Miscellaneous analysis notes

Danny Nygård Hansen

3rd December 2022

1 • Basic concepts

1.1. *The real numbers*

DEFINITION 1.1: *Peano systems*

A *Peano system* is a tuple $(X, 0_X, S)$ such that X is a set, 0_X is some element in X, and $S\colon X \to X$ is a map, that has the following properties:

(i) S is injective.

(ii) For all $x \in X$, $0_X \neq S(x)$.

(iii) For all $A \subseteq X$, if $0_X \in A$ and $x \in A$ implies $S(x) \in A$, then $A = X$.

We define a strict total order $<$ on a Peano system by declaring that $x < S(x)$, and that $x < y$ implies $S(x) < S(y)$, for all $x, y \in X$. Notice that with this ordering, 0_X is the minimum of X.

REMARK 1.2. Recall that an ordered set X is *well-ordered* if any nonempty subset of X has a minimum. Contrary to what many textbook authors seem to suggest, the induction principle (3) is *not* equivalent to being well-ordered, given the other axioms. While any Peano system is well-ordered, as we prove below, the converse does not hold: For instance, if $\mathbb{N}$ are the natural numbers (as we define below), the linear sum $\mathbb{N} \oplus \mathbb{N}$ is well-ordered but does not satisfy the induction principle (compare the ordinal $\omega + \omega$).

The usual 'proof' that being well-ordered implies the induction principle uses that the element 0_X above is the only element of X without a predecessor, and with this modification to the Peano axioms well-ordering does imply the induction principle (and vice-versa). ⌟

Corollary 1.3: *Well-ordering of Peano systems*

Any Peano system $(X, 0_X, S)$ *is well-ordered.*

Proof. Let $A \subseteq X$ be nonempty, and let L be the subset of X consisting of lower bounds of A. Then clearly $0_X \in L$, and if $a \in A$ then $S(a) \notin L$. Hence $L \neq X$, so since X is a Peano system there is an $m \in L$ such that $S(m) \notin L$. We must then have $m \in A$, since otherwise $S(m) \in L$. Thus m is the minimum of A. □

Definition 1.4: *The real numbers*

The *real numbers* is the unique (up to isomorphism) complete ordered field, and is denoted $\mathbb{R}$.

We denote the operations on $\mathbb{R}$ by the usual symbols. In particular, the ordering is denoted $\leq$, and the additive and multiplicative identities are denoted 0 and 1 respectively. Furthermore, if $A \subseteq \mathbb{R}$ then we write e.g.

$$x + A := \{x + a \mid a \in A\},$$
$$cA := \{ca \mid a \in A\},$$

for $x, c \in \mathbb{R}$. We further write $-A$ instead of $-1A$.

A subset A of $\mathbb{R}$ is said to be *inductive* if $0 \in A$, and if $a \in A$ implies $a + 1 \in A$ for all $a \in \mathbb{R}$. Any intersection of inductive subsets is itself inductive, so the intersection of all inductive subsets is inductive, in particular nonempty. This is clearly the smallest inductive subset of $\mathbb{R}$, and is called the *natural numbers* and is denoted $\mathbb{N}$. Notice that the positive axis $[0, \infty)$ is inductive, so all natural numbers are nonnegative and 0 is the minimum of $\mathbb{N}$.

Proposition 1.5

If $S : \mathbb{N} \to \mathbb{N}$ *is given by* $S(n) = n + 1$, *then* $(\mathbb{N}, 0, S)$ *is a Peano system.*

Proof. Clearly $0 \in \mathbb{N}$, and S is injective by cancellation of addition. Also, notice that $S(n) = n + 1 > n \geq 0$ for all $n \in \mathbb{N}$, so 0 is not in the image of S. Finally, let $A \subseteq \mathbb{N}$ be such that $0 \in A$ and $n \in A$ implies $S(n) \in A$. Then A is an inductive subset of $\mathbb{R}$, so $\mathbb{N} \subseteq A$. Hence $(\mathbb{N}, 0, S)$ is a Peano system. □

Furthermore, the *integers* is the set $\mathbb{Z} = \mathbb{N} \cup (-\mathbb{N})$. We further denote by $\mathbb{Z}_+$ the subset of strictly positive integers, which is just the set $\mathbb{N} \setminus \{0\}$. Finally, the *rational numbers* is the set $\mathbb{Q}$ of numbers on the form m/n, where $m, n \in \mathbb{Z}$ and $n > 0$. This is easily seen to be a countable subfield of $\mathbb{R}$.

PROPOSITION 1.6: *The Archimedean property of $\mathbb{R}$*

The set $\mathbb{N}$ is unbounded in $\mathbb{R}$. In particular, for any $r > 0$ there is an $n \in \mathbb{Z}_+$ such that $1/n < r$. Furthermore, for $x, y \in \mathbb{R}$ with $x > 0$ there exists an $n \in \mathbb{Z}_+$ such that $nx > y$.

PROOF. Assume towards a contradiction that $\mathbb{N}$ is bounded. Then it has a supremum $s = \sup \mathbb{N}$, so $n \leq s$ for all $n \in \mathbb{N}$. But then also $n + 1 \leq s$ for all $n \in \mathbb{N}$ since $\mathbb{N}$ is inductive, so $n \leq s - 1$ for all $n \in \mathbb{N}$, contradicting that s is the supremum of $\mathbb{N}$.

If $r > 0$ then there is an $n \in \mathbb{Z}_+$ with $n > 1/r$. Since both r and n are positive, this implies that $1/n < r$.

Finally let $x, y \in \mathbb{R}$ with $x > 0$. There is an $n \in \mathbb{Z}_+$ with $n > y/x$, so since x is positive this implies that $nx > y$ as desired. □

COROLLARY 1.7: *Density of $\mathbb{Q}$ in $\mathbb{R}$*

For each $a, b \in \mathbb{R}$ with $a < b$ there is a rational number in the interval (a, b). In particular, the set $\mathbb{Q}$ is dense in $\mathbb{R}$.

PROOF. First assume that $a \geq 0$. By [TODO Archimedes] there is an $n \in \mathbb{Z}_+$ with $1/n < b - a$. Furthermore, there is an $m \in \mathbb{N}$ such that $m/n > a$, so choose the smallest such m in accordance with [TODO well-ordered]. Then $(m-1)/n \leq a$, and so

$$\frac{m}{n} = \frac{m-1}{n} + \frac{1}{n} < a + (b - a) = b,$$

as desired.

Instead assume that $a < 0$, and choose an $n \in \mathbb{N}$ with $n \geq -a$. The above then yields a rational number $q \in (a + n, b + n)$, so $q - n \in (a, b)$ is the desired rational number. □

LEMMA 1.8: *Bernoulli's inequality*

Let $r \geq -1$. Then for all $n \in \mathbb{N}$,

$$(1 + r)^n \geq 1 + nr,$$

with equality iff $n \in \{0, 1\}$ or $r = 0$.

PROOF. This is obvious if $n = 0$, so assume that the inequality holds for some $n \in \mathbb{N}$. Then

$$(1 + r)^{n+1} = (1 + r)(1 + r)^n \geq (1 + r)(1 + nr) = 1 + (n + 1)r + nr^2 \geq 1 + (n + 1)r,$$

as desired. Notice that the final inequality is strict if $r \neq 0$ and $n \geq 1$. □

Theorem 1.9: *Existence of roots*

For all $a > 0$ and $n \in \mathbb{Z}_+$ there is a unique $r > 0$ such that $r^n = a$.

Proof. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^n$. By [TODO Bernoulli + Archimedes], $f(x) \to \infty$ as $x \to \infty$, so there is an $x_0 > 0$ such that $x_0^n \geq a$. Furthermore, since $f(0) = 0$ and f is continuous, the intermediate value theorem yields an $r \in [0, x_0]$ such that $r^n = a$ as desired. □

1.2. The exponential function

Lemma 1.10

Consider the functional equation

$$f(x+y) = f(x)f(y),$$

where $x, y \in \mathbb{R}$, and $f : G \to \mathbb{R}$ where G is a subgroup of $\mathbb{R}$. Any solution is either identically zero or everywhere nonzero. In particular, any nonzero solution is a group homomorphism $G \to R^*$.

Let $a \in [0, \infty)$. If $G = \mathbb{Q}$, then the equation has a unique solution $f : \mathbb{Q} \to \mathbb{R}$ given the restriction $f(1) = a$.

Furthermore, if $f : G \to \mathbb{R}$ solves the equation and f is continuous at some point in G, then f is continuous. Hence the equation has a unique solution $f : G \to \mathbb{R}$ with $f(1) = a$ among the functions $G \to \mathbb{R}$ that are continuous at any point.

Proof. Let $f : G \to \mathbb{R}$ be a solution. First notice that $f(0) = f(0+0) = f(0)f(0)$, so either $f(0) = 0$ or $f(0) = 1$. In the former case we have $f(x) = f(x+0) = f(x)f(0) = 0$ for all $x \in G$, so f is identically zero. Hence assume that $f(0) = 1$. Then $1 = f(x-x) = f(x)f(-x)$, so $f(x)^{-1} = f(-x)$. In particular, f is everywhere nonzero, hence a group homomorphism $G \to \mathbb{R}^*$.

Assume that $G = \mathbb{Q}$ and let $g : \mathbb{Q} \to \mathbb{R}$ be another solution. For $n \in \mathbb{Z}_+$ we have

$$f(n) = f(\underbrace{1 + \cdots + 1}_{n}) = \underbrace{f(1)\cdots f(1)}_{n} = f(1)^n = g(1)^n = \cdots = g(n).$$

Furthermore, since $f(-n) = f(n)^{-1} = g(n)^{-1} = g(-n)$, f and g agree on $\mathbb{Z}$. Next let $m \in \mathbb{Z}$ and notice that

$$f(\tfrac{m}{n})^n = f(n\tfrac{m}{n}) = f(m) = g(m) = g(\tfrac{m}{n})^n.$$

Hence f and g agree on $\mathbb{Q}$ as claimed.

Next assume that $f : G \to \mathbb{R}$ is a solution that is continuous at some point. If f is identically zero then it is obviously continuous, and if not then it is a group homomorphism. But both G and $\mathbb{R}^*$ are topological groups, so this implies that f is continuous everywhere. □

THEOREM 1.11

Given $b > 0$ there exists a continuous function $E_b\colon \mathbb{R} \to \mathbb{R}$ such that $E_b(m/n) = (b^m)^{1/n}$ for all $m, n \in \mathbb{Z}$ with $n > 0$. This function has the property that

$$E_b(x + y) = E_b(x)E_b(y)$$

for $x, y \in \mathbb{R}$. Furthermore, E_b is strictly increasing if $b > 1$, strictly decreasing if $b < 1$. If $b \neq 1$, then $E_{1/b}(x) = E_b(-x)$ and $E_b(\mathbb{R}) = (0,\infty)$.

The function E_b is the unique function that satisfies $E_b(m/n) = (b^m)^{1/n}$ and that is continuous at 0. We write $b^x = E_b(x)$ for $x \in \mathbb{R}$.

The function E_b is also denoted $\exp_b$ and is called the *exponential function with base b*.

PROOF. We first assume that $b > 1$. The proof of the existence of the function E_b in this case is by the following stages:

(1) Given $m, n, p, q \in \mathbb{Z}$ with $n, q > 0$ and $r = m/n = p/q$, then

$$(b^m)^{1/n} = (b^p)^{1/q}.$$

Thus defining $E_b(r) = (b^m)^{1/n}$ makes sense.

(2) If $r, s \in \mathbb{Q}$, then $E_b(r + s) = E_b(r)E_b(s)$. In particular, E_b is strictly increasing on $\mathbb{Q}$. Furthermore, $E_b(r) \to \infty$ as $r \to \infty$, and $E_b(r) \to 0$ as $r \to -\infty$.

(3) The map E_b is continuous on $\mathbb{Q}$.

(4) For $x \in \mathbb{R}$, let

$$L_x = \{E_b(t) \mid t \in \mathbb{Q}, t \leq x\}.$$

Then $E_b(r) = \sup L_r$ for $r \in \mathbb{Q}$, so it makes sense to define $E_b(x) = \sup L_x$ for all $x \in \mathbb{R}$. Hence E_b is strictly increasing and therefore continuous on $\mathbb{R}$. In particular, $E_b(\mathbb{R}) = (0,\infty)$.

(5) For $x, y \in \mathbb{R}$ we have $E_b(x + y) = E_b(x)E_b(y)$.

Part (1): Notice that $mq = pn$, so

$$\left[(b^m)^{1/n}\right]^{pn} = (b^m)^p = (b^p)^m = \left[(b^p)^{1/q}\right]^{mq} = \left[(b^p)^{1/q}\right]^{pn}.$$

The (positive) pnth root of this number is unique, so $(b^m)^{1/n} = (b^p)^{1/q}$.

Part (2): Write $r = m/n$ and $s = p/q$ for appropriate $m, n, p, q \in \mathbb{Z}$ with $n, q > 0$. Then

$$[(b^m)^{1/n}(b^p)^{1/q}]^{nq} = b^{mq}b^{pn} = b^{mq+pn}.$$

Taking the nqth root implies that

$$E_b(r)E_b(s) = (b^m)^{1/n}(b^p)^{1/q} = (b^{mq+pn})^{1/nq} = E_b(r+s),$$

where we in the last equality use that

$$r + s = \frac{m}{n} + \frac{p}{q} = \frac{mq+pn}{nq}.$$

To see that E_b is increasing on $\mathbb{Q}$, notice that the assumption that $b > 1$ implies that $E_b(s) > 1$ for $s > 0$. If also $r \in \mathbb{Q}$, then since $E_b(r) \neq 0$ by [TODO lemma], we have

$$E_b(r+s) = E_b(r)E_b(s) > E_b(r),$$

so E_b is strictly increasing.

Finally, by [TODO Bernoulli] $b^m \to \infty$ as $m \to \infty$. Since $1 = b^m b^{-m}$, this implies that $b^m \to 0$ as $m \to -\infty$.

Part (3): We begin by showing that E_b is continuous from above at 0. Since E_b is monotonic and $E_b(0) = 1$, it suffices to show that $\lim_{n\to\infty} E_b(1/n) = 1$. Clearly $1/n \downarrow 0$, so assume that $E_b(1/n)$ did not converge to 1. Then there would be some $\varepsilon > 0$ such that $E_b(1/n) \geq 1 + \varepsilon$, i.e. $b \geq (1+\varepsilon)^n$, for all $n \in \mathbb{N}$. But by [TODO Bernoulli's inequality] this is impossible, so we must have $E_b(1/n) \to 1$. Since $E_b(-m/n) = (b^{-m})^{1/n} = E_{1/b}(m/n)$, E_b is also continuous from below at 0. [TODO lemma] then implies that E_b is continuous everywhere.

Part (4): Since E_b is increasing on $(\infty, r] \cap \mathbb{Q}$, it obtains its maximum at r. Hence we obviously have $E_b(r) = \sup L_r$.

If $x \leq y$ then $L_x \subseteq L_y$, and hence $E_b(x) \leq E_b(y)$. Thus E_b is increasing on $\mathbb{R}$. It is in fact strictly increasing, since if $x < y$ and $p < q$ are rational numbers in the interval (x, y), then since E_b is strictly increasing on $\mathbb{Q}$ we have

$$E_b(x) \leq E_b(p) < E_b(q) \leq E_b(y).$$

Furthermore, since E_b continuous on a dense subset of $\mathbb{R}$, it is clearly also continuous on $\mathbb{R}$. Finally, the claim $E_b(\mathbb{R}) = (0, \infty)$ follows from the limit results in part (2).

Part (5): Let (r_n) and (s_n) be sequences in $\mathbb{Q}$ with limits x and y respectively. Then $r_n + s_n$ converges to $x + y$, so

$$E_b(x+y) = \lim_{n\to\infty} E_b(r_n + s_n) = \lim_{n\to\infty} E_b(r_n)E_b(s_n) = E_b(x)E_b(y),$$

as desired.

This proves existence when $b > 1$. If $b = 1$ then existence is obvious, so assume that $b < 1$. In this case we simply define $E_b(x) = E_{1/b}(-x)$. This clearly has the desired properties. Uniqueness is immediate from [TODO lemma]. $\square$

Let $b \neq 1$. Then E_b is a bijection onto $(0,\infty)$, so it has an inverse $L_b\colon (0,\infty) \to \mathbb{R}$. This is called the *logarithm with base b*, and is also denoted $\log_b$. This satisfies a functional equation similar to the one satisfied by E_b: For $x,y \in (0,\infty)$ we have

$$xy = E_b(L_b(x))E_b(L_b(y)) = E_b(L_b(x) + L_b(y)),$$

and applying L_b to both sides yields

$$L_b(xy) = L_b(x) + L_b(y).$$

For $n \in \mathbb{N}$ this in particular implies that $L_b(x^n) = nL_b(x)$.

We next turn to alternative characterisations of the exponential function. We first define the number

$$\mathrm{e} = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

The exponential function $E_\mathrm{e} = \exp_\mathrm{e}$ with base e is simply called the *exponential function* and is denoted exp. Similarly, the logarithm $L_\mathrm{e} = \log_\mathrm{e}$ with base e is called the *natural logarithm* and is denoted log.

PROPOSITION 1.12

For $x \in \mathbb{R}$ we have

$$\exp x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

In particular $\exp' = \exp$ *and* $\log' x = 1/x$ *for* $x \in (0,\infty)$.

