frac{\varepsilon}{n}\left(n \max_{1 \leq i \leq n}(|x_i - a_i|)\right) = \varepsilon\,\|x - a\|_\infty \leq \varepsilon\,\|x - a\|.
\end{aligned}$$

Thus we have the desired inequality

$$f(x) - f(a) - \langle x - a, \nabla f(a)\rangle \leq \varepsilon\,\|x - a\|,$$

which holds with an equality if and only if $x = a$. ∎

Corollary 3.3.2. C^1-functions Have All Directional Derivatives

Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$ be a C^1-function, and let $a \in A$. Then the directional derivatives $\partial_i f(a)$ at a exists for all $v \in \mathbb{R}^n$. In particular,

$$\partial_v f(a) = \langle v, \nabla f(a) \rangle .$$

Proof. Since the result is clear when $v = 0$, assume $v \neq 0$. Now, for every $t \in \mathbb{R}$, choose $x = a + tv$. Then [3.4] can be writeen as

$$\lim_{t \to 0} \frac{f(a+tv) - f(a) - \langle (a+tv) - a, \nabla f(a) \rangle}{\|(a+tv) - a\|} = 0,$$

and by cancelling out few terns and using the linearity at the first place, we obtain

$$\lim_{t \to 0} \frac{f(a+tv) - f(a) - t\langle v, \nabla f(a) \rangle}{|t| \, \|v\|} = 0.$$

Since $\|v\|$ is merely a constant, we further have

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

But this means $\lim_{t \to 0} \frac{f(a+tv)-f(a)}{t}$ exists, and so we have

$$\partial_v f(a) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t} = \langle v, \nabla f(a) \rangle .$$

Differentiability

(3.11) We now begin to discuss about the plain notion of differentiability. Although we have not explicitly stated what it means for a multivariate function to be differentiable, we nevertheless worked with this notion. But before we record the definition, let us first discuss about some consequences that we desire to get from multivariate differentiability. Our motivation for this discussion would be the notion of differentiability for univariate functions.

(a) We want our definition of differentiability to be a direct generalization of univariate case. After all, we are allowd to take n, the dimension, to be any positive integer, and it would be absurd if the multivariate differentiability does not coincide with univariate differentiability when $n = 1$.

(b) There are certain properties that univariate differentiable functions have, and many of them have a natural n-dimensional generalization. For instance, we have seen that the notion of continuity generalizes to multivariate functions well, and, similar to univariate case, we want our differentiable functions to be continuous.

In univariate case, the existence of the derivative at a point is equivalent to saying that the function is differentiable at the point. So it is very tempting that, given $A \subseteq \mathbb{R}^n, f : A \to \mathbb{R}, a \in \operatorname{int}(A)$, if $\partial_i f(A)$ exists for all $v \in \mathbb{R}^n$, then we call f to be *differentiable* at a. However, a major drawback of this definition is that, a function *differentiable* in the above sense can be discontinuous at a – certainly not what we want to see. On the other hand, we have seen a slightly stronger notion: C^1-functions. It turns out that C^1 is even bit stronger than what we want. Indeed, we have recorded this fact in Proposition 3.3. That is, the property described in the proposition is the actual notion of differentiability.

Def'n. Differentiable Function
Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $a \in \operatorname{int}(A)$. We say f is ***differentiable*** at a if there exists $g \in \mathbb{R}^n$ such that

$$\lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, g\rangle}{\|x - a\|} = 0. \tag{3.5}$$

(3.12) Let $A \subseteq \mathbb{R}^n$, let $a \in \operatorname{int}(A)$, and let $f : A \to \mathbb{R}$ be differentiable at a.

(a) We have discussed about equivalent definitions of differentiability in the previous section. The only difference is that we used $\nabla f(a)$ instead of g. But, as we will soon see, whenever f is differentiable at a, $\partial_v f(a)$ exists for all $v \in \mathbb{R}^n$, and the vector g appearing in the above definition is precisely $\nabla f(a)$.

(b) One sees that the above definition coincides with the univariate differentiability when $n = 1$. That is, when $n = 1$, $g \in \mathbb{R}^1 = \mathbb{R}$ is such that

$$\lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, g\rangle}{\|x - a\|} = 0,$$

one obtains the following result:

$$\begin{aligned} 0 &= \lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, g\rangle}{\|x - a\|} = \lim_{x \to a} \frac{f(x) - f(a) - (x - a)g}{|x - a|} \\ &= \lim_{x \to a} \left| \frac{f(x) - f(a) - (x - a)g}{x - a} \right| = \lim_{x \to a} \frac{f(x) - f(a) - (x - a)g}{x - a} = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} - g. \end{aligned}$$

So we obtain that $g = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$ when $n = 1$, which is nothing more than $f'(a)$.

We now record some consequences of our definition of differentiability.

Proposition 3.4. *Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$ be a C^1-function, and let $a \in A$. Then f is differentiable at a, and, in particular, we can take $\nabla f(a)$ to be $g \in \mathbb{R}^n$ that appears in the definition of differentiability. That is,*

$$\lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, \nabla f(a)\rangle}{\|x - a\|} = 0,$$

Proof. This is a direct consequence of Proposition 3.3 and the definition of differentiability. ∎

Proposition 3.5. *Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $a \in \operatorname{int}(A)$. If f is differentiable at a, then $\partial_v f(a)$ exists for all $v \in \mathbb{R}$. In particular,*

$$\partial_v f(a) = \langle v, g\rangle$$

for all $v \in \mathbb{R}^n$, where $g \in \mathbb{R}^n$ is the vector that appears in the definition.

Proof. A possible proof of Proposition 3.5 is very similar to the presented proof of Proposition 3.3. ∎

Corollary 3.5.1. *Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$ be differentiable, and let $a \in \operatorname{int}(A)$. If $g_1, g_2 \in \mathbb{R}^n$ are such that*

$$\lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, g_1\rangle}{\|x - a\|} = \lim_{x \to a} \frac{f(x) - f(a) - \langle x - a, g_2\rangle}{\|x - a\|} = 0,$$

then $g_1 = g_2$. In other words, $g \in \mathbb{R}^n$ appearing in the definition of differentiability is unique.

Proof. From Proposition 3.5, we have

$$\langle v, g_1 \rangle = \partial_v f(a) = \langle v, g_2 \rangle$$

for all $v \in V$. In particular, for each $i \in \{1, \ldots, n\}$,

$$\langle e_i, g_1 \rangle = \langle e_i, g_2 \rangle .$$

But $\langle e_i, \cdot \rangle : \mathbb{R}^n \to \mathbb{R}$ defines a linear functional that gives the ith component of a vector for all $i \in \{1, \ldots, n\}$, so it must be the case that $g_1 = g_2$. ∎

(3.13) Let $A \subseteq \mathbb{R}^n$, let $a \in \operatorname{int}(A)$, and let $f : A \to \mathbb{R}$ be differentiable at a. So far we have been talking about differentiability of a function, but we do not have the notion that corresponds to the derivative of univariate differentiable functions. But [3.5] and (b) of (3.12) suggest to take that appears in [3.5] to be the derivative of f at a. Moreover, we have seen that f has directional derivatives in all directions in Proposition 3.5, so $\nabla f(a)$ exists. Then a verification analogous to the presented proof of Proposition 3.3 shows that $g = \nabla f(a)$: whenever f is differentiable at a,

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

That is, given open $A \subseteq \mathbb{R}^n$ and differentiable $f : A \to \mathbb{R}$, the function $\nabla f : A \to \mathbb{R}$ defined by

$$a \mapsto \nabla f(a)$$

plays the role of derivative. Note that this ∇f is well-defined by Proposition 3.5. The next thing that we desire to prove is that every differentiable function is continuous.