PROOF. Denote the right-hand side by $E(x)$. First let $f(x) = E(x)E(-x)$ and notice that, since $E' = E$, $f' = 0$. Hence $E(x)E(-x) = 1$. Next define $g(x) = E(x)E(y)E(-(x+y))$ for $y \in \mathbb{R}$. Using the above, one easily sees that $g' = 0$, so $g(x) = g(0) = 1$. Hence $E(x)E(y) = E(x+y)$, so [TODO lemma] and [TODO e^x satisfies func eq] implies that $\exp e = E(x)$ as claimed. $\square$

PROPOSITION 1.13

Let $x \in \mathbb{R}$. Then

$$\exp x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n.$$

PROOF. This is obvious if $x = 0$, so assume that $x \neq 0$. Then we have

$$\lim_{n\to\infty} n\log(1+\tfrac{x}{n}) = x \lim_{n\to\infty} \frac{\log(1+\frac{x}{n}) - \log 1}{x/n} = x\log' 1 = x.$$

(Notice that $1 + \frac{x}{n} > 0$ if n is large enough.) Continuity of exp then implies that

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \exp\left(\lim_{n\to\infty} n\log(1+\tfrac{x}{n})\right) = \exp x,$$

as desired. □

1.3. Sequences of real numbers

LEMMA 1.14

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in a metric space (S,ρ). If (a_n) is both Cauchy and has a convergent subsequence, then (a_n) itself is convergent.

PROOF. Let $(a_{n_k})_{k\in\mathbb{N}}$ be a convergent subsequence of (a_n), and let $\varepsilon > 0$. Choose $N_1, N_2 \in \mathbb{N}$ such that

$$m,n \geq N_1 \quad\Rightarrow\quad \rho(a_m, a_n) < \frac{\varepsilon}{2}$$

and

$$k \geq N_2 \quad\Rightarrow\quad \rho(a_{n_k}, a) < \frac{\varepsilon}{2},$$

where $a \in S$ is the limit of (a_{n_k}). For $n \geq N_1 \vee N_2$ we thus have

$$\rho(a_n, a) \leq \rho(a_n, a_m) + \rho(a_m, a) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

showing that $a_n \to a$ as $n \to \infty$. □

PROPOSITION 1.15

Let $(a_n)_{n\in\mathbb{N}}$ be a monotonic sequence in $\mathbb{R}$. Then (a_n) is convergent if and only if it is bounded, in which case it converges to $\sup_{n\in\mathbb{N}} a_n$ if it is increasing and $\inf_{n\in\mathbb{N}} a_n$ if it is decreasing.

PROOF. If (a_n) is convergent then it is bounded, so assume that it is bounded and let $\varepsilon > 0$. For definiteness we assume that it is increasing and let $s = \sup_{n\in\mathbb{N}} a_n$. By definition of s there exists an $N \in \mathbb{N}$ such that $s - a_N < \varepsilon$. Since (a_n) is increasing and s is an upper bound of the sequence, we thus have

$$0 \leq s - a_n < \varepsilon$$

for all $n \geq N$, proving that $a_n \to s$. □

LEMMA 1.16

Every sequence in $\mathbb{R}$ has a monotonic subsequence.

PROOF. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. We say that $n \in \mathbb{N}$ is a *peak* if $a_n \geq a_m$ for all $m \geq n$. If (a_n) has infinitely many peaks, the subsequence consisting of these constitute a decreasing subsequence.

Hence we assume that (a_n) only has finitely many peaks. We construct an increasing sequence $(n_k)_{n\in\mathbb{N}}$ in $\mathbb{N}$ as follows: Let $n_1 \in \mathbb{N}$ be such that all peaks are strictly less than n_1, and assume that $n_1, \ldots, n_{k-1}$ have been chosen such that $a_1 \leq \cdots \leq a_{n_{k-1}}$. Since $a_{n_{k-1}}$ is not a peak there is an $n' > n_{k-1}$ such that $a_{n_{k-1}} < a_{n'}$. Letting $n_k = n'$ we obtain an increasing subsequence (a_{n_k}) of (a_n), proving the claim. □

THEOREM 1.17: *The Bolzano–Weierstrass theorem*

Every subset of $\mathbb{R}^d$ is sequentially compact if and only if it is closed and bounded.

We recall that a topological space X is *sequentially compact* if every sequence in X has a convergent subsequence.

PROOF. We begin with the case $d = 1$. Let $A \subseteq \mathbb{R}$ be closed and bounded, and let $(a_n)_{n\in\mathbb{N}}$ be a sequence in A. Let (a_{n_k}) be a monotonic subsequence of (a_n), and notice that (a_{n_k}) is convergent since it is bounded.

The case for general d follows by induction in d, by noticing that a sequence in $\mathbb{R}^d$ converges if and only if each coordinate sequence converges.

For the converse, let $A \subseteq \mathbb{R}^d$ be sequentially compact. If A were not bounded we could choose $a_n \in A \cap B(0,n)$ for all $n \in \mathbb{N}$, yielding a sequence (a_n) with no convergent subsequence. Furthermore, if (a_n) is a sequence in A converging to a point $a \in \mathbb{R}^d$, then it has a subsequence converging to a point $a' \in A$, and by [lemma] we must have $a = a'$. Thus A is also closed. □

THEOREM 1.18: *Completeness of $\mathbb{R}^d$*

The Euclidean space $\mathbb{R}^d$ is complete.

PROOF. Let $(a_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}^d$. Hence it is bounded, and so it has a convergent subsequence by the Bolzano–Weierstrass theorem. But then (a_n) itself converges by [lemma], so $\mathbb{R}^d$ is complete. □

THEOREM 1.19: *The Heine–Borel theorem*

Every subset of $\mathbb{R}^d$ is compact if and only if it is closed and bounded.

PROOF. Of course every compact set is closed in any Hausdorff space and bounded in any metric space, so we only consider the other implication.

We first show that closed and bounded intervals are compact. Consider the interval $[a,b]$, and let $\mathcal{U}$ be an open cover of $[a,b]$. Define the set

$$A = \Big\{ x \in [a,b] \;\Big|\; [a,x] \text{ has a finite subcover in } \mathcal{U} \Big\}.$$

We clearly have $a \in A$ since a point is covered by a single set in $\mathcal{U}$. If $s = \sup A$ then $a \leq s \leq b$. Suppose that $s < b$ and choose a set $U \in \mathcal{U}$ with $s \in U$. There exist $r,t \in U$ such that $r < s < t$, and so $r \in A$. Let $\mathcal{U}'$ denote a finite subcover of $[a,r]$ in $\mathcal{U}$. Then $\mathcal{U}' \cup \{U\}$ is a finite subcover of $[a,t]$, contradicting the assumption that $s < b$. Hence $s = b$.

Next, choose $V \in U$ with $b \in V$, and let $c \in V$ with $c < b$. Then $c \in A$, and adjoining V to a finite subcover of $[a,c]$ yields a finite subcover of $[a,b]$, so $b \in A$. Thus $[a,b]$ is compact.

Finally, let $K \subseteq \mathbb{R}^d$ be closed and bounded. Since it is bounded it is contained in some cube $[-a,a]^d$. But this cube is a product of compact sets and hence compact, so K is a closed subset of a compact set. The claim follows. □

1.4. *Infinite series*

PROPOSITION 1.20: *Cauchy criterion for series*

Any series $\sum_{n=1}^{\infty} a_n$ with complex terms converges if and only if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$m \geq n \geq N \quad \Rightarrow \quad \left| \sum_{k=n}^{m} a_k \right| < \varepsilon.$$

PROOF. Let $s_n = \sum_{k=1}^{n} a_n$. By completeness of $\mathbb{C}$, the series converges if and only if $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. This is the case just when for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $m \geq n \geq N$ implies that

$$\varepsilon > |s_m - s_n| = \left| \sum_{k=n+1}^{m} a_k \right|,$$

as claimed. □

COROLLARY 1.21

If a series $\sum_{n=1}^{\infty} a_n$ with complex terms converges, then $a_n \to 0$ as $n \to \infty$.

PROOF. Let $m = n + 1$ in [TODO ref]. □

Corollary 1.22

An absolutely convergent series with complex terms is convergent.

Proof. This follows from [TODO ref] by noting that

$$\left|\sum_{k=n}^{m} a_k\right| \leq \sum_{k=n}^{m} |a_k|$$

for all $m, n \in \mathbb{N}$ with $m \geq n$. □

1.5. Unordered sums

If X is a set, then we denote by $\mathcal{F}(X)$ the collection of finite subsets of X. This is directed by set inclusion.

Definition 1.23

Let $g\colon I \to G$ be a map from a set I into an abelian topological group G, and write $g(i) = g_i$. Define a net $S\colon \mathcal{F}(I) \to G$ by $S(J) = \sum_{i\in J} g_i$. If S has a limit, then such a limit is called an *unordered sum* of g over I, and if this is unique then it is denoted

$$\sum_{i\in I} g_i.$$

Notice that since the sum is *unordered*, we cannot generalise this definition to nonabelian groups. Also since the sum is unordered, we would expect that if we rearrange the terms in the sum, its convergence properties should stay the same. More precisely, if the unordered sum $\sum_{i\in I} g_i$ converges to $g \in G$ and $\varphi\colon I \to I$ is a bijection, then we would expect that $\sum_{i\in I} g_{\varphi(i)}$ also converges to g. But in the notation above, $S \circ \varphi$ is clearly a (Willard) subnet of S, so this follows.

Lemma 1.24

Let G be an abelian topological group, and assume that the unordered sum $\sum_{i\in I} g_i$ converges to $g \in G$. If $(J_n)_{n\in\mathbb{N}}$ is an increasing sequence in $\mathcal{F}(I)$ with $I = \bigcup_{n\in\mathbb{N}} J_n$, then the sequence $(\sum_{i\in J_n} g_i)_{n\in\mathbb{N}}$ converges to g.

Proof. Simply notice that the sequence $(\sum_{i\in J_n} g_i)_{n\in\mathbb{N}}$ is a (Willard) subnet of $\sum_{i\in I} g_i$. □

Proposition 1.25

Let G be an abelian topological group. If the unordered sum $\sum_{n\in\mathbb{N}} g_n$ converges *to g, then the series $\sum_{n=1}^{\infty} g_n$ converges to g. Furthermore, every rearrangement of $\sum_{n=1}^{\infty} g_n$ converges to g.*

PROOF. For $n \in \mathbb{N}$ let $J_n = \{1, \dots, n\}$ and consider the sequence $(s_n)_{n\in\mathbb{N}}$ given by $s_n = \sum_{i\in I_n} g_i$. Notice that $s_n = \sum_{i=1}^{n} g_i$, so the claim follows from Lemma 1.24. □

LEMMA 1.26

An unordered sum $\sum_{i\in I} g_i$ in G is Cauchy if and only if the sum has the following property: For every neighbourhood N of 0 there exists a $J \in \mathcal{F}(I)$ such that if $K \in \mathcal{F}(I)$ with $J \cap K = \emptyset$, then $\sum_{i\in K} g_i \in N$. In other words,

$$\forall N \in \mathcal{N}_0 \, \exists J \in \mathcal{F}(I) \forall K \in \mathcal{F}(I) \colon J \cap K = \emptyset \quad \Rightarrow \quad \sum_{i\in K} g_i \in N.$$

PROOF. Suppose the unordered sum is Cauchy. Let N be a neighbourhood of 0. Since the sum is Cauchy there exists a $J \in \mathcal{F}(I)$ such that $J \subseteq L, L'$ implies that

$$\sum_{i\in L} g_i - \sum_{i\in L'} g_i \in N,$$

for $L, L' \in \mathcal{F}(I)$. Hence if $K \in \mathcal{F}(I)$ is disjoint from J, then

$$\sum_{i\in K} g_i = \sum_{i\in J\cup K} g_i - \sum_{i\in J} g_i \in N,$$

as desired.

Conversely assume that the sum has the property above, let U be an open neighbourhood of 0, and let V be a symmetric open neighbourhood of 0 with $V + V \subseteq U$. Let $J \in \mathcal{F}(I)$ be such that $J \cap K = \emptyset$ implies that $\sum_{i\in K} g_i \in V$ for $K \in \mathcal{F}(I)$. If $J \subseteq L, L'$, then

$$\sum_{i\in L} g_i - \sum_{i\in L'} g_i = \sum_{i\in L\setminus J} g_i - \sum_{i\in L'\setminus J} g_i \in V - V \subseteq U.$$

Thus the sum is Cauchy as claimed. □

Since the topology on a topological group G is homogeneous, G is automatically R_0. That is, if $g, h \in G$ and g has a neighbourhood not containing h, then h has a neighbourhood not containing g. In other words, topologically distinguishable points are separated. In particular, the specialisation preorder $\leq$ is symmetric, so that $g \leq h$ implies $g \equiv h$.

LEMMA 1.27

Let G be an abelian topological group that is also first countable. If the unordered *sum $\sum_{i\in I} g_i$ is Cauchy, then the set $\{i \in I \mid g_i \not\equiv 0\}$ is countable.*

PROOF. Let $\mathcal{B}$ be a countable neighbourhood basis at 0. By taking intersections we may assume that the neighbourhoods in $\mathcal{B}$ form a decreasing sequence $(N_n)_{n\in\mathbb{N}}$. For each $n \in \mathbb{N}$, choose $J_n \in \mathcal{F}(I)$ such that $J_n \cap K = \emptyset$ implies that $\sum_{i\in K} g_i \in N_n$. Now let $J = \bigcup_{n\in\mathbb{N}} J_n$. For any $i \in I \setminus J$ the set $\{i\}$ is disjoint from each J_n, so

$$g_i = \sum_{j\in\{i\}} g_j \in N_n.$$

Since the N_n form a neighbourhood basis at 0, every neighbourhood of 0 intersects g_i, i.e. $g_i \in \overline{\{0\}}$. But this means that $g_i \leq 0$ in the specialisation preorder, so $g_i \equiv 0$ since G is R_0. Hence it follows that

$$\{i \in I \mid g_i \not\equiv 0\} \subseteq J,$$

so the former set is countable since J is, as desired. □

1.6. Limits of functions

Let X be a topological space, and let $A \subseteq X$. A subset N is a *neighbourhood* of A if there is an open set U such that $A \subseteq U \subseteq N$. The set of neighbourhoods of A is denoted $\mathcal{N}_A$ and is called the *neighbourhood filter* of A (since it is indeed a filter). If $A = \{x\}$ is a singleton we also write $\mathcal{N}_x$ for the neighbourhood filter of x. A *punctured neighbourhood* of $x \in X$ is a set on the form $N \setminus \{x\}$, where $N \in \mathcal{N}_x$. The set of punctured neighbourhoods of x is denoted $\mathcal{N}'_x$.

A point $a \in X$ is called an *adherent point* of $A \subseteq X$ if every neighbourhood of a intersects A, i.e. if $a \in \overline{A}$. Furthermore, a is called a *limit point* of A if every *punctured* neighbourhood of a intersects A, or equivalently if a is an adherent point of $A \setminus \{a\}$.

DEFINITION 1.28: *Limits of functions*

Let X and Y be topological spaces, let $A \subseteq X$, and let $f : A \to Y$. If $L \subseteq A$ and $a \in X$ is an adherent point of L, then the triple (f, L, a) is called a *limit in* X. If $b \in Y$ is such that

$$\forall N \in \mathcal{N}_b \, \exists M \in \mathcal{N}_a : f(M \cap L) \subseteq N,$$

then we say that b is a *limit value* for the limit (f, L, a). The set of such limit values is denoted $\mathcal{L}(f, L, a)$.

If $\mathcal{L}(f, L, a)$ contains a single element b, then we also write

$$b = \lim_{a,L} f = \lim_{\substack{x \to a \\ x \in L}} f(x).$$

REMARK 1.29. In Definition 1.28 we did not specify whether the neighbourhoods of a were neighbourhoods in A or in X. In fact, this does not matter: The map $M \mapsto M \cap A$ gives a one-to-one correspondence between neighbourhoods of a in X and A, and since L is a subset of A we have

$$f((M \cap A) \cap L) = f(M \cap L).$$

Hence we may use either neighbourhoods in X or in A when taking limits. ⌟

PROPOSITION 1.30: *Uniqueness of limits*

Let X and Y be topological spaces, let $A \subseteq X$, and let $f : A \to Y$. If Y is Hausdorff, then any limit (f, L, a) in X has at most one limit point.

PROOF. Let $b \in Y$ be a limit value for (f, L, a), let $b' \neq b$, and let $N, N' \subseteq Y$ be disjoint neighbourhoods of b and b', respectively. Since b is a limit value, there is a neighbourhood $M \in \mathcal{N}_a$ such that $f(M \cap L) \subseteq N$. If M' is any neighbourhood of a then $M \cap M'$ is also a neighbourhood of a, so it intersects L.[1] But since

$$f((M \cap M') \cap L) \subseteq f(M \cap L) \subseteq N,$$

the set $f(M' \cap L)$ intersects N, so it does not lie in N'. Hence b' is not a limit value for (f, L, a). □

EXAMPLE 1.31. Definition 1.28 provides a very general notion of limit, of which the familiar limiting processes are special cases:

(i) Recall the standard definition of a limit of a function between topological spaces: If a is a limit point of A, then we usually say that $b \in Y$ is a limit of $f(x)$ as $x \to a$ if

$$\forall N \in \mathcal{N}_b \, \exists M' \in \mathcal{N}'_a \colon f(M') \subseteq N.$$

We recover this notion by considering the triple $(f, A \setminus \{a\}, a)$: For notice that if M is a neighbourhood of a, then

$$M \cap (A \setminus \{a\}) = (M \setminus \{a\}) \cap A,$$

[1] Notice that we here make crucial use of the requirement that a be an adherent point of L.

and every punctured neighbourhood of a is on the form $M \setminus \{a\}$. Here we recall from Remark 1.29 that it is immaterial whether we ue neighbourhoods of a in A or in X, so the intersection with A makes no difference.

Notice that it is crucial that a is a limit point of A and not just an adherent point, since a is a limit point of A if and only if a is an adherent point of $A \setminus \{a\}$, which we require in our definition of limits.

Below we shall take this notion of limit as standard, so if Y is Hausdorff and hence limits are unique by Proposition 1.30, we use the shorthand notation

$$b = \lim_a f = \lim_{x \to a} f(x).$$

If the set A is not clear from context, we may disambiguate by writing 'for $x \in A \setminus \{a\}$', or something to that effect.

(ii) Some authors distinguish between *deleted* and *non-deleted* limits. The notion of limits in (i) is that of deleted limits, since we only reture that $f(M') \subseteq N$ for a *punctured* neighbourhood M' of a. By contrast, in a *non-deleted* limit we require that $f(M) \subseteq N$ for an ordinary neighbourhood M of a. We may recover the notion of non-deleted limits by considering triples (f, A, a).

Non-deleted limits are relatively rare in the English-language literature (one notable exception is Tao (2015a) and Tao (2015b), who allows for both types), but it seems to be the prevailing notion in the French-language literature.

(iii) Consider next a function $f: A \to Y$, where $A \subseteq \mathbb{R}$ contains an interval (a, b). In this case we may consider the one-sided limits of $f(x)$ as x approaches either a from above or b from below. These may be described in the above formalism by the triples $(f, (a, b), a)$ and $(f, (a, b), b)$, respectively (indeed, we may choose any subinterval of (a, b) with a or b as a limit point, respectively, and the set of limit values is clearly independent of such a choice). If they exist, we denote the corresponding limit values by

$$\lim_{x \downarrow a} f(x) \quad \text{and} \quad \lim_{x \uparrow b} f(x),$$

respectively. ⌟

Lemma 1.32

Let X and Y be topological spaces, let $A \subseteq X$, and let $f: A \to Y$. If $L_1 \subseteq L_2 \subseteq A$ and $a \in X$ is an adherent point of both L_1 and L_2, then

$$\mathcal{L}(f, L_2, a) \subseteq \mathcal{L}(f, L_1, a).$$

PROOF. Let $b \in \mathcal{L}(f, L_2, a)$ and let $N \in \mathcal{N}_b$. Then there is an $M \in \mathcal{N}_a$ such that

$$f(M \cap L_1) \subseteq f(M \cap L_2) \subseteq N,$$

showing that $b \in \mathcal{L}(f, L_1, a)$ as desired. □

PROPOSITION 1.33

Let X and Y be topological spaces, let $A \subseteq X$, and let $f\colon A \to Y$. If $a \in A$ is a limit point of A, then the following are equivalent:

(i) *f is continuous at a.*

(ii) $f(a) \in \mathcal{L}(f, A, a)$.

(iii) $f(a) \in \mathcal{L}(f, A \setminus \{a\}, a)$.

PROOF. Assume that f is continuous at a, and let $N \in \mathcal{N}_{f(a)}$. By continuity there is an[2] $M \in \mathcal{N}_a$ such that $f(M) \subseteq N$, implying that $f(M \cap A) \subseteq N$. Next, (ii) implies (iii) by [TODO ref lemma].

Finally, if $f(a) \in \mathcal{L}(f, A \setminus \{a\}, a)$ then given $N \in \mathcal{N}_{f(a)}$ there is a neighbourhood $M \in \mathcal{N}_a$ in A such that $f(M \setminus \{a\}) = f(M \cap A \setminus \{a\}) \subseteq N$. Since N contains $f(a)$ we have $f(M) \subseteq N$ so f is continuous at a. □

2 • Differentiation

2.1. *Differentiability*

DEFINITION 2.1: *Differentiability*

Let $A \subseteq \mathbb{R}^d$, let $a \in A$ be a limit point of A, and let $f\colon A \to \mathbb{R}^m$. If there exists a linear map $L \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ such that

$$\lim_{x \to a} \frac{\|f(x) - f(a) - L(x-a)\|}{\|x-a\|} = 0, \tag{2.1}$$

then f is said to be *differentiable* at a. The map L is called a *derivative of f at a*. If this is unique, then we denote it by $f'(a)$. If $E \subseteq A$ and f is differentiable at a for all $a \in E$, then we say that f is *differentiable on E*. If f is differentiable on A, then we simply say that f is *differentiable*.

If f is differentiable and its derivative is unique everywhere, then the map $f'\colon A \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ is called the *derivative of f*.

[2] Recall that we may use either neighbourhoods of a in A or in X, cf. Remark 1.29.

Note that for f to be differentiable on E, every point in E must be a limit point of A. In particular, for f to be differentiable every point in A must be a limit point of A. This is for instance the case when A is open or when A is an interval.

REMARK 2.2: Differentiability on $\mathbb{R}$.
In the case where $d = 1$, the quotient in (2.1) goes to zero if and only if

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

goes to zero, i.e. the difference quotient converges to L. Furthermore, if also $m = 1$ then L is multiplication by a scalar, so in this case Definition 2.1 is equivalent to the usual definition of the derivative. ⌟

LEMMA 2.3

Let $A \subseteq \mathbb{R}^d$, let $a \in A$ be a limit point of A, and let $f = (f_1,\dots,f_m)\colon A \to \mathbb{R}^m$. Then f is differentiable at a if and only if each f_i is differentiable at a, and $f'(a) = (f_1'(a),\dots,f_m'(a))$.

[TODO inner points??]

PROOF. If $L = (L_1,\dots,L_m) \in \mathcal{L}(\mathbb{R}^d,\mathbb{R}^m)$, then notice that

$$f(x)-f(a)-L(x-a) = \begin{pmatrix} f_1(x)-f_1(a)-L_1(x-a) \\ \vdots \\ f_m(x)-f_m(a)-L_m(x-a) \end{pmatrix}.$$

The claim then follows since the quotient in (2.1) converges to zero if and only if each quotient

$$\frac{|f_i(x)-f_i(a)-L_i(x-a)|}{\|x-a\|}$$

converges to zero. □

LEMMA 2.4: *Hadamard's lemma*

Let $A \subseteq \mathbb{R}^d$, let $a \in A$ be a limit point of A, and let $f\colon A \to \mathbb{R}^m$. The following are equivalent:

(i) *f is differentiable at a.*

(ii) *There exists a linear map $L(a) \in \mathcal{L}(\mathbb{R}^d,\mathbb{R}^m)$ and a function $\varepsilon_a\colon A-a \to \mathbb{R}^m$ such that*

$$\lim_{h\to 0}\frac{\|\varepsilon_a(h)\|}{\|h\|} = 0, \quad \text{and} \quad f(a+h)-f(a) = L(a)h+\varepsilon_a(h) \qquad (2.2)$$

for all $h \in A-a$.

(iii) *There exists a function* $\varphi = \varphi_a \colon A \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$*, continuous at* a*, such that*

$$f(x) - f(a) = \varphi_a(x)(x-a)$$

for all $x \in A$.

If any of these conditions are satisfied, then $L(a)$ *and* $\varphi_a(a)$ *are derivatives of* f *at* a*. Conversely, to any derivative* $L(a)$ *of* f *at* a *there exists a function* $\varepsilon_a \colon A - a \to \mathbb{R}^m$ *with the properties in* (2.2).

We will call the function $\varphi = \varphi_a$ an *Hadamard function* for f at a.

PROOF. *(i)* $\Rightarrow$ *(ii)*: Let $L(a)$ be a derivative of f at a and put

$$\varepsilon_a(h) = f(a+h) - f(a) - L(a)h$$

for $h \in A - a$. Then we have

$$\frac{\|\varepsilon_a(h)\|}{\|h\|} = \frac{\|f(a+h) - f(a) - L(a)h\|}{\|h\|} \xrightarrow[h\to 0]{} 0$$

as required, since $L(a)$ is a derivative.

(ii) $\Rightarrow$ *(iii)*: Define

$$\varphi_a(x) = \begin{cases} L(a) + \frac{1}{\|x-a\|^2}\varepsilon_a(x-a)(x-a)^\top, & x \in A \setminus \{a\}, \\ L(a), & x = a. \end{cases}$$

We first of all have

$$\begin{aligned} \varphi_a(x)(x-a) &= L(a)(x-a) + \varepsilon_a(x-a)\frac{(x-a)^\top(x-a)}{\|x-a\|^2} \\ &= L(a)(x-a) + \varepsilon_a(x-a) \\ &= f(x) - f(a) \end{aligned}$$

as required. Next notice that, since h and $\varepsilon_a(h)$ are (column) vectors, we have[3]

$$\frac{\|\varepsilon_a(h)h^\top\|}{\|h\|^2} = \frac{\|\varepsilon_a(h)\|\,\|h\|}{\|h\|^2} = \frac{\|\varepsilon_a(h)\|}{\|h\|} \xrightarrow[h\to 0]{} 0,$$

so φ_a is continuous at a.

[3] More generally, let $v = (v_1, \dots, v_d) \in \mathbb{R}^d$ and $w = (w_1, \dots, w_m) \in \mathbb{R}^m$. Then $(wv^\top)_{ij} = w_i v_j$, so

$$\|wv^\top\|^2 = \sum_{i,j} |w_i v_j|^2 = \sum_{i=1}^m |w_i|^2 \sum_{j=1}^d |v_i|^2 = \|w\|^2\|v\|^2.$$

(iii) $\Rightarrow$ *(i)*: Notice that

$$\begin{aligned}\frac{\|f(x)-f(a)-\varphi_a(a)(x-a)\|}{\|x-a\|} &= \frac{\|\varphi_a(x)(x-a)-\varphi_a(a)(x-a)\|}{\|x-a\|} \\ &\leq \frac{\|\varphi_a(x)-\varphi_a(a)\|\,\|x-a\|}{\|x-a\|} \\ &= \|\varphi_a(x)-\varphi_a(a)\| \xrightarrow[x\to a]{} 0,\end{aligned}$$

by continuity of φ_a at a, and continuity of the operator norm, so f is differentiable at a, and $\varphi_a(a)$ is a derivative of f at a. □

PROPOSITION 2.5

Let $A \subseteq \mathbb{R}^d$, let $a \in A$ be a limit point of A, and let $f\colon A \to \mathbb{R}^m$. If f is differentiable at a, then f is continuous at a.

PROOF. Let φ_a be an Hadamard function for f at a, such that

$$f(x) = f(a) + \varphi_a(x)(x-a) \tag{2.3}$$

for $x \in A$. Since φ_a is continuous at a, so is f. □

THEOREM 2.6: *The chain rule*

Let $A \subseteq \mathbb{R}^d$ and $B \subseteq \mathbb{R}^m$, and let $f\colon A \to \mathbb{R}^m$ and $g\colon B \to \mathbb{R}^p$ with $f(A) \subseteq B$. Assume that $a \in A$ is a limit point of A, that $f(a) \in B$ is a limit point of B with derivative $L(a)$, that f is differentiable at a, and that g is differentiable at $f(a)$ with derivative $M(f(a))$. Then $g \circ f$ is differentiable at a, and $M(f(a)) \circ L(a)$ is a derivative of $g \circ f$ at a.

In particular, if the above derivatives of f, g and $g \circ f$ are uniquely determined, then

$$(g \circ f)'(a) = g'(f(a))f'(a).$$

PROOF. Let φ be an Hadamard function for f at a, and let ψ be an Hadamard function for g at $f(a)$. Hence

$$f(x) - f(a) = \varphi(x)(x-a) \quad \text{and} \quad g(y) - g(f(a)) = \psi(y)(y - f(a))$$

for all $x \in A$ and $y \in B$. Letting $y = f(x)$ we thus have

$$g(f(x)) - g(f(a)) = \psi(f(x))(f(x) - f(a)) = \psi(f(x))\varphi(x)(x-a).$$

Notice that the map $x \mapsto \psi(f(x))\varphi(x)$ is continuous at a by Proposition 2.5, so it is an Hadamard function for $g \circ f$ at a. Thus Lemma 2.4 implies that $g \circ f$ is differentiable at a with derivative

$$\psi(f(a))\varphi(a) = M(f(a)) \circ L(a),$$

as claimed. □

We finally turn to uniqueness of derivatives. First a definition:

Definition 2.7: *Directional limit points*

Let V be a topological vector space, and let $A \subseteq V$. For $a, h \in V$ with $h \neq 0$ we have the following:

(i) We say that a is a *limit point of A in the direction of h* if a is a limit point of the set[4] $A \cap (a, a+\delta h)$ for some $\delta > 0$.

(ii) If a is a limit point of A in the direction of both h and $-h$, then we say that a is a *two-sided limit point of A in the direction of h*.

(iii) If a is a limit point of A in the direction of h for all $h \in V \setminus \{0\}$, then we say that a is a limit point of A *in every direction*.

Notice that the value of δ is irrelevant. Notice furthermore that if h and h' are parallel and point in the same direction (i.e. if $h = sh'$ for some $s > 0$), then a is a limit point of A in the direction of h if and only if a is a limit point of A in the direction of h'. Finally notice that if a is a limit point of A in the direction of some h, then A is a limit point of A in the usual sense (this is clear since $A \cap (a, a+\delta h)$ is a subset of A). [TODO is there any issue in complex vector spaces?]

Lemma 2.8: *Uniqueness of derivatives*

Let $A \subseteq \mathbb{R}^d$, let $a \in A$ be a limit point of A, and let $f \colon A \to \mathbb{R}^m$ be differentiable at a. If $\mathcal{B}$ is a basis for $\mathbb{R}^d$ and a is a limit point of A in the direction of each element in $\mathcal{B}$, then the derivative of f at a is uniquely determined. This is the case in particular if either

(i) *$d = 1$, or*

(ii) *a is an inner point of A.*

Proof. Let $L(a)$ be a derivative of f at a, and let ε_a be as in Lemma 2.4. For $b \in \mathcal{B}$ consider the set $I_b = \{t \geq 0 \mid a + tb \in A\}$, and notice that 0 is a limit point of I since a is a limit point of A in the direction of b. Next notice that, by (2.2), we have

$$L(a)b = \frac{1}{t}\Big(f(a+tb) - f(a)\Big) - \frac{1}{t}\varepsilon_a(tb)$$

[4] Recall that if $x, y \in V$, then we define the 'intervals'

$$[x, y] = \{(t-1)x + ty \mid t \in [0,1]\} \quad \text{and} \quad (x, y) = \{(t-1)x + ty \mid t \in (0,1)\}.$$

for $t \in I \setminus \{0\}$. Also by (2.2) the right-most term above goes to zero as $t \downarrow 0$ on $I \setminus \{0\}$, so this implies that

$$L(a)b = \lim_{t \downarrow 0} \frac{1}{t}\big(f(a+tb) - f(a)\big).$$

In particular, $L(a)b$ is uniquely determined by f and a. But since $\mathcal{B}$ is a basis for $\mathbb{R}^d$, this implies that $L(a)$ is uniquely determined by f and a. □

2.2. Extrema and mean value theorems

Proposition 2.9

Let $f : (a,b) \to \mathbb{R}$, and let $c \in (a,b)$. Assume that f is differentiable at c and attains a local extremum at c. Then $f'(c) = 0$.

Proof. Assume for definiteness that f has a local maximum at c, and choose $\delta > 0$ such that $f(x) \leq f(c)$ for $x \in (c-\delta, c+\delta)$. For $x \in (c-\delta, c)$ we have

$$\frac{f(x) - f(c)}{x - c} \geq 0,$$

and letting $x \uparrow c$ we find that $f'(c) \geq 0$. By considering $x \in (c, c+\delta)$ we similarly find that $f'(c) \leq 0$ as desired. □

Lemma 2.10: *Rolle's theorem*

Let $f : [a,b] \to \mathbb{R}$ be a continuous function that is differentiable on (a,b). If $f(a) = f(b)$, then there is a $c \in (a,b)$ such that $f'(c) = 0$.

Proof. If f is constant, then this is obvious. If f is not constant, then since it is continuous it has a local extremum at some $c \in (a,b)$. By Proposition 2.9 we thus have $f'(c) = 0$ as desired. □

Theorem 2.11: *The generalised mean value theorem*

Let $f, g : [a,b] \to \mathbb{R}$ be continuous functions that are differentiable on (a,b). Then there exists a point $c \in (a,b)$ such that

$$\big(f(b) - f(a)\big)g'(c) = \big(g(b) - g(a)\big)f'(c).$$

Proof. Define $h : [a,b] \to \mathbb{R}$ by $h(x) = (f(b) - f(a))g(x) - (g(b) - g(a))f(x)$, and notice that h is continuous on $[a,b]$, differentiable on (a,b), and that

$$h(a) = f(b)g(a) - g(b)f(a) = h(b).$$

Lemma 2.10 thus implies that $h'(c) = 0$ for some $c \in (a,b)$. This proves the claim. □

COROLLARY 2.12: *The mean value theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function that is differentiable on (a,b). Then there is a $c \in (a,b)$ such that

$$f(b) - f(a) = f'(c)(b-a).$$

PROOF. Let $g(x) = x$ in Theorem 2.11. □

[TODO monotonicity]

2.3. Directional and partial derivatives

DEFINITION 2.13: *Directional and partial derivatives*

Let $A \subseteq \mathbb{R}^d$, and let $f\colon A \to \mathbb{R}^m$. If $a \in A$ is a limit point [TODO do I want to require two-sided limit points, and below limits from both directions?] of A in the direction of $v \in \mathbb{R}^d \setminus \{0\}$, then we say that f has a *directional derivative* at a in the direction of v if the limit

$$\lim_{t \downarrow 0} \frac{f(a+tv) - f(a)}{t}$$

exists. The value of this limit is denoted $D_v f(a)$ and is called the *directional derivative* of f at a in the direction of v.

If $v = e_j$, then we write $D_{e_j} f(a) = D_j f(a)$ and call $D_j f(a)$ the *j-th partial derivative* of f at a. If $D_j f(a)$ exists for all j, then f is said to be *partially differentiable* at a. If $E \subseteq A$ and f is partially differentiable at a for all $a \in E$, then we say that f is *partially differentiable on* E. If f is partially differentiable on A, then we simply say that f is *partially differentiable*.

If f is partially differentiable on a set $E \subseteq A$, then the functions $D_i f\colon E \to \mathbb{R}^m$ are called the *partial derivatives* of f (on E).

PROPOSITION 2.14

Let $A \subseteq \mathbb{R}^d$, and let $f\colon A \to \mathbb{R}^m$. If $a \in A$ is a limit point of A in every direction, and if f is differentiable at a, then the following hold:

(i) *The directional derivative $D_v f(a)$ exists for all $v \in \mathbb{R}^d$, and*

$$D_v f(a) = f'(a)v.$$

(ii) *f is partially differentiable at a, and*

$$f'(a)v = \sum_{j=1}^{d} v_j D_j f(a)$$

for all $v = (v_1, \dots v_d) \in \mathbb{R}^d$. In particular, writing $f = (f_1, \dots, f_m)$ the standard matrix representation $J_f(a)$ of $f'(a)$ is given by $J_f(a) = (D_j f_i(a))_{ij}$.