Proposition 3.6. Differentiability Implies Continuity

Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $a \in \operatorname{int}(A)$. If f is differentiable at a, then f is continuous at a.

Proof. We have the following claims.

(a) There exists $\delta > 0$ such that $B(a; \delta) \subseteq A$ and that

$$|f(x) - f(a)| \leq \|x - a\| (1 + \|\nabla f(a)\|) .$$

for all $x \in B(a; \delta)$.

Proof. Since f is differentiable at a, there exists $\delta > 0$ such that $B(a; \delta) \subseteq A$ and that

$$|f(x) - f(a) - \langle x - a, \nabla f(a) \rangle| \leq \|x - a\| \tag{3.6}$$

for all $x \in B(a; \delta)$. But it follows from [3.6] that

$$\begin{aligned} |f(x) - f(a)| &\leq |f(x) - f(a) - \langle x - a, \nabla f(a) \rangle| + |\langle x - a, \nabla f(a) \rangle| \\ &\leq \|x - a\| + |\langle x - a, \nabla f(a) \rangle| \\ &\leq \|x - a\| + \|x - a\| \|\nabla f(a)\| = \|x - a\| (1 + \nabla f(a)) . \end{aligned}$$ ∎

(b) Whenever $(x_k)_{k=1}^{\infty}$ is a sequence in A such that $\lim_{k\to\infty} x_k = a$, we have $\lim_{k\to\infty} f(x_k) = f(a)$.

Proof. Let $\delta > 0$ be found as in (a). Then $\lim_{k\to\infty} x_k = a$ implies that there exists $k_0 \in \mathbb{N}$ such that $\|x_k - a\| < \delta$ for all $k \geq k_0$. Then for all $k \geq k_0$,

$$|f(x_k) - f(a)| \leq \|x_k - a\| \left(1 + \|\nabla f(a)\|\right). \tag{3.7}$$

But it is clear that the right hand side of [3.7] converges to 0 as $k \to \infty$, thus by invoking the squeeze theorem, we have the desired result. ∎

Corollary 3.6.1. *Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}$ be C^1. Then f is continuous on A.*

(3.14) Space of C^p-functions

For the remaining part of this section, fix a nonempty open set $A \subseteq \mathbb{R}^n$. We denote

$$C^1(A,\mathbb{R}) = \left\{ f \in \mathbb{R}^A : f \text{ is a } C^1\text{-function} \right\}.$$

It is also customary to denote

$$C^0(A,\mathbb{R}) = \left\{ f \in \mathbb{R}^A : f \text{ is continuous} \right\}.$$

We now define $C^2(A,\mathbb{R}), C^3(A,\mathbb{R}), \ldots$ recursively. That is, given that $C^p(A,\mathbb{R})$ is defined for some $p \in \mathbb{N}$, we define

$$C^{p+1}(A,\mathbb{R}) = \left\{ f \in \mathbb{R}^A : f \text{ is differentiable and } \forall i \in \{1,\ldots,n\}\ \partial_i f \in C^p(A,\mathbb{R}) \right\}$$

It easily follows from the definition that $C^p(A,\mathbb{R}) \subseteq C^{p-1}(A,\mathbb{R})$, by using induction.

Proof. We proceed inductively. We have shown that $C^1(A,\mathbb{R}) \subseteq C^0(A,\mathbb{R})$ in Corollary 3.6.1. Now suppose that $C^k(A,\mathbb{R}) \subseteq C^{k-1}(A,\mathbb{R})$ for some $k \in \mathbb{N}$. Then for any $f \in C^{k+1}(A,\mathbb{R})$,

$$\partial_i f \in C^k(A,\mathbb{R}) \subseteq C^{k-1}(A,\mathbb{R})$$

for all $i \in \{1,\ldots,n\}$, so $f \in C^k(A,\mathbb{R})$. Thus $C^k(A,\mathbb{R}) \subseteq C^{k+1}(A,\mathbb{R})$ and the result follows from the principle of mathematical induction. ∎

(3.15) In practical terms, $f \in C^p(A,\mathbb{R})$ means the following: for any indices $1 \leq i_1, \ldots, i_p \leq n$, one can consider the iterated partial derivative $\partial_{i_1}\partial_{i_2}\cdots\partial_{i_p} f$, which is continuous. An immediate property of $C^p(A,\mathbb{R})$ is that it is an algebra.

Proposition 3.7. *Let $A \subseteq \mathbb{R}^n$ be open. Then for any $p \in \mathbb{N} \cup \{0\}$, $C^p(A,\mathbb{R})$ is an algebra over $\mathbb{R}$.*

Iterated Partial Derivatives

(3.16) Consider an open subset $A \subseteq \mathbb{R}^n$ and the space of functions $C^p(A,\mathbb{R})$. When we say $f \in C^p(A,\mathbb{R})$, we mean that the iterated derivative

$$\partial_{i_1} \cdots \partial_{i_p} f$$

exists for all $i_1, \ldots, i_p \in \{1, \ldots, n\}$. At the first glance, it seems like the order which partial derivatives appear is of importance. In other words, we are saying that, given $i_1, \ldots, i_p \in \{1, \ldots, n\}$.

$$\partial_{i_1} \cdots \partial_{i_p} f = \partial_{i_{\sigma(1)}} \cdots \partial_{i_{\sigma(p)}} f$$

need not be true for all permutation $\sigma : \{1, \ldots, p\} \to \{1, \ldots, p\}$. But, as we will see, the final result of the computation turns out to be independent of the order, and this phenomenon has a nice connection to the fact that we can swap the order of integration. The fundamental theorem of calculus would show up as a missing link between these two different results. For simplicity, we first prove the result for the case where the dimension $n = 2$. We desire to prove the following proposition.

Proposition 3.8. *Let $A \subseteq \mathbb{R}^2$ be open and let $f \in C^2(A,\mathbb{R})$. Then*

$$\partial_1 \partial_2 f = \partial_2 \partial_1 f.$$

We first record few lemmas to work towards Proposition 3.8.

Lemma 3.8.1. *Let $u : A \to \mathbb{R}$ be continuous.*

(a) u is integrable on any half-open rectangle $S \in \mathcal{P}_2$ such that $\operatorname{cl}(S) \subseteq A$.

(b) Suppose that $u : A \to \mathbb{R}$ is also continuous, and that we have

$$\int_S u(x)\, \mathrm{d}x = \int_S v(x)\, \mathrm{d}x \qquad [3.8]$$

for all half-open rectangle $S \in \mathcal{P}_2$ such that $\operatorname{cl}(S) \subseteq A$. Then $u = v$.

Proof.

(a) We aregue that u is bounded and continuous. Observe that the continuity of u is continuous on S since u is continuous on $A \supseteq S$. Moreover, observe that $u(S) \subseteq u(\operatorname{cl}(S)) \subseteq \mathbb{R}$. Since S is bounded, $\operatorname{cl}(S)$ is compact, so the continuity of u implies that $u(\operatorname{cl}(S))$ is compact, so bounded. This in term implies that $u(S) \subseteq u(\operatorname{cl}(S))$ is bounded as well, which exactly means that u is bounded. Thus u is bounded and continuous on S, so u is integrable on S by Theorem 2.27.