Notice that the derivative of f at a is uniquely determined by Lemma 2.8, so the notation $f'(a)$ for the derivative is justified.

Proof. *Proof of (i):* Let $I_v = \{t \geq 0 \mid a + tv \in A\}$ and define the function $g \colon I_v \to \mathbb{R}$ by $g(t) = f(a + tv)$. Then since the function $t \mapsto a + tv$ is differentiable at 0 with derivative v, the chain rule implies that

$$\lim_{t \downarrow 0} \frac{f(a + tv) - f(a)}{t} = g'(0) = f'(a)v.$$

Proof of (ii): Partial differentiability follows from (i). Writing $v = \sum_{j=1}^{d} v_j e_j$, it follows from (i) that

$$f'(a)v = \sum_{j=1}^{d} v_j f'(a) e_j = \sum_{j=1}^{d} v_j D_{e_j}(a) = \sum_{j=1}^{d} v_j D_j(a). \qquad \square$$

Theorem 2.15: *Criterion for differentiability*

Let $A \subseteq \mathbb{R}^d$, let a be an inner point of A, and $f \colon A \to \mathbb{R}^m$. If f is partially differentiable in a neighbourhood of a and each partial derivative is continuous at a, then f is differentiable at a.

Proof. We may assume that $m = 1$ by Lemma 2.3. Let U be an open neighbourhood of a on which f is partially differentiable. Let $h = (h_1, \dots, h_d) \in \mathbb{R}^d$, and for $k = 1, \dots, d$ let $h^{(k)} = \sum_{j=1}^{k} h_j e_j$. Notice that $h^{(d)} = h$. Choose h small enough such that $a + h^{(k)} \in U$ for all k. Next let $I_k \subseteq \mathbb{R}$ be an open interval containing 0 and h_k such that the map

$$\begin{aligned} g_k \colon I_k &\to \mathbb{R}, \\ t &\mapsto f(a + h^{(k-1)} + t e_k), \end{aligned}$$

is differentiable. Applying the mean value theorem on g_k yields an s_k in the interval between 0 and h_k such that

$$g_k(h_k) - g_k(0) = g_k'(s_k) h_k = D_k f(a + h^{(k-1)} + s_k e_k) h_k.$$

Next define a map $\varphi_a \colon U \to \mathcal{L}(\mathbb{R}^d, \mathbb{R})$ by

$$\varphi_a(a+h)x = \sum_{k=1}^{d} D_k f(a + h^{(k-1)} + s_k e_k) x_k$$

for $h \in U - a$ and $x = (x_1, \dots, x_d) \in \mathbb{R}^d$. This is clearly well-defined, and since the partial derivatives of f are continuous at a, so is φ_a (notice that s_k also goes to zero when h goes to zero).

Finally notice that

$$\begin{aligned} f(a+h) - f(a) &= \sum_{k=1}^{d} \bigl(f(a + h^{(k)}) - f(a + h^{(k-1)}) \bigr) \\ &= \sum_{k=1}^{d} \bigl(g_k(h_k) - g_k(0) \bigr) \\ &= \sum_{k=1}^{d} D_k f(a + h^{(k-1)} + s_k e_k) h_k \\ &= \varphi_a(a+h)h. \end{aligned}$$

But then φ_a is an Hadamard function for f at a, so f is differentiable at a by Lemma 2.4. □

2.4. Continuous differentiability

Since $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ is a finite-dimensional vector space, it has a unique vector space topology [TODO ref to my notes on topological vector spaces]. Hence it makes sense to talk about the continuity of maps into or out of $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$. In particular, if the derivative f' of f is continuous, then we say that f is *continuously differentiable*. The subset of $C(A, \mathbb{R}^m)$ containing the continuously differentiable functions is denoted $C^1(A, \mathbb{R}^m)$.

Corollary 2.16

Let $U \subseteq \mathbb{R}^d$ be open and let $f \colon U \to \mathbb{R}^m$. Then f is continuously differentiable if and only if f is partially differentiable and all of its partial derivatives are continuous.

Proof. If f is partially differentiable, then $D_j f$ equals the function $a \mapsto f'(a)e_j$ by Proposition 2.14(ii). If f is differentiable then it is partially differentiable, and if f' is continuous then so are the $D_j f$.

Conversely, if f is partially differentiable with continuous partial derivatives, then Theorem 2.15 implies that f is differentiable. Continuity of f' now follows since its coordinate functions are continuous. □

2.5. Higher-order derivatives

THEOREM 2.17: *Clairaut's Theorem*

Let $A \subseteq \mathbb{R}^d$ be open, let a be an inner point of A, and let $f\colon A \to \mathbb{R}^m$. For $i,j \in \{1,\dots,d\}$, if D_jD_if and D_iD_jf exist in a neighbourhood of a and are continuous at a, then

$$D_jD_if(a) = D_iD_jf(a).$$

PROOF. By permuting indices, we may assume that $i = 1$ and $j = 2$, and indeed that $d = 2$. Furthermore, by Lemma 2.3 it suffices to prove the theorem for $m = 1$.

Let U be an open neighbourhood of a on which D_jD_if and D_iD_jf exist, and let $x \in U$. We may assume that U is on the form $U_1 \times U_2$ for open sets $U_1, U_2 \subseteq \mathbb{R}$. Writing $a = (a_1, a_2)$ and $x = (x_1, x_2)$, define

$$r(x) = \frac{f(x) - f(a_1, x_2) - f(x_1, a_2) + f(a)}{(x_1 - a_1)(x_2 - a_2)}$$

for $x \neq a$. Let $I \subseteq U_1$ be an open interval containing a_1 and x_1, and define $g\colon I \to \mathbb{R}$ by $g(t) = f(t, x_2) - f(t, a_2)$. Then

$$r(x) = \frac{g(x_1) - g(a_1)}{(x_1 - a_1)(x_2 - a_2)}.$$

Since g is differentiable on I, the mean value theorem yields an $s_1 \in \mathbb{R}$ between a_1 and x_1 such that

$$r_(x) = \frac{g'(s_1)}{x_2 - a_2} = \frac{D_1f(s_1, x_2) - D_1f(s_1, a_2)}{x_2 - a_2}.$$

Similarly, let $J \subseteq U_2$ be an open interval containing a_2 and x_2, and define $h\colon J \to \mathbb{R}$ by $h(t) = D_1f(s_1, t)$. Again the mean value theorem yields an $s_2 \in \mathbb{R}$ between x_2 and a_2 such that

$$r(x) = h'(s_2) = D_2D_1f(s_1, s_2).$$

Let $s = (s_1, s_2)$. Since $\|s - a\| \leq \|x - a\|$, when $x \to a$ we also have $s \to a$. Continuity of D_2D_1f at a thus implies that

$$\lim_{x \to a} r(x) = \lim_{s \to a} D_2D_1f(s) = D_2D_1f(a).$$

Finally notice that the above argument is symmetric in the indices 1 and 2, so we similarly find that $\lim_{x \to a} r(x) = D_1D_2f(a)$. $\square$

3 • Integration

3.1. Functions of bounded variation

A *partition* of an interval $[a,b]$ is a collection $P = \{x_0, \dots, x_n\}$ of real numbers such that

$$a = x_0 < \cdots < x_n = b.$$

In turn, a *tagged partition* of $[a,b]$ is a pair (P,T) where P is a partition of $[a,b]$ and $T = \{t_1, \dots, t_n\}$ is a multiset of numbers such that $t_i \in [x_{i-1}, x_i]$ for all $i = 1, \dots, n$. Let $\mathcal{P}'[a,b]$ denote the set of tagged partitions of $[a,b]$. We define a direction on $\mathcal{P}'[a,b]$ by $(P,T) \preceq (P',T')$ if $P \subseteq P'$. Notice that T and T' do not appear in this definition. This also induces a direction on the set $\mathcal{P}[a,b]$ of all (non-tagged) partitions of $[a,b]$.

Given a partition $P = \{x_0, \dots, x_n\}$ of $[a,b]$ and a function $f\colon [a,b] \to \mathbb{R}$ we write $\Delta f_i = f(x_i) - f(x_{i-1})$ for $i = 1, \dots, n$. We furthermore write $\Delta x_i = x_i - x_{i-1}$. Furthermore, define

$$\|P\| = \max_{1 \leq i \leq n} \Delta x_i \quad \text{and} \quad \Sigma_f(P) = \sum_{i=1}^{n} |\Delta f_i|.$$

The number $\|P\|$ is called the *norm* of P. Notice that the map $P \mapsto \|P\|$ is decreasing, while the map $P \mapsto \Sigma_f(P)$ is increasing.

Definition 3.1: *Total variation*

Consider a function $f\colon [a,b] \to \mathbb{R}$. The *total variation* of f on $[a,b]$ is the number

$$V_f(a,b) = \sup_{P \in \mathcal{P}[a,b]} \Sigma_f(P).$$

If $V_f(a,b) < \infty$, then we say that f is of *bounded variation* on $[a,b]$. The set of all functions that are of bounded variation on $[a,b]$ is denoted $BV[a,b]$.

If f is of bounded variation on $[a,b]$, then it is clear that f is also of bounded variation on any subinterval of $[a,b]$. If $c \in (a,b)$ it is also easy to show that

$$V_f(a,b) = V_f(a,c) + V_f(c,b). \tag{3.1}$$

If $g\colon [a,b] \to \mathbb{R}$ is another function of bounded variation on $[a,b]$ and $c \in \mathbb{R}$, then it is clear from the definition that $f+g$ and cf are also of bounded variation. It is also simple to show that the product fg is of bounded variation. Hence $BV[a,b]$ is an $\mathbb{R}$-algebra.

Also note that monotonic functions are of bounded variation on any compact interval.

Lemma 3.2

Let $f\colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $V(x) = V_f(a,x)$ for $x \in (a,b]$ and $V(a) = 0$. Then the functions V and $V - f$ are increasing on $[a,b]$.

Proof. The function V is clearly increasing, so consider the function $D = V - f$. Let $x, y \in [a,b]$ with $x < y$, and notice that $f(y) - f(x) \leq V_f(x,y)$. Recalling (3.1) it follows that

$$D(y) - D(x) = V(y) - V(x) - (f(y) - f(x)) = V_f(y,x) - (f(y) - f(x)) \geq 0. \qquad \square$$

Proposition 3.3

A function $f\colon [a,b] \to \mathbb{R}$ is of bounded variation if and only if it is the difference of two (strictly) increasing functions.

Proof. By Lemma 3.2 we can write f as the difference of two increasing functions as $f = V - (V - f)$. Adding a strictly increasing function to both V and $V - f$ yields the claim. $\square$

Proposition 3.4

Let $f\colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $V(x) = V_f(a,x)$ for $x \in (a,b]$ and $V(a) = 0$. If $c \in [a,b]$, then f is continuous at c if and only if V is continuous at c.

Proof. We prove the claim in the case where c is an inner point of $[a,b]$. First assume that V is continuous at c. Since V is monotonic by Lemma 3.2, the left- and right-hand limits $V(c-)$ and $V(c+)$ exist. For $x \in (c,b]$ we have

$$|f(x) - f(c)| \leq V(x) - V(c) \xrightarrow[x \downarrow c]{} V(c+) - V(c) = 0,$$

so f is right-continuous at c. Similarly for left-continuity.

We prove the converse, so assume that f is continuous at c, and let $\varepsilon > 0$. There exists a $\delta > 0$ such that $0 < |x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$. For any partition P of $[c,b]$ we have

$$0 \leq V(x_1) - V(c) = V_f(c,b) - V_f(x_1,b).$$

To bound the right-hand side, choose $P = \{x_0, \dots, x_n\}$ such that

$$V_f(c,b) - \varepsilon < \sum_{i=1}^{n} |\Delta f_i|.$$

This inequality is conserved when adding points to P, so we may assume that $x_1 - x_0 < \delta$, implying that $|\Delta f_1| < \varepsilon$. The inequality above then becomes

$$V_f(c,b) - \varepsilon < \varepsilon + \sum_{i=2}^{n} |\Delta f_i| \le \varepsilon + V_f(x_1,b),$$

so that $V_f(c,b) - V_f(x_1,b) < 2\varepsilon$. Hence $V(x_1) - V(c) \le 2\varepsilon$, so $V(c+) = V(c)$ since ε was arbitrary and V is increasing. A similar argument yields $V(c-) = V(c)$, so V is continuous at c. ◻

3.2. Integration

Next consider two functions $f, \alpha \colon [a,b] \to \mathbb{R}$. For each tagged partition (P,T) of $[a,b]$ we define the *Riemann–Stieltjes sum*

$$S_{f,\alpha}(P,T) = \sum_{i=1}^{n} f(t_i)\Delta\alpha_i.$$

This induces a net $S_{f,\alpha} \colon \mathcal{P}'[a,b] \to \mathbb{R}$.

Definition 3.5: *Riemann–Stieltjes integral*

Let $f, \alpha \colon [a,b] \to \mathbb{R}$ be bounded functions. We say that f is *Riemann-integrable* with respect to α (or simply α*-integrable*) on $[a,b]$ if the net $S_{f,\alpha}$ has a limit $A \in \mathbb{R}$. In this case A is called the *Riemann–Stieltjes integral* of f with respect to α on $[a,b]$ and is denoted

$$\int_a^b f \,\mathrm{d}\alpha \quad \text{or} \quad \int_a^b f(x)\,\mathrm{d}\alpha(x).$$

We denote the set of α-integrable functions on $[a,b]$ by $\mathcal{R}_\alpha[a,b]$.

We call f the *integrand* and α the *integrator*. In the case where $\alpha(x) = x$, we use the notations

$$S_f, \quad \int_a^b f \quad \text{and} \quad \int_a^b f(x)\,\mathrm{d}x.$$

The sums S_f are then simply called *Riemann sums* and the integral the *Riemann integral* of f on $[a,b]$. With this choice of α, an α-integrable function is called *Riemann integrable* on $[a,b]$, and the set of such functions is denoted $\mathcal{R}[a,b]$.

Remark 3.6. In the ordinary Riemann integral it is not necessary to assume that f is bounded, since integrability in this case implies boundedness. We claim that this assumption can be lifted whenever the integrator α is injective. In fact, we only need to assume that there is a partition Q of $[a,b]$ such that α is injective on each subinterval of Q.

If the net $S_{f,\alpha}$ has a limit $A \in \mathbb{R}$, then there is a tagged partition (P, T) of $[a, b]$ such that

$$|A - S_{f,\alpha}(P', T')| < 1, \quad \text{implying that} \quad |S_{f,\alpha}(P', T')| < M := |A| + 1,$$

whenever $(P, T) \preceq (P', T')$. By replacing P with $P \cup Q$ we may assume that α is injective on each subinterval of P, since α is clearly also injective on each subinterval of $P \cup Q$.

Write $P = \{x_0, \dots, x_n\}$ and $T = \{t_1, \dots, t_n\}$. Let $x \in [x_{k-1}, x_k]$ and consider the multiset $T_x = \{t_1, \dots, t_{k-1}, x, t_{k+1}, \dots, t_n\}$ and the corresponding tagged partition (P, T_x). We then have

$$f(x)\Delta\alpha_k = S_{f,\alpha}(P, T_x) - \sum_{i \neq k} f(t_i)\Delta\alpha_i.$$

Since α is injective on $[x_{k-1}, x_k]$ we have $\Delta\alpha_k \neq 0$, so the above implies that

$$|f(x)| = \frac{1}{|\Delta\alpha_k|}\left|S_{f,\alpha}(P, T_x) - \sum_{i \neq k} f(t_i)\Delta\alpha_i\right| < \frac{1}{|\Delta\alpha_k|}\left(M + \left|\sum_{i \neq k} f(t_i)\Delta\alpha_i\right|\right),$$

where the inequality follows since $(P, T) \preceq (P, T_x)$ for all $x \in [x_{k-1}, x_k]$. The right-hand side is thus as upper bound for $|f|$ on $[x_{k-1}, x_k]$, and there are finitely many such intervals, so f is bounded on $[a, b]$ as claimed. ⌟

Below we fix an interval $[a, b]$ and (bounded) integrators α and β on it.

Proposition 3.7: *Linearity of the integral*

Let $f, g \in \mathcal{R}_\alpha[a, b]$ and $c_1, c_2 \in \mathbb{R}$. Then:

(i) *$c_1 f + c_2 g$ is α-integrable on $[a, b]$ and*

$$\int_a^b (c_1 f + c_2 g)\,\mathrm{d}\alpha = c_1 \int_a^b f\,\mathrm{d}\alpha + c_2 \int_a^b g\,\mathrm{d}\alpha.$$

In particular, $\mathcal{R}_\alpha[a, b]$ is a vector space.

(ii) *f is $(c_1\alpha + c_2\beta)$-integrable on $[a, b]$ and*

$$\int_a^b f\,\mathrm{d}(c_1\alpha + c_2\beta) = c_1 \int_a^b f\,\mathrm{d}\alpha + c_2 \int_a^b f\,\mathrm{d}\beta.$$

Proof. This follows immediately from the bilinearity of the map $(f, \alpha) \mapsto S_{f,\alpha}$ along with basic properties of nets. □

PROPOSITION 3.8

Consider $f, \alpha \colon [a,b] \to \mathbb{R}$ and let $c \in (a,b)$. If two of the three integrals in (3.2) exist, then so does the third and we have

$$\int_a^c f \, \mathrm{d}\alpha + \int_c^b f \, \mathrm{d}\alpha = \int_a^b f \, \mathrm{d}\alpha. \tag{3.2}$$

PROOF. Let (P,T) be a tagged partition of $[a,b]$ such that $c \in P$, and let $P_1 = P \cap [a,c]$ and $P_2 = P \cap [c,b]$, and $T_1 = T \cap [a,c]$ and $T_2 = T \cap [c,b]$. Then

$$S_{f,\alpha}(P,T) = S_{f,\alpha}(P_1,T_1) + S_{f,\alpha}(P_2,T_2). \tag{3.3}$$

If two of the three integrals in (3.2) exist, then two of the nets in (3.3) converge to the respective integrals. The third net then also converges to the relevant integral: Notice that we assume without loss of generality that $c \in P$, since adding c to a partition simply yields a finer partition. Also notice that any partition P' of e.g. $[a,c]$ is on the form $P' = P \cap [a,c]$ for some partition P of $[a,b]$. □

PROPOSITION 3.9: *Integration by parts*

Given functions $f, \alpha \colon [a,b] \to \mathbb{R}$, assume that $f \in \mathcal{R}_\alpha[a,b]$. Then $\alpha \in \mathcal{R}_f[a,b]$ and

$$\int_a^b f \, \mathrm{d}\alpha + \int_a^b \alpha \, \mathrm{d}f = f(b)\alpha(b) - f(a)\alpha(a).$$

In particular, exchanging f and α we have

$$f \in \mathcal{R}_\alpha[a,b] \quad \textit{if and only if} \quad \alpha \in \mathcal{R}_f[a,b].$$

PROOF. Let (P,T) be a tagged partition of $[a,b]$ with $P = \{x_0, \ldots, x_n\}$ and $T = \{t_1, \ldots, t_n\}$. Writing $A = f(b)\alpha(b) - f(a)\alpha(a)$, notice that

$$A = \sum_{i=1}^n f(x_i)\alpha(x_i) - \sum_{i=1}^n f(x_{i-1})\alpha(x_{i-1}),$$

and that

$$S_{\alpha,f}(P,T) = \sum_{i=1}^n \alpha(t_i)\Delta f_i = \sum_{i=1}^n f(x_i)\alpha(t_i) - \sum_{i=1}^n f(x_{i-1})\alpha(t_i).$$

Hence we have

$$\begin{aligned} A - S_{\alpha,f}(P,T) &= \sum_{i=1}^n f(x_i)(\alpha(x_i) - \alpha(t_i)) + \sum_{i=1}^n f(x_{i-1})(\alpha(t_i) - \alpha(x_{i-1})) \\ &= S_{f,\alpha}(P \cup T, P'), \end{aligned}$$

where P' is obtained from P by duplicating[5] appropriate elements such that each subinterval of $P \cup T$ contains the corresponding element from P'. Notice that if P and T have any elements in common, these are not duplicated in the union $P \cup T$. However, in this case the corresponding terms in the sum above vanish, so the last equality does in fact hold. Since $P \cup T$ is finer than P, the claim follows by taking the limit of $S_{\alpha,f}$. □

PROPOSITION 3.10: *Change of variables in Riemann–Stieljes integrals*

Let $f \in \mathcal{R}_\alpha[a,b]$, and let $\varphi\colon I \to [a,b]$ be a monotonic (or equivalently continuous) bijection where I is an interval with endpoints c and d. Assume that $a = \varphi(c)$ and $b = \varphi(d)$. Then $f \circ \varphi \in \mathcal{R}_{\alpha\circ\varphi}[c,d]$ and

$$\int_a^b f \,\mathrm{d}\alpha = \int_c^d f \circ \varphi \,\mathrm{d}(\alpha \circ \varphi).$$

PROOF. Since φ is bijective it is strictly monotonic, so it induces an order isomorphism $\mathcal{P}'(I) \to \mathcal{P}'[a,b]$ given by $(P,T) \mapsto (\varphi(P), \varphi(T))$.

Now let $(P,T) \in \mathcal{P}'(I)$ with $P = \{y_0, \ldots, y_n\}$ and $T = \{s_1, \ldots, s_n\}$. Assume for definiteness that φ is increasing so that $I = [c,d]$, and write $x_i = \varphi(y_i)$ and $t_i = \varphi(s_i)$. Then we have

$$S_{f\circ\varphi,\alpha\circ\varphi}(P,T) = \sum_{i=1}^n (f \circ \varphi)(s_i)\Delta(\alpha\circ\varphi)_i = \sum_{i=1}^n f \circ \varphi(t_i)\Delta\alpha_i = S_{f,\alpha}(\varphi(P), \varphi(T)).$$

Since the map $(P,T) \mapsto (\varphi(P), \varphi(T))$ is an order isomorphism, each side above converges to the corresponding integral, proving the claim. □

PROPOSITION 3.11: *Reduction to a Riemann integral*

Let $\alpha \in C^1[a,b]$ and $f \in \mathcal{R}_\alpha[a,b]$. Then $f\alpha' \in \mathcal{R}[a,b]$, and

$$\int_a^b f \,\mathrm{d}\alpha = \int_a^b f\alpha'.$$

PROOF. Consider the Riemann(–Stieltjes) sums

$$S_{f\alpha'}(P,T) = \sum_{i=1}^n f(t_i)\alpha'(t_i)\Delta x_i \quad \text{and} \quad S_{f,\alpha}(P,T) = \sum_{i=1}^n f(t_i)\Delta\alpha_i.$$

[5] Recall that if (P,T) is a tagged partition, then T is a *multiset*.

By the mean value theorem we can write $\Delta\alpha_i = \alpha'(s_i)\Delta x_i$ for appropriate $s_i \in (x_{i-1}, x_i)$. It follows that

$$S_{f,\alpha}(P,T) - S_{f\alpha'}(P,T) = \sum_{i=1}^{n} f(t_i)(\alpha'(s_i) - \alpha'(t_i))\Delta x_i.$$

By uniform continuity of α', given $\varepsilon > 0$ there exists a $\delta > 0$ such that $|x-y| < \delta$ implies $|\alpha'(x) - \alpha'(y)| < \varepsilon$ for all $x, y \in [a,b]$. Choosing a tagged partition (P,T) with $\|P\| < \delta$ we thus have

$$|S_{f,\alpha}(P,T) - S_{f\alpha'}(P,T)| \leq \|f\|_{\sup}\varepsilon(b-a).$$

Since ε was arbitrary, the left-hand side converges to zero. It follows that

$$\left|\int_a^b f \,\mathrm{d}\alpha - S_{f\alpha'}(P,T)\right| \leq \left|\int_a^b f \,\mathrm{d}\alpha - S_{f,\alpha}(P,T)\right| + |S_{f,\alpha}(P,T) - S_{f\alpha'}(P,T)|,$$

which converges to zero. Hence $S_{f\alpha'}(P,T)$ converges, so $f\alpha' \in \mathcal{R}[a,b]$, and its integral equals the α-integral of f as claimed. □

3.3. *Increasing integrators*

Definition 3.12

Let $f, \alpha\colon [a,b] \to \mathbb{R}$ be bounded functions, and assume that α is increasing. Let $P = \{x_0, \dots, x_n\}$ be a partition of $[a,b]$, and let

$$M_i(f) = \sup\Big\{f(x) \;\Big|\; x \in [x_{i-1}, x_i]\Big\},$$
$$m_i(f) = \inf\Big\{f(x) \;\Big|\; x \in [x_{i-1}, x_i]\Big\}.$$

The numbers

$$U_{f,\alpha}(P) = \sum_{i=1}^{n} M_i(f)\Delta\alpha_i \quad \text{and} \quad L_{f,\alpha}(P) = \sum_{i=1}^{n} m_i(f)\Delta\alpha_i$$

are called the *upper and lower Stieltjes sums* of f with respect to α for the partition P.

Since α is increasing we have $\Delta\alpha_i \geq 0$, so it is immediate that

$$L_{f,\alpha}(P) \leq S_{f,\alpha}(P,T) \leq U_{f,\alpha}(P)$$

for any tagged partition (P,T) of $[a,b]$. It is also clear that, if $P \subseteq P'$, then

$$U_{f,\alpha}(P) \geq U_{f,\alpha}(P') \quad \text{and} \quad L_{f,\alpha}(P) \leq L_{f,\alpha}(P'), \tag{3.4}$$

and that for any pair of partitions P_1 and P_2 we have

$$L_{f,\alpha}(P_1) \leq U_{f,\alpha}(P_2). \tag{3.5}$$

Definition 3.13

Let $f, \alpha \colon [a,b] \to \mathbb{R}$ be bounded functions with α increasing. Then the numbers

$$\overline{\int_a^b} f \, \mathrm{d}\alpha = \inf\bigl\{ U_{f,\alpha}(P) \bigm| P \in \mathcal{P}[a,b] \bigr\}$$

and

$$\underline{\int_a^b} f \, \mathrm{d}\alpha = \sup\bigl\{ L_{f,\alpha}(P) \bigm| P \in \mathcal{P}[a,b] \bigr\}$$

are called the *upper and lower Stieltjes integrals* of f with respect to α on $[a,b]$.

We also use the notations $\overline{I}(f,\alpha)$ and $\underline{I}(f,\alpha)$ for the upper and lower integrals, respectively, when the interval $[a,b]$ is understood. It follows immediately from the definition and (3.5) that $\underline{I}(f,\alpha) \leq \overline{I}(f,\alpha)$.

Theorem 3.14: *Riemann's condition*

Let $f, \alpha \colon [a,b] \to \mathbb{R}$ be bounded functions with α increasing. Then the following conditions are equivalent:

(i) $f \in \mathcal{R}_\alpha[a,b]$.

(ii) f *satisfies* Riemann's condition *with respect to α on $[a,b]$: For every $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that*

$$U_{f,\alpha}(P) - L_{f,\alpha}(P) < \varepsilon. \tag{3.6}$$

(iii) $\underline{I}(f,\alpha) = \overline{I}(f,\alpha)$.

In this case we have

$$\underline{\int_a^b} f \, \mathrm{d}\alpha = \int_a^b f \, \mathrm{d}\alpha = \overline{\int_a^b} f \, \mathrm{d}\alpha.$$

Proof. *(i) ⇒ (ii)*: Let $\varepsilon > 0$, and choose a partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$ such that

$$\left| \sum_{i=1}^n f(t_i) \Delta\alpha_i - \int_a^b f \, \mathrm{d}\alpha \right| < \varepsilon$$

for all $t_i \in [x_{i-1}, x_i]$. It follows that

$$\left| \sum_{i=1}^n (f(t_i) - f(t_i')) \Delta\alpha_i \right| < 2\varepsilon$$

for all $t_i, t_i' \in [x_{i-1}, x_i]$. For any $\delta > 0$ there exist t_i, t_i' such that

$$f(t_i) - f(t_i') > M_i(f) - m_i(f) - \delta.$$

From this it follows that

$$\begin{aligned}U_{f,\alpha}(P) - L_{f,\alpha}(P) &= \sum_{i=1}^{n} (M_i(f) - m_i(f))\Delta\alpha_i \\ &< \sum_{i=1}^{n} (f(t_i) - f(t_i'))\Delta\alpha_i + \delta(\alpha(b) - \alpha(a)) \\ &< 3\varepsilon\end{aligned}$$

for an appropriate choice of δ. Since ε was arbitrary, this proves (ii).

(ii) $\Rightarrow$ (iii): If P is any partition of $[a,b]$ we have

$$L_{f,\alpha}(P) \leq \underline{\int_a^b} f \,\mathrm{d}\alpha \leq \overline{\int_a^b} f \,\mathrm{d}\alpha \leq U_{f,\alpha}(P).$$

Thus (3.6) implies that $0 \leq \overline{I}(f,\alpha) - \underline{I}(f,\alpha) < \varepsilon$ for every $\varepsilon > 0$, proving (iii).

(iii) $\Rightarrow$ (i): Let $\varepsilon > 0$. There exists a partition P of $[a,b]$ such that

$$\underline{I}(f,\alpha) - \varepsilon < L_{f,\alpha}(P) \leq S_{f,\alpha}(P,T) \leq U_{f,\alpha}(P) < \overline{I}(f,\alpha) + \varepsilon$$

for any choice of points T such that (P,T) is a tagged partition. Denoting the common value of $\underline{I}(f,\alpha)$ and $\overline{I}(f,\alpha)$ by A, this shows that $|S_{f,\alpha}(P',T') - A| < \varepsilon$ for all tagged partitions (P',T') with $P \subseteq P'$. Hence $f \in \mathcal{R}_\alpha[a,b]$, and the integral of f with respect to α equals A. □

3.4. *Mean value theorems for Riemann–Stieltjes integrals*

PROPOSITION 3.15: *The first mean value theorem*

Let $\alpha\colon [a,b] \to \mathbb{R}$ be increasing, and let $f \in \mathcal{R}_\alpha[a,b]$. Let $m = \inf_{x\in[a,b]} f(x)$ and $M = \sup_{x\in[a,b]} f(x)$. Then there exists a $c \in \mathbb{R}$ with $m \leq c \leq M$ such that

$$\int_a^b f \,\mathrm{d}\alpha = c \int_a^b \mathrm{d}\alpha = c(\alpha(b) - \alpha(a)).$$

In particular, if f is continuous then $c = f(x_0)$ for some $x_0 \in [a,b]$.

PROOF. If $\alpha(a) = \alpha(b)$ then both sides are zero, so assume that $\alpha(a) < \alpha(b)$. From the inequalities

$$m(\alpha(b) - \alpha(a)) \leq \int_a^b f \,\mathrm{d}\alpha \leq M(\alpha(b) - \alpha(a))$$

follow that

$$m \leq \frac{1}{\alpha(b) - \alpha(a)} \int_a^b f \, \mathrm{d}\alpha \leq M,$$

so defining c as the middle term, the claim follows. □

PROPOSITION 3.16: *The second mean value theorem*

Let $f, \alpha \colon [a, b] \to \mathbb{R}$ with f monotonic and α continuous. Then there exists a point $x_0 \in [a, b]$ such that

$$\int_a^b f \, \mathrm{d}\alpha = f(a) \int_a^{x_0} \mathrm{d}\alpha + f(b) \int_{x_0}^b \mathrm{d}\alpha.$$

PROOF. By replacing f with $-f$ we may assume that f is increasing. Then f is of bounded variation, so Proposition 3.9 implies that $f \in \mathcal{R}_\alpha[a, b]$. Another application of this proposition along with Proposition 3.15 yields a $x_0 \in [a, b]$

$$\begin{aligned}
\int_a^b f \, \mathrm{d}\alpha &= f(b)\alpha(b) - f(a)\alpha(a) - \int_a^b \alpha \, \mathrm{d}f \\
&= f(b)\alpha(b) - f(a)\alpha(a) - \alpha(x_0)(f(b) - f(a)) \\
&= f(a)(\alpha(x_0) - \alpha(a)) + f(b)(\alpha(b) - \alpha(x_0)),
\end{aligned}$$

as desired. □

We attempt to understand the two mean value theorems. In the case where $\alpha(a) < \alpha(b)$, the first essentially says that the 'α-average' of f, given by

$$\frac{1}{\alpha(b) - \alpha(a)} \int_a^b f \, \mathrm{d}\alpha,$$

lies between the minimum and maximum of f, as we would expect. Notice that the theorem yields an element c of the *codomain* of f that serves as the 'mean value' of f.

As for the second theorem, if f is monotonic then the integral of f is a weighted sum of the values of f at the endpoints. The theorem says that there is an element x_0 of the *domain* that splits up the interval $[a, b]$ such that the corresponding integrals gives the desired weights. This element x_0 in some sense serves as a 'mean' of f.

3.5. *Integrators of bounded variation*

THEOREM 3.17

Let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $V(x) = V_\alpha(a,x)$ for $x \in (a,b]$ and $V(a) = 0$. Then $\mathcal{R}_\alpha[a,b] \subseteq \mathcal{R}_V[a,b]$.

PROOF. Let $f \in \mathcal{R}_\alpha[a,b]$, and choose $M > 0$ such that $|f| \leq M$. Choose a partition $P = \{x_0,\dots,x_n\}$ of $[a,b]$ such that $V(b) < \sum_{i=1}^n |\Delta\alpha_i| + \varepsilon$. Then

$$\begin{aligned}\sum_{i=1}^n (M_i(f) - m_i(f))(\Delta V_i - |\Delta\alpha_i|) &\leq 2M \sum_{i=1}^n (\Delta V_i - |\Delta\alpha_i|) \\ &= 2M\biggl(V(b) - \sum_{i=1}^n |\Delta\alpha_i|\biggr) \\ &< 2M\varepsilon.\end{aligned}$$

Also choose P such that $|\sum_{i=1}^n (f(t_i) - f(t_i'))\Delta\alpha_i| < \varepsilon$ for all $t_i, t_i' \in [x_{i-1}, x_i]$. Next let $\delta > 0$. For $i = 1,\dots,n$, if $\Delta\alpha_i \geq 0$ choose t_i, t_i' such that

$$f(t_i) - f(t_i') > M_i(f) - m_i(f) - \delta.$$

If instead $\Delta\alpha_i < 0$, choose t_i, t_i' such that

$$f(t_i') - f(t_i) > M_i(f) - m_i(f) - \delta.$$

It follows that

$$\sum_{i=1}^n (M_i(f) - m_i(f))|\Delta\alpha_i| < \sum_{i=1}^n (f(t_i) - f(t_i'))\Delta\alpha_i + \delta V(b) < 2\varepsilon$$

for an appropriate choice of δ. Combining these inequalities yields

$$U_{f,V}(P) - L_{f,V}(P) = \sum_{i=1}^n (M_i(f) - m_i(f))\Delta V_i < 2(M+1)\varepsilon,$$

and since ε was arbitrary, this shows that $f \in \mathcal{R}_V[a,b]$. □

Since $\alpha = V - (V - \alpha)$ and both V and $V - \alpha$ are increasing, this allows us to reduce questions about integrators of bounded variation to questions about monotonic integrators. In particular it lets us use Riemann's condition to prove integrability with respect to integrators of bounded variation.

COROLLARY 3.18

Let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $f \in \mathcal{R}_\alpha[a,b]$. Then $f \in \mathcal{R}_\alpha[c,d]$ for every subinterval $[c,d]$ of $[a,b]$.