(b) For the sake of contradiction, suppose that $v \neq u$, which allows us to pick $a \in A$ such that $v(a) \neq u(a)$. For convenience, denote $\alpha = u(a), \beta = v(a)$. Without loss of generality, assume $\alpha > \beta$. Since u, v are continuous at a, we have $\delta_1, \delta_2 > 0$ such that

$$\forall x \in A \left[\|x - a\| < \delta_1 \implies |u(x) - u(a)| < \frac{1}{3}(\alpha - \beta) \right]$$

and that

$$\forall x \in A \left[\|x - a\| < \delta_2 \implies |v(x) - v(a)| < \frac{1}{3}(\alpha - \beta) \right].$$

Since A is open, there exists $r>0$ such that $B(a;r)\subseteq A$. Now, let $\delta=\min(\delta_1,\delta_2,r)$ and let

$$S=\left(a_1,a_1+\frac{\delta}{2}\right]\times\left(a_2,a_2+\frac{\delta}{2}\right],$$

where $(a_1,a_2)=a$. A quick calculation shows that

$$\mathrm{cl}(S)=\left[a_1,a_1+\frac{\delta}{2}\right]\times\left[a_2,a_2+\frac{\delta}{2}\right]\subseteq B(a;\delta)\subseteq B(a;r)\subseteq A.$$

Moreover, for any $x\in S$, we have that $|x-a|<\delta$, so

$$\begin{aligned}\frac{\alpha-\beta}{3}>|u(x)-u(a)| &\implies -\frac{\alpha-\beta}{3}<-|u(x)-u(a)|\leq u(x)-u(a)=(x)-\alpha\\ &\implies u(x)>\alpha-\frac{\alpha-\beta}{3}.\end{aligned}$$

and

$$\frac{\alpha-\beta}{3}>|v(x)-v(a)|\leq v(x)-v(a)=v(x)-\beta\implies v(x)<\beta+\frac{\alpha-\beta}{3}.$$

But from the fact that $\alpha>\beta$, one has

$$u(x)>\alpha-\frac{\alpha-\beta}{3}=\frac{2\alpha}{3}+\frac{\beta}{3}>\frac{2\beta}{3}+\frac{\alpha}{3}>v(x)$$

for all $x\in S$. Therefore, since (a) guarantees that u,v are integrable on S, we have

$$\int_S u(x)\,\mathrm{d}x>\left(\alpha-\frac{\alpha-\beta}{3}\right)\mathrm{vol}_n(S)>\left(\beta+\frac{\alpha-\beta}{3}\right)\mathrm{vol}_n(S)>\int_S v(x)\,\mathrm{d}x,$$

which contradicts [3.8]. Thus we conclude $u=v$. ∎

Lemma 3.8.2. *Let $f\in C^2(A,\mathbb{R})$ and let $u=\partial_1\partial_2 f:A\to\mathbb{R}$. Let $S=(a_1,a_2]\times(b_1,b_2]\in\mathcal{P}_2$ satisfy $\mathrm{cl}(S)\subseteq A$. Then n is integrable on S and, in particular,*

$$\int_S u(x)\,\mathrm{d}x=\left(f((a_2,b_2))-f((a_2,b_1))\right)-\left(f((a_1,b_2))-f((a_1,b_1))\right).$$

Proof. We separate the proof into few steps.

(a) By (a) of Lemma 3.8.1, u is integrable on S.

(b) For all $t\in(b_1,b_2]$, let $u_t:(a_1,a_2]\to\mathbb{R}$ be the horizontal slice of u. That is,

$$u_t(s)=u((s,t))$$

for all $s\in(a_1,a_2]$. This u_t *inherits* continuity and boundedness from u, so u_t is integrable.

(c) Since u,u_t are integrable, we invoke Theorem 2.33 to obtain that

$$\int_S u(x)\,\mathrm{d}x=\int_{(b_1,b_2]}\int_{(a_1,a_2]}u_t(s)\,\mathrm{d}s\,\mathrm{d}t=\int_{b_1}^{b_2}\int_{a_1}^{a_2}u_t(s)\,\mathrm{d}s\,\mathrm{d}t. \qquad [3.9]$$

(d) To evaluate what we obtained in [3.9], we proceed as follows. First denote $g = \partial_2 f : S \to \mathbb{R}$. This means

$$u = \partial_1 \partial_2 f = \partial_1 g. \tag{3.10}$$

But note the following: when we are doing horizontal slices, [3.10] manifests itself as an 1-dimensional derivative. That is, by defining the horizontal slice $g_t : (a_1, a_2] \to \mathbb{R}$ for all $t \in (b_1, b_2]$ by

$$g_t(s) = g((s,t)),$$

we obtain that

$$u_t = g_t'$$

for all $t \in (b_1, b_2]$. Since u_t is integrable on $(a_1, a_2]$,

$$\int_{a_1}^{a_2} u_t(s)\,\mathrm{ds} = g_t(a_2) - g_t(a_1) = g((a_2,t)) - g((a_1,t))$$

by the fundamental theorem of calculus. Therefore, [3.9] becomes

$$\int_S u(x)\,\mathrm{dx} = \int_{b_1}^{b_2} g((a_2,t)) - g((a_1,t))\,\mathrm{dt}. \tag{3.11}$$

(e) To evaluate [3.11], we repeat (d) in terms vertical slices of f at a_1, a_2. Define $\varphi_1, \varphi_2 : (b_1, b_2] \to \mathbb{R}$ be the vertical slice of f at a_1, a_2. That is,

$$\varphi_1(t) = f((a,t)), \varphi_2(t) = f((a_2,t))$$

for all $t \in (b_1, b_2]$. Recall how we defined g: $g = \partial_2 f$. That is, by defining $\psi_1, \psi_2 : (b_1, b_2] \to \mathbb{R}$ by

$$\psi_1(t) = g((a_1,t)), \psi_2(t) = g((a_2,t)),$$

we have that $\psi_1 = \varphi_1', \psi_2 = \varphi_2'$. Therefore,

$$\begin{aligned}
\int_S u(x)\,\mathrm{dx} &= \int_{b_1}^{b_2} g((a_2,t)) - g((a_1,t))\,\mathrm{dt} = \int_{b_1}^{b_2} \psi_2(t) - \psi_1(t)\,\mathrm{dt} \\
&= \int_{b_1}^{b_2} \psi_2(t)\,\mathrm{dt} - \int_{b_1}^{b_2} \psi_1(t)\,\mathrm{dt} = (\varphi_2(b_2) - \varphi_2(b_1)) - (\varphi_1(b_2) - \varphi_1(b_1)) \\
&= (f((a_2,b_2)) - f((a_2,b_1))) - (f((a_1,b_2)) - f((a_1,b_1))).
\end{aligned}$$

∎

Lemma 3.8.3. *Let $f \in C^2(A, \mathbb{R})$ and let $v = \partial_2 \partial_1 f$. Let $S = (a_1, a_2] \times (b_1, b_2] \in \mathcal{P}_2$ satisfy $\operatorname{cl}(S) \subseteq A$. Then v is integrable on S and, in particular,*

$$\int_S v(x)\,\mathrm{dx} = (f((a_2,b_2)) - f((a_2,b_1))) - (f((a_1,b_2)) - f((a_1,b_1))).$$

Proof. A possible proof of Lemma 3.8.3 is analogous to the presented proof of Lemma 3.8.2. ∎

Proof of Proposition 3.8. By Lemma 3.8.2 and 3.8.3,

$$\int_S \partial_1 \partial_2 f(x)\,\mathrm{dx} = (f((a_2,b_2)) - f((a_2,b_1))) - (f((a_1,b_2)) - f((a_1,b_1))) = \int_S \partial_2 \partial_1 f(x)\,\mathrm{dx}$$

for all half-open rectangle $S = (a_1, a_2] \times (b_1, b_2] \in \mathcal{P}_2$ satisfying $\mathrm{cl}(S) \subseteq A$. We also know that $\partial_1 \partial_2 f, \partial_2 \partial_1 f$ are continuous since $f \in C^2(A, \mathbb{R})$. Thus (b) of Lemma 3.8.1 applies:

$$\partial_1 \partial_2 f = \partial_2 \partial_1 f.$$

We can give a more general result, known as the *Clairaut's theorem* or *Schwart's theorem*.