PROOF. By Theorem 3.17 and Proposition 3.7, it suffices to prove the claim when α is increasing. Furthermore, by Proposition 3.8 it suffices to show that f is α-integrable on $[a,x]$ for all $x \in (a,b]$. Given $\varepsilon > 0$, by Theorem 3.14(ii) there is a partition P of $[a,b]$ such that

$$U_{f,\alpha}(P) - L_{f,\alpha}(P) < \varepsilon.$$

By (3.4), adjoining x to P preserves the inequality. Writing $P = \{x_0, \dots, x_n\}$ we may thus assume that $x = x_k$ for some $k \in \{1, \dots, n\}$. Letting $P' = P \cap [a,x]$ we thus have

$$\begin{aligned} U_{f,\alpha}(P') - L_{f,\alpha}(P') &= \sum_{i=1}^{k} (M_i(f) - m_i(f))\Delta\alpha_i \leq \sum_{i=1}^{n} (M_i(f) - m_i(f))\Delta\alpha_i \\ &= U_{f,\alpha}(P) - L_{f,\alpha}(P) < \varepsilon, \end{aligned}$$

where the first inequality follows since every term in the second sum is non-negative. Thus f is integrable on $[a,x]$. □

PROPOSITION 3.19

Let $f, \alpha \colon [a,b] \to \mathbb{R}$ be functions with f continuous and α of bounded variation. Then f is α-integrable, and α is f-integrable.

PROOF. We may assume that α is increasing. Let $\varepsilon > 0$. Uniform continuity of f furnishes a $\delta < 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$ for $x, y \in [a,b]$. Let $P = \{x_0, \dots, x_n\}$ be a partition with $\|P\| < \delta$. Then $M_i(f) - m_i(f) \leq \varepsilon$, implying that

$$U_{f,\alpha}(P) - L_{f,\alpha}(P) = \sum_{i=1}^{n} (M_i(f) - m_i(f))\Delta\alpha_i \leq \varepsilon(\alpha(b) - \alpha(a)),$$

and since ε was arbitrary, it follows from Riemann's condition that $f \in \mathcal{R}_\alpha[a,b]$. The final claim follows from Proposition 3.9. □

For functions $f \colon [a,b] \to \mathbb{R}$ and $g \colon f([a,b]) \to \mathbb{R}$, when is the composition $g \circ f$ integrable? In the Lebesgue theory all that is required is that f and g are measurable, but as far as I know, no such result is available for the Riemann integral. However, in the case where g is continuous, Proposition 3.20 below shows that $g \circ f$ is indeed integrable. We first prove this and then prove a couple of important corollaries.

PROPOSITION 3.20

Let $\alpha \colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $f \in \mathcal{R}_\alpha[a,b]$. Choose $m, M \in \mathbb{R}$

such that $m \leq f \leq M$. If $\varphi\colon [m,M] \to \mathbb{R}$ is continuous, then $\varphi \circ f \in \mathcal{R}_\alpha[a,b]$.

Proof. We may assume that α is increasing. Put $g = \varphi \circ f$ and let $\varepsilon > 0$. Uniform continuity of φ yields a $\delta > 0$ such that $|s-t| < \delta$ implies $|\varphi(s)-\varphi(t)| < \varepsilon$ for $s,t \in [m,M]$. Also choose δ such that $\delta \leq \varepsilon$. Let $P = \{x_0,\dots,x_n\}$ be a partition of $[a,b]$ such that

$$U_{f,\alpha}(P) - L_{f,\alpha}(P) < \delta^2.$$

Let A consist of those numbers $i \in \{1,\dots,n\}$ such that $M_i(f) - m_i(f) < \delta$, and let B consist of the remaining i. For $i \in A$ we then have $M_i(g) - m_i(g) \leq \varepsilon$.

Let $K > 0$ be such that $|\varphi| \leq K$. For $i \in B$ we then have $M_i(g) - m_i(g) \leq 2K$. Furthermore, we have

$$\sum_{i\in B} \Delta\alpha_i \leq \frac{1}{\delta}\sum_{i\in B}(M_i(f) - m_i(f))\Delta\alpha_i \leq \frac{1}{\delta}\bigl(U_{f,\alpha}(P) - L_{f,\alpha}(P)\bigr) < \delta.$$

It thus follows that

$$\begin{aligned} U_{g,\alpha}(P) - L_{g,\alpha}(P) &= \sum_{i\in A}(M_i(g) - m_i(g))\Delta\alpha_i + \sum_{i\in B}(M_i(g) - m_i(g))\Delta\alpha_i \\ &\leq \varepsilon(\alpha(b) - \alpha(a)) + 2K\delta \\ &\leq (\alpha(b) - \alpha(a) + 2K)\varepsilon. \end{aligned}$$

Since ε was arbitrary, it follows that $g \in \mathcal{R}_\alpha[a,b]$. ☐

For any function $f\colon X \to \mathbb{R}$, recall that the *positive and negative parts* of f are the functions $f^+ = f \vee 0$ and $f^- = -(f \wedge 0)$ respectively, and that $f = f^+ - f^-$. We notice that f^+ and f^- are both non-negative.

Corollary 3.21

Let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation. Then a function $f\colon [a,b] \to \mathbb{R}$ is α-integrable if and only if f^+ and f^- are, in which case

$$\int_a^b f\,\mathrm{d}\alpha = \int_a^b f^+\,\mathrm{d}\alpha - \int_a^b f^-\,\mathrm{d}\alpha.$$

Proof. This follows immediately from Proposition 3.20, since the maps $x \mapsto x \vee 0$ and $x \mapsto -(x \wedge 0)$ are continuous. ☐

Corollary 3.22

Let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $f,g \in \mathcal{R}_\alpha[a,b]$. Then the functions $|f|$ and fg are also α-integrable. In particular, $\mathcal{R}_\alpha[a,b]$ is an $\mathbb{R}$-algebra.

If α is increasing we also have

$$\left|\int_a^b f \,\mathrm{d}\alpha\right| \leq \int_a^b |f| \,\mathrm{d}\alpha. \tag{3.7}$$

Recall from Proposition 3.7 that $\mathcal{R}_\alpha[a,b]$ is always a vector space.

PROOF. Integrability of $|f|$ follows from Proposition 3.20 since $x \mapsto |x|$ is continuous. The inequality (3.7) follows since $f \leq |f|$, and since the α-integral is increasing when α is.

For the product fg, notice that

$$2fg = (f+g)^2 - f^2 - g^2,$$

and that the function $x \mapsto x^2$ is continuous. □

PROPOSITION 3.23

Let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation, let $f, g \in \mathcal{R}_\alpha[a,b]$, and define

$$F(x) = \int_a^x f \,\mathrm{d}\alpha$$

for $x \in [a,b]$. Then $g \in \mathcal{R}_F[a,b]$ and

$$\int_a^b fg \,\mathrm{d}\alpha = \int_a^b g \,\mathrm{d}F.$$

REMARK 3.24. It suffices to assume that fg is integrable instead of f. Conversely, notice that $fg \in \mathcal{R}_\alpha[a,b]$ by Corollary 3.22 under the assumptions in Proposition 3.23.

If f is continuous and α is continuously differentiable, then this result follows from the first fundamental theorem of calculus, which we prove in Theorem 3.25(iii). For in this case F is continuously differentiable with $F'(x) = f(x)\alpha'(x)$ (as we show below, we may assume that α is increasing), so applying Proposition 3.11 twice yields

$$\int_a^b g \,\mathrm{d}F = \int_a^b gF' = \int_a^b fg\alpha' = \int_a^b fg \,\mathrm{d}\alpha.$$

However, the claim holds more generally, as we now show. ⌟

PROOF OF PROPOSITION 3.23. We may assume that α is increasing: Define V as in Theorem 3.17, and let

$$F_1(x) = \int_a^x f \,\mathrm{d}V, \quad \text{and} \quad F_2(x) = \int_a^x f \,\mathrm{d}(V - \alpha),$$

so that $F = F_1 - F_2$. Then assuming that the proposition holds for increasing integrators, we have

$$\begin{aligned}\int_a^b fg\,\mathrm{d}\alpha &= \int_a^b fg\,\mathrm{d}V - \int_a^b fg\,\mathrm{d}(V-\alpha) = \int_a^b g\,\mathrm{d}F_1 - \int_a^b g\,\mathrm{d}F_2 \\ &= \int_a^b g\,\mathrm{d}(F_1 - F_2) = \int_a^b g\,\mathrm{d}F.\end{aligned}$$

Hence we assume below that α is increasing.

Let (P,T) be a tagged partition of $[a,b]$ with $P = \{x_0,\dots,x_n\}$ and $T = \{t_1,\dots,t_n\}$. Then

$$S_{g,F}(P,T) = \sum_{i=1}^n g(t_i)\int_{x_{i-1}}^{x_i} f(t)\,\mathrm{d}\alpha(t) = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(t)g(t_i)\,\mathrm{d}\alpha(t),$$

and

$$\int_a^b fg\,\mathrm{d}\alpha = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(t)g(t)\,\mathrm{d}\alpha(t).$$

Letting $M = \sup_{x\in[a,b]}|f(x)|$ we thus have

$$\begin{aligned}\left|S_{g,F}(P,T) - \int_a^b fg\,\mathrm{d}\alpha\right| &= \left|\sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(t)\big(g(t_i) - g(t)\big)\,\mathrm{d}\alpha(t)\right| \\ &\leq M\sum_{i=1}^n \int_{x_{i-1}}^{x_i} |g(t_i) - g(t)|\,\mathrm{d}\alpha(t) \\ &\leq M\sum_{i=1}^n \int_{x_{i-1}}^{x_i} \big(M_i(g) - m_i(g)\big)\,\mathrm{d}\alpha \\ &= M\big(U_{g,\alpha}(P) - L_{g,\alpha}(P)\big).\end{aligned}$$

Since g is α-integrable, this can be made as small as desired. This proves the claim. □

3.6. *The fundamental theorems of calculus*

THEOREM 3.25: *The integral as a function of the interval*

Let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation, and let $f \in \mathcal{R}_\alpha[a,b]$. Define a function $F\colon [a,b] \to \mathbb{R}$ by

$$F(x) = \int_a^x f\,\mathrm{d}\alpha. \tag{3.8}$$

Then the following hold:

(i) *F is of bounded variation.*

(ii) *Every point of continuity of α is also a point of continuity of F.*

(iii) *Assume that α is increasing. If f is continuous and α differentiable at $x \in [a,b]$, then F is differentiable at x with $F'(x) = f(x)\alpha'(x)$.*

The result Theorem 3.25(iii) is known as *the first fundamental theorem of calculus*. Notice that the integral in (3.8) exists by Corollary 3.18, and so do the integrals in the proof below.

PROOF. *Proof of (i)*: Since $BV[a,b]$ is a vector space, we may assume that α is increasing. By Corollary 3.21 we may further assume that f is positive. Then F is increasing, hence of bounded variation.[6]

Proof of (ii): By Proposition 3.4 we may assume that α is increasing. Let $x,y \in [a,b]$ with $x \neq y$, and let I denote the closed interval between x and y. Proposition 3.15 now furnishes a $c_{xy} \in \mathbb{R}$ with

$$\inf_{t\in[a,b]} f(t) \leq \inf_{t\in I} f(t) \leq c_{xy} \leq \sup_{t\in I} f(t) \leq \sup_{t\in[a,b]} f(t),$$

such that

$$F(y) - F(x) = \int_x^y f \,\mathrm{d}\alpha = c_{xy}(\alpha(y) - \alpha(x)). \tag{3.9}$$

Assume now that α is continuous at x. Since $y \mapsto c_{xy}$ is bounded on $[a,b]$ (since f is), letting $y \to x$ thus yields continuity of F at x.

Proof of (iii): If f is continuous at x, then $c_{xy} \to f(x)$ as $y \to x$, since c_{xy} is bounded by the supremum and infimum of f in a neighbourhood of x. Now (3.9) implies that

$$\frac{F(y) - F(x)}{y - x} = c_{xy}\frac{\alpha(y) - \alpha(x)}{y - x} \xrightarrow[y\to x]{} f(x)\alpha'(x),$$

as desired. ☐

REMARK 3.26. In the case $\alpha(x) = x$, (iii) has a proof that does not use Proposition 3.15: Simply note that

$$\frac{F(y) - F(x)}{y - x} - f(x) = \frac{1}{y - x}\int_x^y (f(t) - f(x))\,\mathrm{d}t,$$

and notice that the integrand can be made less than any $\varepsilon > 0$ if $|t - x| < \delta$ for an appropriate $\delta > 0$. I am not sure that this proof can be generalised. ⌟

[6] E.g. Apostol (1974, Theorem 7.32) attempts to prove this using Proposition 3.15, but his argument is not, as far as I can tell, correct as stated. Hence we use a different argument, more in the spirit of Lebesgue.

THEOREM 3.27: *The second fundamental theorem of calculus*

Let $f \in \mathcal{R}[a,b]$. If there exists a continuous function $F\colon [a,b] \to \mathbb{R}$ that is differentiable on (a,b) with $F' = f$, then

$$\int_a^b f = F(b) - F(a).$$

PROOF. Let $P = \{x_0, \dots, x_n\}$ be a partition of $[a,b]$. The mean value theorem furnishes points $t_i \in (x_{i-1}, x_i)$ such that $\Delta F_i = F'(t_i)\Delta x_i = f(t_i)\Delta x_i$. It follows that

$$\left| F(b) - F(a) - \int_a^b f \right| = \left| \sum_{i=1}^n f(t_i)\Delta x_i - \int_a^b f \right| < \varepsilon$$

if P is fine enough. Since ε was arbitrary, this proves the theorem. □

3.7. *Further results on Riemann integrals*

COROLLARY 3.28: *The second mean value theorem for Riemann integrals*

Let $f, g\colon [a,b] \to \mathbb{R}$ with f monotonic and $g \in \mathcal{R}[a,b]$. Then there exists a point $x_0 \in [a,b]$ such that

$$\int_a^b fg = f(a) \int_a^{x_0} g + f(b) \int_{x_0}^b g.$$

Both Apostol (1974, Theorem 7.37) and Folland (2007, Lemma 8.41) assume that g is continuous (the latter even assumes that f is right-continuous), but as far as I can tell the argument below goes through without this assumption.

PROOF. Define $\alpha(x) = \int_a^x g$. Then α is continuous by Theorem 3.25(ii), so Proposition 3.16 yields an $x_0 \in [a,b]$ such that

$$\int_a^b f \, \mathrm{d}\alpha = f(a) \int_a^{x_0} \mathrm{d}\alpha + f(b) \int_{x_0}^b \mathrm{d}\alpha.$$

Furthermore, since both f and g is Riemann-integrable (f since it is monotonic), Proposition 3.23 implies that

$$\begin{aligned} \int_a^b fg = \int_a^b f \, \mathrm{d}\alpha &= f(a) \int_a^{x_0} \mathrm{d}\alpha + f(b) \int_{x_0}^b \mathrm{d}\alpha \\ &= f(a) \int_a^{x_0} g + f(b) \int_{x_0}^b g, \end{aligned}$$

as desired. □

[TODO change of variables]

3.8. Limit and continuity theorems

Proposition 3.29

Let $f\colon [a,b]\times[c,d] \to \mathbb{R}$ be continuous, and let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation. Then the function $F\colon [c,d] \to \mathbb{R}$ given by

$$F(y) = \int_a^b f(x,y)\,\mathrm{d}\alpha(x)$$

is continuous.

Proof. We may assume that α is increasing. By uniform continuity of f, given $\varepsilon > 0$ there is a $\delta > 0$ such that $\|z-z'\| < \delta$ implies $|f(z)-f(z')| < \varepsilon$ for $z,z' \in [a,b]\times[c,d]$. Given $y,y' \in [c,d]$ with $|y-y'| < \delta$ we thus have

$$|F(y)-F(y')| \leq \int_a^b |f(x,y)-f(x,y')|\,\mathrm{d}\alpha(x) \leq \varepsilon(\alpha(b)-\alpha(a)).$$

Since ε was arbitrary, this shows that F is continuous. $\square$

Proposition 3.30

Let $f\colon [a,b]\times[c,d] \to \mathbb{R}$ be bounded, and let $\alpha\colon [a,b] \to \mathbb{R}$ be of bounded variation. Assume that $f(\cdot,y) \in \mathcal{R}_\alpha[a,b]$ for all $y \in [c,d]$, that $f(x,\cdot)$ is continuous on $[c,d]$ and differentiable on (c,d) for all $x \in [a,b]$, and that $D_2 f$ is continuous on $[a,b]\times(c,d)$. Then the function $F\colon [c,d] \to \mathbb{R}$ given by

$$F(y) = \int_a^b f(x,y)\,\mathrm{d}\alpha(x)$$

is differentiable on (c,d) and

$$F'(y) = \int_a^b D_2 f(x,y)\,\mathrm{d}\alpha(x).$$

Proof. We may assume that α is increasing. Let $y,y_0 \in (c,d)$ with $y \neq y_0$. By the mean value theorem we have

$$\frac{F(y)-F(y_0)}{y-y_0} = \int_a^b \frac{f(x,y)-f(x,y_0)}{y-y_0}\,\mathrm{d}\alpha(x) = \int_a^b D_2 f(x,y_x)\,\mathrm{d}\alpha(x)$$

for some $y_x \in (c,d)$ lying between y and y_0, depending on x. Let $I \subseteq (c,d)$ be a non-trivial compact interval containing y. Then $D_2 f$ is uniformly continuous on $[a,b]\times I$, so given $\varepsilon > 0$ there is a $\delta > 0$ such that $\|z-z'\| < \delta$ implies

$|D_2 f(z) - D_2 f(z')| < \varepsilon$ for $z, z' \in [a,b] \times I$. For $y, y_0 \in I$ with $|y - y_0| < \delta$ we also have $|y_x - y_0| < \delta$ for all $x \in [a,b]$, and so

$$\begin{aligned}\left|\int_a^b D_2 f(x, y_x)\,\mathrm{d}\alpha(x) - \int_a^b D_2 f(x, y_0)\,\mathrm{d}\alpha(x)\right| &\leq \int_a^b |D_2 f(x, y_x) - D_2 f(x, y_0)|\,\mathrm{d}\alpha(x) \\ &\leq \varepsilon(\alpha(b) - \alpha(a)).\end{aligned}$$

Since ε was arbitrary, this shows that F is differentiable at y_0 with derivative as claimed. □

Proposition 3.31

Let α be of bounded variation on $[a,b]$, and let $(f_n)_{n \in \mathbb{N}}$ be a sequence of α-integrable functions on $[a,b]$ that converge uniformly to a function f. Then f is also α-integrable on $[a,b]$, and

$$\int_a^b f\,\mathrm{d}\alpha = \lim_{n \to \infty} \int_a^b f_n\,\mathrm{d}\alpha.$$

In particular, $\mathcal{R}_\alpha[a,b]$ is a closed subspace of $C[a,b]$ equipped with the uniform norm.