Theorem 3.9. Clairaut's Theorem

Lt $A \subseteq \mathbb{R}^n$ be open with $n \geq 2$ and let $f \in C^2(A, \mathbb{R})$. Then for any $i, j \in \{1, \dots, n\}$, one has

$$\partial_i \partial_j f = \partial_j \partial_i f.$$

Corollary 3.9.1. Symmetry of Mixed Partials

Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^p(A, \mathbb{R})$ for some $p \in \mathbb{N}$. Then for any permutation $\sigma : \{1, \dots, p\} \to \{1, \dots, p\}$, one has

$$\partial_{i_1} \cdots \partial_{i_p} f = \partial_{i_{\sigma(1)}} \cdots \partial_{i_{\sigma(p)}} f$$

for all $i_1, \dots, i_p \in \{1, \dots, n\}$.

Hessian Matrix and Quadratic Approximation

Def'n. Hessian Matrix Associated with a Function

Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A, \mathbb{R})$, and let $a \in A$. Then we define the ***Hessian matrix*** associated with f at a, denoted as $[Hf](a) \in M_{n \times n}(\mathbb{R})$, by

$$([Hf](a))_{ij} = \partial_i \partial_j f(a).$$

(3.17) Fix an open $A \subseteq \mathbb{R}^n$ throughout this section. Let $f \in C^2(A, \mathbb{R})$ and let $a \in A$. Consider the matrix

$$H = [Hf](a) \in M_{n \times n}(\mathbb{R}).$$

(a) As a consequence of Theorem 3.9, H is a symmetric matrix (i.e. $H = H^T$).

(b) Recall that, given any $n \times n$ matrix, there exists a quadratic function from $\mathbb{R}^n$ to $\mathbb{R}$ associated with the matrix in a canonical way. That is, we define the quadratic function $q : \mathbb{R}^n \to \mathbb{R}$ associated with H by

$$q(v) = \langle Hv, v \rangle$$

for all $v \in \mathbb{R}^n$. It is immediate from the definition that

$$\langle Hv, v \rangle = \sum_{i,j=1}^n H_{ij} v_i v_j = \sum_{i,j}^n \partial_i \partial_j f(a) v_i v_j.$$

This quadratic function appears in the following theorem, which shows that any C^2-function has a quadratic approximation.

Theorem 3.10. Quadratic Approximation for C^2-functions

Let $f \in C^2(A, \mathbb{R})$ and let $a \in A$. Denote $g = \nabla f(a) \in \mathbb{R}^n$ and $H = [Hf](a) \in M_{n\times n}(\mathbb{R})$. Then one has

$$\lim_{x \to a} \frac{f(x) - f(a) - \langle x-a, g\rangle - \langle H(x-a), x-a\rangle}{\|x-a\|^2} = 0. \qquad [3.12]$$

For time sake, we are going to accept Theorem 3.10. A possible proof is based on the mean value theorem, and, for this time, one should apply it twice.

(3.18) Given a C^2-function, the Hessian matrix plays the role of second derivative, and this idea can be facilitated by considering the second derivative test for a multivariate function. Recall that the second derivative test for a univariate function is as follows. Let $I \subseteq \mathbb{R}$ be an open interval, let $f : I \to \mathbb{R}$, and let $a \in I$ be a critical point for f (i.e. $f'(a) = 0$). Then one has the following.

(a) If $f''(a) < 0$, then a is a point of local maximum for f.

(b) If $f''(a) > 0$, then a is a point of local minimum for f.

(c) If $f''(a) = 0$, then a is not a point of local extremum for f.

But given a C^2-function, the Hessian matrix associated with the function at any given point is not a real number unless we are working with the case $n = 1$. Therefore, we have to find corresponding notions that can replace the statements regarding the sign of the second derivative. We shall see that the positive definiteness and the negative definiteness – which are indeed well defined since Hessian matrix of a C^2-function is symmetric – are the notions that we are looking for. But first, let us begin by introducing the following defintiion.

Def'n. Point of Local Extremum for a Function

Let $f : A \to \mathbb{R}$.

(a) We say $a \in A$ is a point of ***local minimum*** for f if there exists $r > 0$ such that

$$\forall x \in A \cap B(x; r)\,[f(x) \geq f(a)].$$

(b) We say $a \in A$ is a point of ***local maximum*** for f if there exists $r > 0$ such that

$$\forall x \in A \cap B(x; r)\,[f(x) \leq f(a)].$$

One also says a point $a \in A$ is a point of ***local extremum*** if it is a point of local minimum or a point of local maximum.

Def'n. Critical Point, Regular Point of a Function

Let $f \in C^1(A, |R)$. We say $a \in A$ is a ***critical point*** of f if $\nabla f(a) = 0$. Otherwise, we say a is a ***regular point***.

Proposition 3.11. *Let $f \in C^1(A, \mathbb{R})$ and let $a \in A$ be a point of local extremum for f. Then $\nabla f(a) = 0$.*

Proof. This is a consequence of the univariate version of this proposition, since partial derivatives are essencially 1-dimensional. ∎

(3.19) Here is a bit of review of linear algebra. Consider a real symmetric matrix $A \in M_{n\times n}(\mathbb{R})$. Then A is Hermitian, so the characteristic polynomial of A splits over $\mathbb{R}$. This allows the following definition.

Recall. Positive Definite, Negative Definite Matrix

Let $A \in M_{n\times n}(\mathbb{R})$ be symmetric and let $c_1,\ldots,c_k \in \mathbb{R}$ be the distinct eigenvalues of A.

(a) We say A is ***positive definite*** if $c_i > 0$ for all $i \in \{1,\ldots,k\}$.

(b) We say A is ***negative definite*** if $c_i < 0$ for all $i \in \{1,\ldots,k\}$.

Theorem 3.12. Second Derivative Test

Let $f \in C^2(A,\mathbb{R})$ and let $a \in A$ be a critical point for f. Denote

$$H = [Hf](a) \in M_{n\times n}(\mathbb{R}),$$

the Hessian matrix associated with f at a.

(a) If H is positive definite, then a is a point of local minimum.

(b) If H is negative definite, then a is a point of local maximum.

(c) Otherwise (i.e. H has a positive eigenvalue and a negative eigenvalue), a is not a point of local extremum.

The case (c) in Theorem 3.12 is often referred as a *saddle point*: a is a saddle point for f if a is a critical point but is not a local extremum. We verify few lemmas first.

Lemma 3.12.1.

Let $A \in M_{n\times n}(\mathbb{R})$ be symmetric and let $c_1,\ldots,c_n \in \mathbb{R}$ be the eigenvalues of A, listed with multiplicities. Denote

$$m = \min_{1\le i\le n}(c_i),\, M = \max_{1\le i\le n}(c_i).$$

Then

$$m\|v\|^2 \le \langle Av, v\rangle \le M\|v\|^2$$

for all $v \in \mathbb{R}^n$.