[TODO α must be increasing! Or at least BV.]

Proof. We may assume that α is increasing. Let $\varepsilon_n = \|f_n - f\|_\infty$ such that

$$f_n - \varepsilon_n \leq f \leq f_n + \varepsilon_n$$

for $n \in \mathbb{N}$. It follows that

$$\int_a^b (f_n - \varepsilon_n)\,\mathrm{d}\alpha \leq \underline{\int_a^b} f\,\mathrm{d}\alpha \leq \overline{\int_a^b} f\,\mathrm{d}\alpha \leq \int_a^b (f_n + \varepsilon_n)\,\mathrm{d}\alpha,$$

and hence,

$$0 \leq \overline{\int_a^b} f\,\mathrm{d}\alpha - \underline{\int_a^b} f\,\mathrm{d}\alpha \leq 2\varepsilon_n(\alpha(b) - \alpha(a)).$$

Thus the upper and lower integrals of f are equal, so f is α-integrable. Finally we have

$$\left|\int_a^b f_n\,\mathrm{d}\alpha - \int_a^b f\,\mathrm{d}\alpha\right| \leq \int_a^b |f_n - f|\,\mathrm{d}\alpha \leq \varepsilon_n(\alpha(b) - \alpha(a)),$$

proving the claim. □

3.9. Line integrals

Recall that a *path* in a topological space X is a continuous map $\gamma\colon [a,b] \to X$, and that γ is *closed* if $\gamma(a) = \gamma(b)$. A subset $\Gamma \subseteq X$ is called a *curve* in X if there is a path α in X whose image is Γ. The image of a path γ is called its *trace* and is denoted γ^*.

DEFINITION 3.32: *Equivalence of paths*

Let $\alpha\colon [a,b] \to X$ and $\beta\colon [c,d] \to X$ be paths in a topological space X. If there is an increasing homeomorphism $\varphi\colon [c,d] \to [a,b]$ such that $\beta = \alpha \circ \varphi$, then α and β are said to be *properly equivalent*.

If α and β are closed paths with $\alpha(a) \neq \beta(c)$, then we also say that they are properly equivalent if there is a point $e \in (c,d)$ such that α and γ are properly equivalent in the above sense, where $\gamma\colon [e, d-c+e] \to X$ is given by

$$\gamma(t) = \begin{cases} \beta(t), & t \in [e,d], \\ \beta(t-d+c), & t \in [d, d-c+e]. \end{cases}$$

If the map φ above is decreasing, then we say that α and β are *improperly equivalent*. The paths α and β are *equivalent* if they are either properly or improperly equivalent.

Note that the condition that φ be an increasing (decreasing) homeomorphism is equivalent to it being continuous, strictly increasing (decreasing) and surjective. Also note that equivalent paths trace out the same curve in X.

DEFINITION 3.33: *Line integrals*

Let $\gamma\colon [a,b] \to \mathbb{R}^d$ be a path, and let $f\colon \gamma^* \to \mathbb{R}^d$ be a vector field. Given a tagged partition (P,T) of $[a,b]$ then, with notation as above, we form the sums

$$S_{f,\gamma}(P,T) = \sum_{i=1}^{n} f(\gamma(t_i)) \cdot (\gamma(x_i) - \gamma(x_{i-1})).$$

Define the *line integral* of f with respect to γ as the limit of the net $S_{f,\gamma}$, if the limit exists. We denote this integral by $\int f \cdot \mathrm{d}\gamma$.

Notice that if α and β are properly equivalent paths, then

$$\int f \cdot \mathrm{d}\alpha = \int f \cdot \mathrm{d}\beta.$$

If α and β are instead improperly equivalent, then the two integrals are equal but with opposite signs.

PROPOSITION 3.34

Let $\gamma\colon [a,b] \to \mathbb{R}^d$ be a path, and let $f\colon \gamma^* \to \mathbb{R}^d$ be a bounded function. Then

$$\int f \cdot \mathrm{d}\gamma = \sum_{k=1}^{d} \int_a^b f_k \circ \gamma \,\mathrm{d}\gamma_k$$

whenever each Riemann–Stieltjes integral on the right exists. If in addition γ is piecewise C^1, then

$$\int f \cdot \mathrm{d}\gamma = \int_a^b f(\gamma(t)) \cdot \gamma'(t) \,\mathrm{d}t.$$

PROOF. Notice that

$$S_{f,\gamma}(P,T) = \sum_{k=1}^{d} \sum_{i=1}^{n} f_k(\gamma(t_i))(\gamma_k(t_i) - \gamma_k(t_{i-1})).$$

Since the inner sums on the right-hand side approximate the Riemann–Stieltjes integrals $\int_a^b f_k \circ \gamma \,\mathrm{d}\gamma_k$, the first claim follows by taking limits. The second claim follows by [reference]. □

THEOREM 3.35: *Integral of a gradient*

Let $U \subseteq \mathbb{R}^d$ be open, and let $\varphi \in C^1(U)$. For every pair of points $x, y \in U$ and every piecewise C^1 path $\gamma\colon [a,b] \to U$ with $\gamma(a) = x$ and $\gamma(b) = y$ we have

$$\int \nabla\varphi \cdot \mathrm{d}\gamma = \varphi(y) - \varphi(x).$$

If $f = \nabla\varphi$, then φ is called a *potential function* for f.

PROOF. Let $a = t_0 < \cdots < t_n = b$ be a partition of $[a,b]$ such that γ' is continuous on each subinterval. By the chain rule,

$$(\varphi \circ \gamma)'(t) = \nabla\varphi(\gamma(t)) \cdot \gamma'(t)$$

on each open subinterval (t_{i-1}, t_i). By [reference],

$$\begin{aligned}\int \nabla\varphi \cdot \mathrm{d}\gamma &= \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \nabla\varphi(\gamma(t)) \cdot \gamma'(t) \,\mathrm{d}t = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (\varphi \circ \gamma)'(t) \,\mathrm{d}t \\ &= \varphi(\gamma(b)) - \varphi(\gamma(a)) = \varphi(y) - \varphi(x),\end{aligned}$$

as desired. □

Theorem 3.36

Let $U \subseteq \mathbb{R}^d$ be an open, connected set, and let $f : U \to \mathbb{R}^d$ be a continuous function. Fix a point $x_0 \in U$. For each $x \in U$ and each pair of polygonal paths $\alpha, \beta : [a,b] \to U$ joining x_0 and x, assume that

$$\int f \cdot \mathrm{d}\alpha = \int f \cdot \mathrm{d}\beta.$$

Then there exists a function $\varphi \in C^1(U)$ such that $f = \nabla\varphi$.

Notice that since U is connected, such polygonal paths exist between any pair of points.

Proof. Let $x \in U$, and let $\alpha : [a,b] \to U$ be a polygonal curve joining x_0 and x. Define

$$\varphi(x) = \int f \cdot \mathrm{d}\alpha.$$

By hypothesis, the number $\varphi(x)$ does not depend on the particular choice of α. We show that each partial derivative $D_k\varphi(x)$ exists and equals $f_k(x)$.

Let $B(x,\delta) \subseteq U$ for some $\delta > 0$, and let $\lambda \in [-\delta/2, \delta/2]$. Define a path $\gamma : [0,1] \to B(x,\delta)$ by $\gamma(t) = (1-t)x + t(x + \lambda e_k)$, where e_k is the kth standard basis vector. Then

$$\varphi(x + \lambda e_k) - \varphi(x) = \int f \cdot \mathrm{d}\gamma.$$

Furthermore, $\gamma_k'(t) = \lambda$ and $\gamma_i'(t) = 0$ for $i \neq k$. Thus γ is C^1, and so

$$\begin{aligned}\varphi(x + \lambda e_k) - \varphi(x) &= \sum_{i=1}^{d} \int_0^1 f_i(\gamma(t))\gamma_i'(t)\,\mathrm{d}t \\ &= \lambda \int_0^1 f_k(\gamma(t))\,\mathrm{d}t = \lambda \int_0^1 g(t,\lambda)\,\mathrm{d}t,\end{aligned}$$

where $g(t,\lambda) = f_k((1-t)x + t(x + \lambda e_k))$. Since g is continuous on $[0,1] \times [-\delta/2, \delta/2]$, Proposition 3.29 implies that

$$\lim_{\lambda \to 0} \int_0^1 g(t,\lambda)\,\mathrm{d}t = \int_0^1 g(t,0)\,\mathrm{d}t = \int_0^1 f_k(x)\,\mathrm{d}t = f_k(x),$$

proving that $D_k\varphi(x) = f_k(x)$. Thus $\nabla\varphi(x) = f(x)$ for all $x \in U$, and $\varphi \in C^1(U)$ since f is continuous. □

4 • Convergence

It is well-known that a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space S converges to some $x \in S$ if and only if (x_n) has a *subsequence* that converges to x.

In this section we highlight a similar feature of convergence in measure and convergence in mean: If $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in measure, and if there is a subsequence that converges *pointwise a.e.* to some function f, then (f_n) also converges to f in measure. Similarly for convergence in mean.

Furthermore, Markov's inequality implies that convergence in mean is stronger than convergence in measure. In particular, a sequence that is Cauchy in mean is also Cauchy in measure. Hence when we show that convergence in measure and in mean are complete, it suffices to show that being Cauchy in measure implies the existence of a pointwise a.e. convergent subsequence.

DEFINITION 4.1: *Convergence in measure*

Let $(X,\mathcal{E},\mu)$ be a measure space. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$, and let further $f\in\mathcal{M}(\mathcal{E})$. We say that (f_n) *converges to f in μ-measure* if for every $\varepsilon>0$,

$$\lim_{n\to\infty}\mu\Big(\{|f_n-f|>\varepsilon\}\Big)=0.$$

Furthermore, (f_n) is called a *Cauchy sequence in μ-measure* if, for every $\varepsilon>0$,

$$\lim_{m,n\to\infty}\mu\Big(\{|f_m-f_n|>\varepsilon\}\Big)=0.$$

We prove that convergence in μ-measure is complete.

LEMMA 4.2

Let $(X,\mathcal{E},\mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$. If there exists a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ of strictly positive numbers such that

$$\sum_{n=1}^{\infty}\varepsilon_n<\infty \quad\textit{and}\quad \sum_{n=1}^{\infty}\mu\Big(\{|f_{n+1}-f_n|>\varepsilon_n\}\Big)<\infty, \tag{4.1}$$

then there exists a function $f\in\mathcal{M}(\mathcal{E})$ such that (f_n) converges to f both μ-a.e. and in μ-measure.

PROOF. For $n\in\mathbb{N}$, denote the set $\{|f_{n+1}-f_n|>\varepsilon_n\}$ by E_n, and define sets $F_k=\bigcup_{n=k}^{\infty}E_n$ and $F=\bigcap_{k\in\mathbb{N}}F_k$. Notice that $F=\limsup_{n\to\infty}E_n$, so the first Borel–Cantelli lemma implies that $\mu(F)=0$.

For $m\geq n$ we find that

$$|f_m-f_n|\leq\sum_{i=n}^{m-1}|f_{i+1}-f_i|\leq\sum_{i=n}^{\infty}|f_{i+1}-f_i|. \tag{4.2}$$

If furthermore $x\in F_k^c$ and $m\geq n\geq k$, then

$$|f_m(x)-f_n(x)|\leq\sum_{i=n}^{\infty}\varepsilon_i.$$

The right-hand side converges to zero as $n \to \infty$, which shows that $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$ for $x \in F_k^c$, hence for $x \in F^c$. Letting $f = \lim_{n\to\infty} f_n \mathbf{1}_{F^c}$ we thus find that (f_n) converges to $f \in \mathcal{M}(\mathcal{E})$ μ-a.e.

Next we show that $f_n \to f$ in μ-measure as $n \to \infty$. Letting $m \to \infty$ in (4.2) we find that

$$|f_n - f| \leq \sum_{i=n}^{\infty} |f_{i+1} - f_i|$$

μ-a.e. Now let $\varepsilon > 0$, and choose an $N \in \mathbb{N}$ such that $\sum_{i=N}^{\infty} \varepsilon_i \leq \varepsilon$. For $n \geq N$, $|f_n - f| > \varepsilon$ then implies that

$$\sum_{i=n}^{\infty} \varepsilon_i \leq \varepsilon < |f_n - f| \leq \sum_{i=n}^{\infty} |f_{i+1} - f_i|$$

μ-a.e., which in turn implies that $\varepsilon_i < |f_{i+1} - f_i|$ μ-a.e. for some $i \geq n$. Hence it follows that

$$\mu\Big(\{|f_n - f| > \varepsilon\}\Big) \leq \mu\Big(\bigcup_{i=n}^{\infty} E_i\Big) \leq \sum_{i=n}^{\infty} \mu(E_i),$$

which converges to zero by (4.1). □

Theorem 4.3: *Completeness of convergence in measure*

Let $(X, \mathcal{E}, \mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$ that is Cauchy in μ-measure. Then there exists a function $f \in \mathcal{M}(\mathcal{E})$ such that $f_n \to f$ in μ-measure. Furthermore, (f_n) has a subsequence that converges to f μ-a.e.

Proof. We prove the following lemma:

> *Let $(X, \mathcal{E}, \mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$ that is Cauchy in μ-measure. If (f_n) has a subsequence that converges μ-a.e. to function $f \in \mathcal{M}(\mathcal{E})$, then (f_n) also converges to f in μ-measure.*

Let (f_{n_k}) be such a subsequence. For $\varepsilon > 0$ we then have

$$\{|f_n - f| > \varepsilon\} \subseteq \{|f_n - f_{n_k}| > \varepsilon/2\} \cup \{|f_{n_k} - f| > \varepsilon/2\},$$

and the measures of the sets on the right-hand side go to zero as $n \to \infty$ (since $n_k \geq n$). This proves the lemma.

To prove the theorem, choose a subsequence $(g_k) = (f_{n_k})$ such that

$$\mu\Big(\{|g_{k+1} - g_k| > 2^{-k}\}\Big) \leq 2^{-k}.$$

Lemma 4.2 then implies the existence of a function $f \in \mathcal{M}(\mathcal{E})$ such that $g_k \to f$ both μ-a.e. and in μ-measure. The claim then follows from the above lemma. □

Definition 4.4: *Convergence in mean*

Let $(X, \mathcal{E}, \mu)$ be a measure space. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$, and let further $f \in \mathcal{M}(\mathcal{E})$. If $p \in (0,\infty)$ we say that (f_n) *converges to f in the μ-p-th mean* if

$$\lim_{n\to\infty} \int_X |f_n - f|^p \,\mathrm{d}\mu = 0.$$

Furthermore, (f_n) is called a *Cauchy sequence in the μ-p-th mean* if

$$\lim_{m,n\to\infty} \int_X |f_m - f_n|^p \,\mathrm{d}\mu = 0.$$

Remark 4.5. If (f_n) converges to f in the μ-p-th mean, then $f_n - f \in \mathcal{L}^p(\mu)$ for $n \geq N$ for some $N \in \mathbb{N}$. Furthermore, if $f \in \mathcal{L}^p(\mu)$, then $f_n = (f_n - f) + f \in \mathcal{L}^p(\mu)$ for $n \geq N$.

On the other hand, if $f_n \in \mathcal{L}^p(\mu)$ for large enough n, then $f = (f - f_n) + f_n \in \mathcal{L}^p(\mu)$. In particular, $\mathcal{L}^p(\mu)$ is a closed subspace of $\mathcal{M}(\mathcal{E})$. ⌟

Theorem 4.6: *Completeness of convergence in mean*

Let $(X, \mathcal{E}, \mu)$ be a measure space, let $p \in (0,\infty)$, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$ that is Cauchy in the μ-p-th mean. Then there exists a function $f \in \mathcal{M}(\mathcal{E})$ such that $f_n \to f$ in the μ-p-th mean. Furthermore, (f_n) has a subsequence that converges to f μ-a.e.

In particular, $\mathcal{L}^p(\mu)$ is complete.

Proof. We prove the following lemma:

> *Let $(X, \mathcal{E}, \mu)$ be a measure space, let $p \in (0,1)$, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mathcal{E})$ that is Cauchy in the μ-p-th mean. If (f_n) has a subsequence that converges μ-a.e. to function $f \in \mathcal{M}(\mathcal{E})$, then (f_n) also converges to f in the μ-p-th mean.*

Let $(f_{n_k})_{k\in\mathbb{N}}$ be a subsequence that converges μ-a.e. to f. For $n \in \mathbb{N}$, Fatou's lemma implies that

$$\begin{aligned}\int_X |f_n - f|^p \,\mathrm{d}\mu &= \int_X \liminf_{k\to\infty} |f_n - f_{n_k}|^p \,\mathrm{d}\mu \leq \liminf_{k\to\infty} \int_X |f_n - f_{n_k}|^p \,\mathrm{d}\mu \\ &\leq \sup_{m\geq n} \int_X |f_n - f_m|^p \,\mathrm{d}\mu,\end{aligned}$$

which converges to zero as $n \to \infty$ as desired.

We now prove the theorem. Markov's inequality implies that (f_n) is also a Cauchy sequence in μ-measure, so [reference] yields a function $f \in \mathcal{M}(\mathcal{E})$ such that $f_n \to f$ in μ-measure, and such that (f_n) has a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ that converges to f μ-a.e. The lemma then implies that (f_n) converges to f in the μ-p-th mean.

For the last claim, if $f_n \in \mathcal{L}^p(\mu)$ for all $n \in \mathbb{N}$, then since $\mathcal{L}^p(\mu)$ is closed we also have $f \in \mathcal{L}^p(\mu)$. □

5 • Portmanteau theorems

If P is a Borel probability measure on a topological space X, then a subset $A \subseteq X$ is called a *P-continuity set* if $P(\partial A) = 0$.

Recall that a topological space X is called a *perfect space* or a *G_δ space* if every closed subset of X is a G_δ set, i.e. a countable intersection of open sets. (Equivalently, if every open subset is an F_σ set, a countable union of closed sets.) Furthermore, if A and B are closed subsets of X, a *Urysohn function* for the pair (A,B) is a continuous function $f\colon X \to [0,1]$ with $f(A) = 1$ and $f(B) = 0$.[7] By Urysohn's lemma, a topological space is normal (i.e. every pair of disjoint closed sets can be separated by disjoint open sets) if and only if there is a Urysohn function for every pair of disjoint closed subsets. Finally, a (not necessarily Hausdorff) normal G_δ space is called a *perfectly normal space*. If it is also Hausdorff it also called a *T_6-space* or a *perfectly T_4 space*.

It is easy to show that (pseudo-)metrisable spaces are perfectly normal. Another notable class of perfectly normal (Hausdorff) spaces are the CW complexes. [TODO: Reference, proof somewhere else maybe?]

Theorem 5.1

Let $(P_n)_{n\in\mathbb{N}}$ and P be probability measures on a perfectly normal space X. Then the following conditions are equivalent:

(i) $P_n \Rightarrow P$.

(ii) $\limsup_{n\to\infty} P_n(F) \leq P(F)$ *for all closed* $F \subseteq X$.

(iii) $\liminf_{n\to\infty} P_n(G) \geq P(G)$ *for all open* $G \subseteq X$.

(iv) $P_n(A) \to P(A)$ *for all P-continuity sets* $A \subseteq X$.

Proof. *(i) ⇒ (ii):* Let F be a closed subset of X. Since X is perfect, there is a decreasing sequence $(F_k)_{k\in\mathbb{N}}$ of open sets such that $F = \bigcap_{k\in\mathbb{N}} F_k$. Furthermore, since X is normal there is for each $k \in \mathbb{N}$ a Urysohn function g_k for the pair

[7] The ordering of A and B is only relevant insofar as it is relevant on which set the Urysohn function in question vanishes. If $f\colon X \to [0,1]$ is a Urysohn function for the pair (A,B), then the function $1-f$ is a Urysohn function for the pair (B,A).

(F, F_k^c). We clearly have $\mathbf{1}_F \leq g_k \leq \mathbf{1}_{F_k}$ so

$$\limsup_{n\to\infty} P_n(F) \leq \limsup_{n\to\infty} \int_X g_k \,\mathrm{d}P_n = \int_X g_k \,\mathrm{d}P \leq P(F_k).$$

Since $F_k \downarrow F$ as $k \to \infty$, continuity of P implies the claim.

(ii) ⇔ (iii): This follows easily by taking complements.

(ii) & (iii) ⇒ (iv): For $A \subseteq X$ we have

$$\begin{aligned} P(A^\circ) &\leq \liminf_{n\to\infty} P_n(A^\circ) \leq \liminf_{n\to\infty} P_n(A) \\ &\leq \limsup_{n\to\infty} P_n(A) \leq \limsup_{n\to\infty} P_n(\overline{A}) \leq P(\overline{A}). \end{aligned}$$

If A is a P-continuity set then $P(A^\circ) = P(\overline{A})$, which implies (iv).

(iv) ⇒ (i): Given $f \in C_b(X)$, by linearity we may assume that $0 \leq f \leq 1$. Then

$$\int_X f \,\mathrm{d}P = \int_0^\infty P(f \geq t)\,\mathrm{d}t = \int_0^1 P(f \geq t)\,\mathrm{d}t,$$

and similarly for P_n. Since f is continuous, we have $\partial\{f \geq t\} \subseteq \{f = t\}$. Now notice that $\{f = t\}$ is a P-null set except for countably many $t \in (0,1)$, since P is a finite measure.[8] Hence $\{f \geq t\}$ is a P-continuity set for all but countably many t. It then follows from (iv) and the dominated convergence theorem that

$$\int_X f \,\mathrm{d}P_n = \int_0^1 P_n(f \geq t)\,\mathrm{d}t \xrightarrow[n\to\infty]{} \int_0^1 P(f \geq t)\,\mathrm{d}t = \int_X f \,\mathrm{d}P$$

as claimed. □

REMARK 5.2. In the case that X is pseudometrisable, we may pick a pseudometric ρ. The sets F_k can then be given explicitly by

$$F_k = \left\{ x \in S \;\middle|\; \rho(x, F) < \frac{1}{k} \right\}.$$

Furthermore, we may take the Urysohn functions g_k to be given by $g_k(x) = (1 - k\rho(x,F)) \vee 0$. Notice that these are Lipschitz since

$$\begin{aligned} |g_k(x) - g_k(y)| &= \big|(1 - k\rho(x,F)) \vee 0 - (1 - k\rho(y,F)) \vee 0\big| \\ &\leq |k\rho(x,F) - k\rho(y,F)| \leq k\rho(x,y) \end{aligned}$$

for all $x, y \in X$. In particular they are uniformly continuous. Hence weak convergence in a metrisable space may be characterised by the bounded, uniformly continuous functions, or even the bounded Lipschitz functions. Notice also that this is independent of the metric. ⌟

[8] Indeed, f is a random variable whose discrete support is precisely this set of ts. But the discrete support of any random variable is countable.

6 • Dynkin systems and monotone classes

DEFINITION 6.1: *Dynkin systems, π-systems*

Let X be a set. A collection $\mathcal{D}$ of subsets of X is a *Dynkin system* in X if it has the following properties:

(i) $X \in \mathcal{D}$,

(ii) $B \setminus A \in \mathcal{D}$ whenever $A, B \in \mathcal{D}$ and $A \subseteq B$, and

(iii) $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$ for any increasing sequence $(A_n)_{n\in\mathbb{N}}$ of sets in $\mathcal{D}$.

Furthermore, a collection $\mathcal{S}$ of subsets of X is called a *π-system* if it is closed under intersections.

It is easy to show that if $\mathcal{D}$ is both a Dynkin system in X and a π-system, then it is in fact a σ-algebra in X.

DEFINITION 6.2: *Monotone classes, set algebras*

Let X be a set. A collection $\mathcal{M}$ of subsets of X is a *monotone class* if it is closed under countable increasing unions and countable decreasing intersections.

Furthermore, a collection $\mathcal{A}$ of subsets of X is called a *set algebra* in X if it is closed under finite unions and complements.

We note that a set algebra $\mathcal{A}$ in X is automatically closed under finite intersections, and that it also contains both $\emptyset$ and X. It is easy to show that if $\mathcal{M}$ is both a monotone class and a set algebra in X, then it is in fact a σ-algebra in X.

Notice that we have two pairs of properties that ensure that a collection of sets is a σ-algebra. On the one hand we should think of Dynkin systems and monotone classes as being analogous, and similarly for π-systems and set algebras on the other. The latter pair of properties are algebraic, while the first pair are somehow analytic (or continuous), in that they involve infinitely many operations.

It turns out that if $\mathcal{S}$ is a π-system, then the Dynkin system $\delta(\mathcal{S})$ generated by $\mathcal{S}$ is also a π-system. Similarly, if $\mathcal{A}$ is a set algebra, then the monotone class $M(\mathcal{A})$ generated by $\mathcal{A}$ is also a set algebra. We have the following two results whose proofs are basically identical:

THEOREM 6.3: *Dynkin's lemma*

Let $\mathcal{S}$ be a π-system in a set X. Then $\delta(\mathcal{S})$ is also a π-system, and in particular

$$\delta(\mathcal{S}) = \sigma(\mathcal{S}). \tag{6.1}$$

PROOF. The identity (6.1) follows if $\delta(\mathcal{S})$ is a π-system: For then it is also a σ-algebra, and then $\sigma(\mathcal{S}) \subseteq \delta(\mathcal{S})$.

For $A \in \delta(\mathcal{S})$ define

$$\mathcal{D}_A = \{B \in \delta(\mathcal{S}) \mid A \cap B \in \delta(\mathcal{S})\}.$$

This is easily seen to be a Dynkin system. Also notice that $B \in \mathcal{D}_A$ if and only if $A \in \mathcal{D}_B$ for all $A, B \in \delta(\mathcal{S})$. Furthermore, if $A \in \mathcal{S}$ then $\mathcal{S} \subseteq \mathcal{D}_A$, and so $\delta(\mathcal{S}) \subseteq \mathcal{D}_A$. In other words,

$$B \in \mathcal{D}_A \quad \text{for } A \in \mathcal{S} \text{ and } B \in \delta(\mathcal{S}).$$

By symmetry we then have

$$A \in \mathcal{D}_B \quad \text{for } A \in \mathcal{S} \text{ and } B \in \delta(\mathcal{S}),$$

and since $\mathcal{D}_B$ is a Dynkin system it follows that $\delta(\mathcal{S}) \subseteq \mathcal{D}_B$. Hence $\delta(\mathcal{S})$ is a π-system as desired. □

THEOREM 6.4: *The monotone class lemma*

Let $\mathcal{A}$ be a set algebra in a set X. Then $M(\mathcal{A})$ is also a set algebra, and in particular

$$M(\mathcal{A}) = \sigma(\mathcal{A}). \tag{6.2}$$

PROOF. The identity (6.2) follows if $M(\mathcal{A})$ is a set algebra: For then it is also a σ-algebra, and then $\sigma(\mathcal{A}) \subseteq M(\mathcal{A})$.

For $A \in M(\mathcal{A})$ define

$$\mathcal{M}_A = \{B \in M(\mathcal{A}) \mid A \setminus B,\ B \setminus A, \text{ and } A \cap B \text{ are in } M(\mathcal{A})\}.$$

This is easily seen to be a monotone class. Also notice that $B \in \mathcal{M}_A$ if and only if $A \in \mathcal{M}_B$ for all $A, B \in M(\mathcal{A})$. Furthermore, if $A \in \mathcal{A}$ then $\mathcal{A} \subseteq \mathcal{M}_A$, and so $M(\mathcal{A}) \subseteq \mathcal{M}_A$. In other words,

$$B \in \mathcal{M}_A \quad \text{for } A \in \mathcal{A} \text{ and } B \in M(\mathcal{A}).$$

By symmetry we then have

$$A \in \mathcal{M}_B \quad \text{for } A \in \mathcal{A} \text{ and } B \in M(\mathcal{A}),$$

and since $\mathcal{M}_B$ is a monotone class it follows that $M(\mathcal{A}) \subseteq \mathcal{M}_B$. Hence $M(\mathcal{A})$ is a set algebra as desired. □

7 • Complex analysis

[Should it be here?]

If $V \subseteq \mathbb{C}$ is open and $f : V \to \mathbb{C}$ is differentiable at every point in V, then we say that f is *holomorphic* on V. The set of functions holomorphic on V is denoted $\mathcal{H}(V)$.

THEOREM 7.1: *The Cauchy–Goursat Lemma*

If $f \in \mathcal{H}(V)$, then

$$\int_{\partial T} f(x)\,\mathrm{d}z = 0$$

for every triangle $T \subseteq V$.

PROOF. Notice that any triangle T can be subdivided into four smaller triangles $T^1, \dots, T^4$ whose corners are the corners and midpoints of the sides of T. We then clearly have

$$\int_{\partial T} g(z)\,\mathrm{d}z = \sum_{i=1}^{4} \int_{\partial T^i} g(z)\,\mathrm{d}z$$

for all $g \in C(T)$.

Let $T_0 \subseteq V$ be a triangle, and consider the integral

$$I = \int_{\partial T_0} f(z)\,\mathrm{d}z.$$

By the above considerations we have

$$|I| \leq 4 \left| \int_{\partial T_0^i} f(z)\,\mathrm{d}z \right|$$

for at least one value of i. For this value of i let $T_1 = T_0^i$. Continuing this process yields a sequence $(T_n)_{n \in \mathbb{N}}$ of triangles such that

$$|I| \leq 4^n \left| \int_{\partial T_n} f(z)\,\mathrm{d}z \right|$$

for $n \in \mathbb{N}_0$.

Furthermore, each of the four triangles in a subdivision of a triangle T have side lengths half of those of T, so

$$\operatorname{diam} T_n = 2^{-n} \operatorname{diam} T_0$$

for $n \in \mathbb{N}_0$. Thus there exists a point $z_0 \in \mathbb{C}$ such that $\bigcap_{n \in \mathbb{N}_0} T_n = \{z_0\}$ since (T_n) is a sequence of closed sets whose diameters tend to zero. It follows that

$$\sup_{z \in \partial T_n} |z - z_0| \leq 2^{-n} \operatorname{diam} T_0.$$

Given $\varepsilon > 0$ there exists an $r > 0$ such that

$$|f(z) - f(z_0) - f'(z_0)(z - z_0)| \leq \varepsilon |z - z_0|$$

for $z \in B(z_0, r)$, and there further exists an $N \in \mathbb{N}$ such that $n \geq N$ implies $T_n \subseteq B(z_0, r)$. Now notice that the function $z \mapsto f(z_0) + f'(z_0)(z - z_0)$ has an antiderivative, so its integral along ∂T_n is zero. Denoting the length of the curve ∂T_n by L_n we have $L_n \leq 2 \operatorname{diam} T_n$. Hence,

$$\begin{aligned} \left| \int_{\partial T_n} f(z) \, \mathrm{d}z \right| &\leq \int_{\partial T_n} |f(z) - f(z_0) - f'(z_0)(z - z_0)| \, \mathrm{d}z \\ &\leq L_n \varepsilon \sup_{z \in \partial T_n} |z - z_0| \\ &\leq 2^{-2n+1} \varepsilon (\operatorname{diam} T_0)^2 \end{aligned}$$

for $n \geq N$, and so

$$|I| \leq 2\varepsilon (\operatorname{diam} T_0)^2.$$

Since ε was arbitrary, it follows that $I = 0$ as desired. □

8 • The extended real line

Let $+\infty$ (or simply ∞) and $-\infty$ denote elements disjoint from $\mathbb{R}$. The *extended real line*, as a set, is then the union $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. We extend the ordering on $\mathbb{R}$ to $\overline{\mathbb{R}}$ by declaring that $-\infty < a$ for all $a \neq -\infty$ and that $\infty > a$ for all $a \neq \infty$. This clearly makes $\overline{\mathbb{R}}$ into a totally ordered set. We further equip it with order topology, i.e. the topology generated by open rays $\{x \in \overline{\mathbb{R}} \mid a < x\}$ and $\{x \in \overline{\mathbb{R}} \mid x < b\}$ for all $a, b \in \overline{\mathbb{R}}$. One easily sees that this is a Hausdorff topology (in fact it is T_5, which all order topologies are), so in particular singletons are closed and $\mathbb{R}$ is open.

Since $\overline{\mathbb{R}}$ is a topological space we can consider the Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$. Before characterising this we recall two elementary results:

(1) If $(X, \mathcal{T})$ is a topological space and $A \subseteq X$ is any subspace, then $\mathcal{B}(A) = \mathcal{B}(X)_A$. That is, the Borel σ-algebra generated by the subspace topology on A agrees with the restriction of the Borel σ-algebra on X to A.

(2) If $(X, \mathcal{E})$ is a measurable space and $A \in \mathcal{E}$, then

$$\mathcal{E} = \{E \cup F \mid E \in \mathcal{E}_A, F \in \mathcal{E}_{A^c}\}.$$

The inclusion '$\supseteq$' is obvious, and the other inclusion follows since if $B \in \mathcal{E}$, then

$$B = (B \cap A) \cup (B \cap A^c),$$

and $B \cap A \in \mathcal{E}_A$ and $B \cap A^c \in \mathcal{E}_{A^c}$ by the definition of the subspace σ-algebra.

This easily implies the following result:

PROPOSITION 8.1

The Borel σ-algebra on $\overline{\mathbb{R}}$ is given by

$$\mathcal{B}(\overline{\mathbb{R}}) = \left\{ A \cup S \;\middle|\; A \in \mathcal{B}(\mathbb{R}), S \subseteq \{\pm\infty\} \right\} = \{ B \subseteq \overline{\mathbb{R}} \mid B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R}) \}.$$

PROOF. To prove the first equality we only need to verify that $\mathcal{B}(\{\pm\infty\}) = 2^{\{\pm\infty\}}$. But this is obvious since both sets $\{\infty\}$ and $\{-\infty\}$ are closed. The second equality easily follows from the first. (Note that we cannot simply use the fact that we can characterise $\mathcal{B}(\mathbb{R})$ in terms of $\mathcal{B}(\overline{\mathbb{R}})$, which we can do since $\mathbb{R}$ is a subset of $\overline{\mathbb{R}}$, since we are trying to do the exact opposite.) □

References

Apostol, Tom (1974). *Mathematical Analysis*. 2nd ed. Pearson. 492 pp. ISBN: 0-201-00288-4.

Duistermaat, J. J. and J. A. C. Kolk (2004). *Multidimensional Real Analysis I: Differentiation*. 1st ed. Cambridge University Press. 422 pp. ISBN: 978-0-511-19558-7.

Folland, Gerald B. (2007). *Real Analysis: Modern Techniques and Their Applications*. 2nd ed. Wiley. 386 pp. ISBN: 0-471-31716-0.

Rudin, Walter (1976). *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill. 342 pp. ISBN: 0-07-085613-3.

Tao, Terence (2015a). *Analysis I*. 3rd ed. Springer. 350 pp. ISBN: 978-981-10-1789-6.

— (2015b). *Analysis II*. 3rd ed. Springer. 220 pp. ISBN: 978-981-10-1804-6.