Proof. Since A is symmetric, there is an orthonormal eigenbasis for V, say $\beta = (v_1,\ldots,v_n)$. Let $c_1,\ldots,c_n \in \mathbb{R}$ be the eigenvalues corresponding to $v_1,\ldots,v_n$, respectively, and let

$$v = \sum\nolimits_{i=1}^{n} \alpha_i v_i \in \mathbb{R}^n$$

for some $\alpha_1,\ldots,\alpha_n \in \mathbb{R}$. Then

$$\|v\|^2 = \sum\nolimits_{i=1}^{n} \alpha_i^2.$$

On the other hand,

$$\langle Av, v\rangle = \left\langle \sum\nolimits_{i=1}^{n} c_i\alpha_i v_i, \sum\nolimits_{j=1}^{n} \alpha_j v_j\right\rangle = \sum\nolimits_{i=1}^{n} c_i\alpha_i^2.$$

Thus

$$m\|v\|^2 = \min_{1\le i\le m} c_i \sum\nolimits_{i=1}^{n} \alpha_i^2 \le \sum\nolimits_{i=1}^{n} c_i\alpha_i^2 \le \max_{1\le i\le m} c_i \sum\nolimits_{i=1}^{n} \alpha_i^2 = M\|v\|^2.$$

∎

Proof of Theorem 3.12. Since f is a C^2-function, Theorem 3.10 applies, where the limit that appears in [3.12] is equivalent to the following: given any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall x \in B(a;\delta)\left[|f(x) - f(a) - \langle H(x-a), x-a\rangle| \le \varepsilon\|x-a\|^2\right]. \tag{3.13}$$

Notice that the gradient vector disappeared in [3.13]. This is not a mistake: a is a critical point for f so $\nabla f(a) = 0$. Reformulating [3.13] gives that, for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$-\varepsilon \|x-a\|^2 \le f(x) - f(a) - \langle H(x-a), x-a\rangle \le \varepsilon \|x-a\|^2 \qquad [3.14]$$

for all $x \in B(a;\delta)$.

(a) Suppose that H is positive definite and take $\varepsilon = \frac{m}{2}$, where $c_1, \ldots, c_n \in \mathbb{R}$ are the eigenvalues of H and $m = \min_{1\le i\le n}(c_i)$, listed with multiplicities. Then there exists $\delta > 0$ such that [3.14] holds for all $x \in B(a;\delta)$, and by rearranging the first inequality in [3.14], we obtain that

$$\begin{aligned} f(x) &\ge f(a) + \langle H(x-a), x-a\rangle - \varepsilon \|x-a\|^2 \\ &\ge f(a) + m\|x-a\| - \frac{m}{2}\|x-a\| = f(a) + \frac{m}{2}\|x-a\| \ge f(a). \end{aligned}$$

for all $x \in B(a;\delta)$. Therefore, a is a point of local minimum for f.

(b) This part can be done analogously to (a).

(c) We have few claims to verify.

(i) Let $c \in \mathbb{R}$ be an eigenvalue of A and let $v \in \mathbb{R}^n$ be a unit eigenvector corresponding to c. Then

$$\lim_{t\to 0} \frac{f(a+tv) - f(a)}{t^2} = c.$$

Proof. By using [3.13], we have that

$$\lim_{t\to 0} \frac{f(a+tv) - f(a) - \langle H(tv), tv\rangle}{\|tv\|^2} = 0.$$

Again, we are using the fact that $\nabla f(a) = 0$. But one has that

$$\langle H(tv), tv\rangle = \langle ctv, tv\rangle = ct^2\langle v, v\rangle = ct^2,$$

and

$$\|tv\|^2 = |t|^2\|v\|^2 = t^2.$$

Therefore,

$$\lim_{t\to 0} \frac{f(a+tv) - f(a) - ct^2}{t^2} = \lim_{t\to 0} \frac{f(a+tv) - f(a) - \langle H(tv), tv\rangle}{\|tv\|^2} = 0,$$

which means

$$\lim_{t\to 0} \frac{f(a+tv) - f(a)}{t^2} = \lim_{t\to 0} \frac{ct^2}{t^2} = c.$$

∎

(ii) Let $c \in \mathbb{R}$ be a positive eigenvalue of A and let $v \in \mathbb{R}^n$ be a unit eigenvector corresponding to c. Suppose that we are given a sequence $(x_k)_{k=1}^{\infty}$ on $\mathbb{R}^n$ of the form

$$x_k = a + t_k v,$$

where $(t_k)_{k=1}^{\infty}$ is a convergent sequence on $\mathbb{R}\setminus\{0\}$ with $\lim_{k\to\infty} t_k = 0$. Then $\lim_{k\to\infty} x_k = a$ and there exists $k_0 \in \mathbb{N}$ such that $f(x_k) > f(a)$ for all $k \ge k_0$.

Proof. It is clear that that $\lim_{k\to\infty} x_k = a$. Notice that

$$\lim_{k\to\infty}\frac{f(x_k)-f(a)}{t_k^2}=\lim_{k\to\infty}\frac{f(a+t_kv)-f(a)}{t_k^2}=\lim_{t\to 0}\frac{f(a+tv)-f(a)}{t_k^2}=c \tag{3.15}$$

by (i). Since the limit is positive and $t_k^2>0$ for all $k\in\mathbb{N}$, it must be the case that there exists $k_0\in\mathbb{N}$ such that $f(x_k)-f(a)>0$ for all $k\ge k_0$, which exactly means $f(x_k)>f(a)$ for all $k\ge k_0$. ∎

(iii) Let $c\in\mathbb{R}$ be a negative eigenvalue of A and let $v\in\mathbb{R}^n$ be a unit eigenvector corresponding to c. Suppose that we are given a sequence $(y_k)_{k=1}^\infty$ on $\mathbb{R}^n$ of the form

$$y_k = a + s_k v,$$

where $(s_k)_{k=1}^\infty$ is a convergent sequence on $\mathbb{R}\setminus\{0\}$ with $\lim_{k\to\infty} s_k = 0$. Then $\lim_{k\to\infty} y_k = a$ and there exists $k_0\in\mathbb{N}$ such that $f(y_k)<f(a)$ for all $k\ge k_0$.

Proof. This part can be verified similarly to (ii). ∎

It is clear from (ii) and (iii) that a is not a point of local extremum. ∎

Vector Valued C^1-functions and the Chain Rule

(3.20) We discuss about C^1-functions again, but in a slightly different setting: instead of loking at real valued C^1-functions, we consider more general class of functions, namely *vector valued C^1-functions*. Our goal is to obtain an analogous result of the chain rule for univariate real valued functions,

$$(f\circ g)'(b) = f'(g(b))\,g'(b). \tag{3.16}$$

A difference would be that, [3.16] is written in terms of multiplication of real numbers, whereas the chain rule for multivariate vector valued functions would involve matrix multiplication. Fix dimensions $n,m,p\in\mathbb{N}$ throughout.

> **Def'n. C^1-function**
> Let $A\subseteq\mathbb{R}^n$ be open. We say $f:A\to\mathbb{R}^m$ is a ***C^1-function*** when each component function $f_i:A\to\mathbb{R}$ of f is a C^1-function.

Given an open subset $A\subseteq\mathbb{R}^n$, we also denote the set of C^1-functions by

$$C^1(A,\mathbb{R}^m)=\left\{f\in(\mathbb{R}^m)^A : f \text{ is } C^1\right\}.$$

Recall that, given a multivariate real valued functions, the gradient vector plays the role of derivative. We can generalize this by using matrices for multivariate vector valued functions. After all, the standard inner product on $\mathbb{R}^n$ can be interpreted as a multiplication of a row vector and a column vector, and we shall see that the following definition coincides with the gradient vector when $m=1$.

Def'n. Jacobian Matrix of a Function

Let $A \subseteq \mathbb{R}^n$ be open and let $f \in C^1(A, \mathbb{R}^m)$. For every $a \in A$, we define the ***Jacobian matrix*** of f at a, denoted as $[Jf](a)$, to be

$$[Jf](a) = \begin{bmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_n f_m(a) \end{bmatrix} \in M_{m\times n}(\mathbb{R}).$$

(3.21) Jacobian Matrix

Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R})$, and let $a \in A$.

(a) One cmn define a function $[Jf] : A \to M_{m\times n}(\mathbb{R})$ by putting

$$[Jf](b) = \begin{bmatrix} \partial_1 f_1(b) & \cdots & \partial_n f_1(b) \\ \vdots & & \vdots \\ \partial_1 f_m(b) & \cdots & \partial_n f_m(b) \end{bmatrix}$$

for all $b \in A$.

(b) The Jacobian matrix of f at a, $[Jf](a)$, can be thought as *gradient vectors of component functions of f stacked together*. That is, the ith row of $[Jf](a)$ is precisely $\nabla f_i(a)$. In particular, when $m = 1$,

$$[Jf](a) = \nabla f(a).$$

(c) When $n = 1$, the Jacobian matrix $[Jf](a)$ is really a column veector of length m, where the ith entry is the derivative of ith component of f, which is a univariate real valued function. That is,

$$[Jf](a) = \begin{bmatrix} f_1'(a) \\ \vdots \\ f_m'(a) \end{bmatrix} \in M_{m\times 1}(\mathbb{R}).$$

Similar to real valued C^1-functions which has linear approximation, vector valued C^1-functions can be linearly approximated, as the following proposition shows.

Proposition 3.13. Linear Approximation of Vector Valued C^1-functions

Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R}^m)$, and let $a \in A$. Then

$$\lim_{x\to a} \frac{\|f(x) - f(a) - [Jf](a)(x-a)\|}{\|x-a\|} = 0. \qquad [3.17]$$

Proof. Similar to before, we are going to work with $\varepsilon - \delta$ reformulation of [3.17]. That is, we fix $\varepsilon > 0$ and show there exists $\delta > 0$ such that $B(a;\delta) \subseteq A$ and that

$$\|f(x) - f(a) - [Jf](a)(x-a)\| \leq \varepsilon \|x-a\|$$

for all $x \in B(a;\delta)$. Denote

$$v(x) = f(x) - f(a) - [Jf](a)(x-a)$$

for all $x \in A$. One can write each component v_i of v by

$$v_i(x) = f_i(x) - f_i(a) - \langle \nabla f_i(a), x-a \rangle$$

for all $x \in A$. Then, for each $i \in \{1, \ldots, m\}$, from the proof of Theorem 3.3, we know that there exists $\delta_i > 0$ such that

$$|v_i| \leq \frac{\varepsilon}{2m} \|x-a\|$$

for all $x \in B(a;\delta)$. Thus by taking $\delta = \min_{1\le i\le m}(\delta_i)$,

$$\|f(x)-f(a)-[Jf](a)(x-a)\| = \|v\| \le \|v\|_1 = \sum_{i=1}^m |v_i| \le \sum_{i=1}^m \frac{\varepsilon}{2m}\|x-a\| < \varepsilon\|x-a\|,$$

as required. ∎

(3.22) Differentiability of C^1-functions

Recall that, in a multivariate real valued setting, f being a C^1-function implies that f is differentiable. We have analogous result for multivariate vector valued case. But we are not going to repeat the whole discussion again; rather, we present the following proposition to show case what it would entail.

Proposition 3.14. *Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}^m$. Let $a \in A$ and suppose that we have $M \in M_{m\times n}(\mathbb{R})$ satisfying*

$$\lim_{x\to a} \frac{\|f(x)-f(a)-M(x-a)\|}{\|x-a\|} = 0. \qquad [3.18]$$

Then each component function $f_i : A \to \mathbb{R}$ is differentiable at a. In paticular,

$$M_{ij} = \partial_j f_i(a)$$

for all $i \in \{1,\ldots,m\}$, $j \in \{1,\ldots,n\}$.

Proof. Let $v : A \to \mathbb{R}^m$ be defined by

$$v(x) = f(x) - f(a) - M(x-a)$$

for all $x \in A$, so we have

$$v_i(x) = f_i(x) - f_i(a) - \mathrm{Row}_i(M)(x-a) = f_i(x) - f_i(a) - \langle \mathrm{Row}_i(M), (x-a)\rangle$$

for all $x \in A$. Since one has

$$\frac{|v_i(x)|}{\|x-a\|} \le \frac{\|v(x)\|}{\|x-a\|}$$

for all $x \in A$, by the squeeze theorem

$$\lim_{x\to a} \frac{|v_i|}{\|x-a\|} = \frac{|f_i(x)-f_i(a)-\langle \mathrm{Row}_i(M), x-a\rangle|}{\|x-a\|} = 0. \qquad [3.19]$$

since we have [3.18]. But this is precisely what it means for f_i to be differentiable at a. Furthermore, it is clear from [3.19] that

$$\nabla f_i(a) = \mathrm{Row}_i(M),$$

one has

$$M = \begin{bmatrix} \nabla f_1(a) \\ \vdots \\ \nabla f_m(a) \end{bmatrix} = \begin{bmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_n f_m(a) \end{bmatrix},$$

as claimed. ∎

Theorem 3.15. Chain Rule

Suppose that we have

(a) an open subset $A \subseteq \mathbb{R}^n$ and $f : C^1(A, \mathbb{R}^m)$; and

(b) an open subset $B \subseteq \mathbb{R}^p$ and $g \in C^1(A, \mathbb{R}^n)$.

Suppose further that image$(g) \subseteq A$. *This allows one to consider the composition*

$$u = f \circ g : B \to \mathbb{R}^m.$$

Then $u \in C^1(B, \mathbb{R}^m)$ *and, in particular,*

$$[Ju](b) = [Jf]((g(b)))[Jg](b)$$

for all $b \in B$.

Constrained Optimization and Lagrange Multiplier

(EX 3.23) Consider the following example which demonstrates the type of optimization problem that we consider in this section. Let

$$A = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\right\},$$

the unit closed ball in $\mathbb{R}^3$, and let $f : A \to \mathbb{R}$ be defined by

$$f(x,y,z) = xy^2z^2$$

for all $(x,y,z) \in A$. Since f is continuous and A is compact, the extreme value theorem tells us that we have points of global extremum of f on A. In particular, we desire to find a point of global maximum for f. How do we achieve this? The first step, arguably, would be finding out critical points for f in $\operatorname{int}(A)$, since we have

$$\begin{aligned} a \text{ is a point of global maximum} &\implies a \text{ is a point of local maximum} \\ &\implies a \text{ is a critical point} \end{aligned}$$

for all $a \in \operatorname{int}(A)$. But by a direct calculation, we can show that every critical point for f is not a point of local maximum. That is, we have to find a point of global maximum on $A \operatorname{int}(A)$. But the problem here is that, since $A \setminus \operatorname{int}(A)$ has no interior points, we cannot apply the approach that we find critical points first. To get around this, we are going to utilize the fact that $A \operatorname{int}(A)$ is a *level set* of a function $\varphi : \mathbb{R}^3 \to \mathbb{R}$. So consider developing towards this idea.

Def'n. Path
Let $I \subseteq \mathbb{R}$ be an open interval. We say $\gamma : I \to \mathbb{R}^n$ is a ***path*** if γ is continuous.

(3.24) Let $I \subseteq \mathbb{R}$ be an open interval and let $\gamma : I \to \mathbb{R}^n$ be a path. Write

$$\gamma = (\gamma_1, \ldots, \gamma_n)$$

where each γ_i is the ith component of γ.

(a) We say γ is a ***differentiable*** path if each γ_i is differentiable. We also define the ***velocity vector*** of γ, denoted as γ', to be

$$\gamma' = (\gamma_1', \ldots, \gamma_n') : I \to \mathbb{R}^n.$$

That is, $\gamma'(t) = [J\gamma](t)$ for all $t \in I$ when viewed as a column matrix.

(b) We say γ is a C^1-path if each γ_i is C^1.

Proposition 3.16. *Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}$ be C^1. Let $I \subseteq \mathbb{R}$ be an open interval and let $\gamma : I \to \mathbb{R}^n$ be a C^1-path such that* image$(\gamma) \subseteq A$. *Then*

$$u = f \circ \gamma : I \to \mathbb{R}$$

is a C^1-function, and one has

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

for all $t \in I$.

Proof. This is a direct consequence of the chain rule. ∎

Def'n. Level Set of a Function
Let $A \subseteq \mathbb{R}^n$, let $\varphi : A \to \mathbb{R}$, and let $c \in \text{image}(f)$. We define the ***level set*** of φ corresponding to c by

$$L = \{x \in A : \varphi(x) = c\}.$$

Def'n. Tangent Vector, Normal Vector to a Level Set of a C^1-function
Let $A \subseteq \mathbb{R}^n$ be open and let $\varphi \in C^1(A, \mathbb{R})$. Let $c \in \text{image}(\varphi)$ and consider the level set L of φ corresponding to c. Suppose that we have a regular point $a \in L$.

(a) We say $w \in \mathbb{R}^n$ is a ***normal vector*** to L at a if there exists $\lambda \in \mathbb{R}$ such that $w = \lambda \nabla f(a)$.

(b) We say $v \in \mathbb{R}^n$ is a ***tangent vector*** to L at a if there exists an open interval $I \subseteq \mathbb{R}$ and a C^1-path $\gamma : I \to \mathbb{R}^n$ such that

 (i) image$(\gamma) \subseteq L$; and
 (ii) $0 \in I$ and we have $\gamma(0) = a$ and $\gamma'(0) = v$.

An immediate consequence of the above definition is that $v \in \mathbb{R}^n$ is a tangent vector of L at a if and only if $v, \nabla \varphi(a)$ are orthogonal.

Proposition 3.17. *Consider the setting in the above definition. We have*

$$v \in \mathbb{R}^n \text{ is a tangent vector to } L \text{ at } a \iff \langle v, \nabla \varphi(a) \rangle = 0.$$

Proof.

(a) Consider $u : I \to \mathbb{R}$ defined as $u = \varphi \circ \gamma$. Then by Proposition 3.16, we have

$$u'(0) = \langle \nabla \varphi(\gamma(0)), \gamma'(0) \rangle = \langle \nabla \varphi(a), v \rangle.$$

But u is a constant function, since $\gamma(t) = c$ for all $t \in I$, which means

$$u(t) = \varphi(\gamma(t)) = \varphi(c)$$

for all $t \in I$. Therefore $u'(t) = 0$ for all $t \in I$, and in particular, $u'(0) = 0$. So

$$\langle \nabla \varphi(a), v \rangle = u'(0) = 0.$$

(b) We accept this direction as a fact, since it is more involved. ∎

(3.25) Tangent Vectors Are Orthogonal to Every Normal Vector

From the defintion, it is clear that

$$V = \{w \in \mathbb{R}^n : w \text{ is a normal vector to } L \text{ at } a\}$$

is a subspace of $\mathbb{R}^n$, as it is spanned by $\nabla f(a)$. On the other hand, we have that, $v \in \mathbb{R}^n$ is a tangent vector to L at a if and only if

$$\langle \nabla f(a), v \rangle = 0$$

by Proposition 3.17. But $V = \text{span}(\nabla f(a))$, which means that $v \in \mathbb{R}^n$ is a tangent vector to L at a if and only if v is orthogonal to every vector in V. That is, if we denote

$$T = \{v \in \mathbb{R}^n : v \text{ is a tangent vector of } L \text{ at } a\},$$

then $T = V^\perp$.

Corollary 3.17.1. *Consider Proposition 3.17 and (3.25). We have $T = V^\perp$.*

(3.26) To discuss about the Lagrange multiplier, consider the following setting. Let $A \subseteq \mathbb{R}^n$ be open.

(a) We are given a *constraint function* $\varphi \in C^1(A, \mathbb{R})$, and consider a level set of φ, $L = \{x \in A : \varphi(x) = c\}$ for some $c \in \text{image}(\varphi)$.

(b) We are given an *objective function* $f \in C^1(A, \mathbb{R})$.

Our problem is to find the points of local extremum for the restriction of f to L, $f|_L : L \to \mathbb{R}$.

Theorem 3.18. Lagrange Multiplier

Consider (3.26). If $a \in L$ is a point of local extremum for $f|_L$ and if a is a regular point for φ, then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(a) = \lambda \nabla \varphi(a).$$

Proof. We are going to verify the result by assuming that a is a point of local maximum for $f|_L$. The case which a is a point of local minimum for $f|_L$ can be dealt similarly. We are going to utilize Corollary 3.17.1. That is, it suffices to show that

$$\langle \nabla f(a), v \rangle = 0$$

for all $v \in \mathbb{R}^n$ which is tangent to L at a. So fix a tangent vector $v \in \mathbb{R}^n$. That is, we have an open interval $I \subseteq \mathbb{R}$ with $0 \in I$ and a C^1-path $\gamma : I \to \mathbb{R}^n$ such that $\gamma(0) = a$, $\gamma'(0) = v$, and $\text{image}(\gamma) \subseteq L$. Define

$$u = f \circ \gamma : I \to \mathbb{R}.$$

Observe the following.

(a) By using Proposition 3.16, we obtain that u is C^1 and

$$u'(t) = \left\langle \nabla f(\gamma(t)), \gamma'(t) \right\rangle$$

for all $t \in I$. In particular,

$$u'(0) = \left\langle \nabla f(\gamma(0)), \gamma'(0) \right\rangle = \langle \nabla f(a), v \rangle.$$

(b) We claim that $0 \in I$ is a point of local maximum for u.

Proof. Since $\text{image}(\gamma) \subseteq L$, we have that

$$u = f \circ \gamma = f|_L \circ \gamma$$

so

$$u(t) = (f|_L)(\gamma(t))$$

for all $t \in I$. In particular,

$$u(0) = (f|_L)(\gamma(0)) = (f|_L)(a).$$

But a is a point of local maximum for $f|_L$, so there exists $\delta_f > 0$ such that $(f|_L)(a) \geq (f|_L)(x)$ for all $x \in B(a;\delta_f) \cap L$. Since γ is continuous, it follows that $\gamma^{-1}(B(a;\delta_f) \cap L)$ is an open subset of $\mathbb{R}$, and clearly $0 \in \gamma^{-1}(B(a;\delta_f) \cap L)$. Therefore we can find $\delta > 0$ such that $(-\delta, \delta) \subseteq \gamma^{-1}(B(a;\delta_f) \cap L)$. But for any $t \in (-\delta, \delta)$, $\gamma(t) \in B(a;\delta_f) \cap L$, so

$$(f|_L)(a) \geq (f|_L)(\gamma(t))$$

for all $t \in (-\delta, \delta)$. But $\gamma(0) = a$, so

$$u(0) = f(\gamma(0)) = (f|_L)(a) \geq (f|_L)(\gamma(t)) = u(t)$$

for all $t \in (-\gamma, \gamma)$. Thus 0 is a point of local extremum for u.

By (b), we have that $u'(0) = 0$, so by (a),

$$\langle \nabla f(a), v \rangle = u'(0) = 0,$$

as desired. ∎

Integration by Substitution

(3.27) Recall that we sometimes 1-dimensional integral $\int_{a_1}^{a_2} f(x)\,\mathrm{dx}$ by saying *we proceed by a substitution* $x = \varphi(y)$, *where* $\mathrm{dx} = \varphi'(y)\,\mathrm{dy}$. The function φ is a differentialbe function from some other interval $B = (b_1, b_2)$ to $A = (a_1, a_2)$. Since we do not define the precise meaning of dx and dy, what we said above is simply a memorizing device of the following:

$$\int_{a_1}^{a_2} f(x)\,\mathrm{dx} = \int_{b_1}^{b_2} f(\varphi(y))\,\varphi'(y)\,\mathrm{dy}, \qquad [3.20]$$

which is the change of variable formula. We desire to give a multivariate version of [3.20], and the *determinant* of Jacobian matrix would play the role of $\varphi'(y)$.

(3.28) We are going to work with a bounded continuous function $f : A \to \mathbb{R}$ for some open and Jordan measurable $A \subseteq \mathbb{R}^n$. Although the multivariate version of [3.20] works on any Jordan measurable subset, we nevertheless insist on open subsets simply because we are going to work with C^1-functions somewhere, and C^1-functions are naturally defined on an open set. After all, we know that, given any Jordan measurable $A \subseteq \mathbb{R}^n$, its interior $\text{int}(A)$ is also Jordan measurable, and for any integrable function $f : A \to \mathbb{R}$, we precisely have the equality $\int_A f = \int_{\text{int}(A)} f$.

Def'n. Jacobian Determinant (Jacobian) of a C^1-function at a Point

Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}^n$ be C^1, where $m \in \mathbb{N}$. We define the ***Jacobian determinant*** (or simply ***Jacobian***) of f at a to be the determinant of Jacobian matrix of f at a, $\det([Jf](a))$, for all $a \in A$.

We write $|J|_f : A \to \mathbb{R}$ to denote the function which gives the absoute value of the Jacobian of f when evaluated at any $a \in A$,

$$[J]_f(a) = |\det([Jf](a))|$$

for all $a \in A$. It is an immediate consequece of the defintion that $|J|_f$ is continuous for any C^1-function f: the definition of the determinant allows one to write $\det([Jf](a))$ in terms of partial derivatives of f – which are continuous by defintion – by using addition, subtraction, and multiplication.

Def'n. Diffeomorphism

Let $A, B \subseteq \mathbb{R}^n$ be open. We say $f : A \to B$ is a ***diffeomorphism*** if

(a) f is bijective; and

(b) $f^{-1} : B \to A$ and f are differentiable.

Moreover, if f, f^{-1} are C^k, then we say f is a C^k***-diffeomorphism***.

Theorem 3.19. Integration by Substitution

Let $A, B \subseteq \mathbb{R}^n$ be open and Jordan measurable and let $\Phi : B \to A$ be a C^1-diffeomorphism, where $|J|_\Phi : B \to \mathbb{R}$ is bounded. Let $f : A \to \mathbb{R}$ be bounded and continuous. Then $g : B \to \mathbb{R}$ by

$$g(y) = f(\Phi(y))$$

for all $y \in B$ is also bounded and continuous, and one has equality of integrals

$$\int_A f = \int_B g.$$

That is, one has

$$\int_A f(x)\,\mathrm{d}x = \int_B f(\Phi(y))\,|J|_\Phi(y)\,\mathrm{d}y. \tag{3.21}$$

An immediate yet important result that follows from Theorem 3.19 is that one can retain the Jordan content of a Jordan measurable set by using a diffeomorphism.

Corollary 3.19.1.

Let $A, B \subseteq \mathbb{R}^n$ be open and Jordan measurable and let $\Phi : B \to A$ be a diffeomorphism with bounded $[J]_\Phi : B \to \mathbb{R}$. Then

$$\tilde{\mathrm{vol}}_n(A) = \int_B |J|_\Phi(y)\,\mathrm{d}y.$$

(3.29) Scaling a Jordan Measurable Set

The result of Corollary 3.19.1 is quite nontrivial. To demonstrate this, suppose that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation. That is, we have a unique invertible $M \in M_{n\times n}(\mathbb{R})$ such that

$$\Phi(y) = My$$

for all $y \in \mathbb{R}^n$. It turns out that, given a Jordan measurable $A \subseteq \mathbb{R}^n$, $\Phi(A)$ is also Jordan measurable, so the restriction $\Phi|_A : A \to B$ is a diffeomorphism where $B = \Phi(A)$. Therefore, we can apply Corollary 3.19.1 with respect to $\Phi|_A$. But one has that, whenever a function is a restriction of a linear transformation, its

Jacobian matrix is the matrix representation of the linear transformation with respect to the standard basis for $\mathbb{R}^n$. So one has that

$$|J|_{\Phi|_A}(y) = |[J\Phi|_A](y)| = |\det(M)|$$

for all $y \in B$. Thus,

$$\tilde{\text{vol}}_n(A) = \int_B |J|_{\Phi|_A}(y)\,\mathrm{d}y = \int_B |\det(M)|\,\mathrm{d}y = |\det(M)|\,\tilde{\text{vol}}_n(B).$$

That is, applying an invertible linear transformation to a Jordan measurable set *scales* the Jordan content by the absolute value of the determinant of the transformation.

(EX 3.30) Integrating in Polar Coordinates

An important application of Theorem 3.19 is *integrating in polar coordinates*. To establish this, consider the following setting: we are working with 2-dimensional Euclidean spac, and we have two real numbers $r_1, r_2 \in \mathbb{R}$ such that $0 < r_1 < r_2$ (think r_1, r_2 as *radii*). Further define:

(a) $A = \left\{(s,t) \in \mathbb{R}^2 : r_1 < \sqrt{s^2 + r^2} < r_2\right\} \setminus \{(s,0) : r_1 < s < r_2\}$ (i.e. *annulus with a ray removed*);

(b) $B = \left\{(r,\theta) \in \mathbb{R}^2 : r_1 < r < r_2 \wedge 0 < \theta < 2\pi\right\}$ (i.e. a rectangle); and

(c) a transformation $\Phi : B \to A$ defined by

$$(r,\theta) \mapsto (r\cos(\theta), r\sin(\theta))$$

for all $(r,\theta) \in B$.

Then Φ is a C^1-diffeomorphism (verify this). Moreover, for any $(r,\theta) \in B$, a direct computation shows that

$$|J|_\Phi(r,\theta) = \det\begin{bmatrix}\cos(\theta) & -r\sin(\theta)\\ \sin(\theta) & r\cos(\theta)\end{bmatrix} = r.$$

Thus, by using [3.21],

$$\int_A f(x)\,\mathrm{d}x = r\int_B f(r\cos(\theta), r\sin(\theta))\,\mathrm{d}(r,\theta).$$