Analysis II notes

David Pierce

June 20, 2003

These notes are for Math 272 at METU. I intend to edit them and add to them from time to time; the latest version is in the course directory, `<http://www.math.metu.edu.tr/~dpierce/courses/272/>`. The catalog description of the course is:

> Riemann–Stieltjes Integral. Infinite series and products. Sequences of functions. Inverse Function Theorem. Multiple Integrals.

These topics are covered in [1, chs 7, 8, 9, 12, 13 and 14]. This book will be my main reference, although [2] may also be useful. I shall also cover functions of bounded variation (in [1, ch. 6]), a topic left over from Math 271. My proof of Theorem 4.13 below is based on that of [3, ch. 1, § 1, Proposition 6]; my §§ 5 and 6 are influenced by [4, chs 2 and 3].

I prepare the notes, first of all, for my own use. They are only an outline of what is to be discussed in class. In particular, for the student, reading these notes is probably not an adequate substitute for coming to class. Your own notes, taken properly in class, will be richer and more complete than these notes.

The parts of these notes labelled 'proof' are generally only sketches of proofs. I leave it to the reader both:

- to recognize where details are missing, and
- to supply those details.

I might give the details in class, especially if I am asked to. I myself might ask for the details on an exam.

Some proofs are omitted entirely and are thus left to be given in class or to be done as exercises. These proofs too might be asked for on an exam.

I do intend to write my proofs (or proof-sketches) in complete sentences, with the usual sorts of punctuation (commas, semicolons, full stops/periods). Like any other writing (in English and Turkish and many other languages), the proofs are to be read left to right, top to bottom. Students should follow this example in writing their own complete proofs.

Every time a new class of functions (for example) is introduced, one should ask: What are some examples of functions that belong to this class? Which functions do *not* belong to this class?

Likewise, when a theorem is introduced, one should ask: What sorts of functions does the theorem apply to? What does the theorem *not* tell us? For example, if the theorem is an implication (an if-then statement), then can we prove the converse, or is there a counter-example?

Examples and counter-examples could be requested on an exam.

In studying the *proof* of a theorem, one should ask: What previous lemmas and theorems does the proof rely on? Is the proof in the style of earlier proofs; does it use familiar techniques; or does it introduce a new approach? Will a similar proof work to prove something else? Is there an alternative proof?

References

[1] Tom M. Apostol. *Mathematical analysis.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., second edition, 1974.

[2] Robert G. Bartle. *The elements of real analysis.* John Wiley & Sons, New York-London-Sydney, second edition, 1976.

[3] Raghavan Narasimhan. *Complex analysis in one variable.* Birkhäuser Boston Inc., Boston, MA, 1985.

[4] Michael Spivak. *Calculus on manifolds. A modern approach to classical theorems of advanced calculus.* W. A. Benjamin, Inc., New York-Amsterdam, 1965.

Contents

1 Variation

Here are some conventions, definitions and notations to be used throughout this section:

Let I be a compact interval of $\mathbb{R}$; so I is $[a, b]$ for some numbers a and b. Let f be a real-valued function on I. A **partition** of I is a subset that:

- is finite, and
- contains the endpoints a and b.

Let P be a partition of I. We may also write P as

$$\{x_0, x_1, \dots, x_n\}$$

for some positive integer n, where

$$a = x_0 < x_1 < \cdots < x_n = b.$$

If $0 < k \leqslant n$, then for $x_k - x_{k-1}$ we may write

$$\Delta x_k,$$

and for $f(x_k) - f(x_{k-1})$ we may write

$$\Delta f_k.$$

For $\sum_{k=1}^{n} |\Delta f_k|$ we may write

$$\sum(P).$$

The set of partitions of I can be denoted

$$\mathcal{P}I.$$

(This is *not* the power-set of I!) Suppose there is a number M such that for *all* P in $\mathcal{P}I$ we have $\sum(P) \leqslant M$. Then f is said to be **of bounded variation** on I. In this case, the supremum of the set $\{\sum(P) : P \in \mathcal{P}I\}$ exists; this supremum is called the **total variation** of f on I, and can be denoted

$$V_f I$$

or just V_f. Note that total variation is always non-negative. We may say that f itself is **bounded on I by A** if $|f(x)| \leqslant A$ for all x in I.

What sorts of functions are of bounded variation? How bad can a function be while still being of bounded variation?

Theorem 1.1. *Suppose f is monotone on I. Then f is of bounded variation on I, and in fact*

$$V_f I = |f(b) - f(a)|.$$

In particular, constant functions $x \mapsto a$ and the identity-function $x \mapsto x$ are of bounded variation on any compact interval.

Recall that monotonicity does not imply continuity. So, functions of bounded variation need not be continuous.

The converse of Theorem 1.1 fails: consider $x \mapsto x^2 : [-1,1] \to \mathbb{R}$.

Theorem 1.2. *If f is of bounded variation on I, then f is bounded on I, and in fact*

$$|f(x)| \leqslant V_f + |f(a)|$$

for all x in I.

Proof. Suppose f is of bounded variation. If $x \in I$, let P_x be the partition $\{a, x, b\}$ of I. Then

$$|f(x)| - |f(a)| \leqslant |f(x) - f(a)| \leqslant |f(x) - f(a)| + |f(b) - f(x)| = \sum(P_x) \leqslant V_f,$$

which yields the claim. □

The converse fails: consider $x \mapsto \begin{cases} x \sin \frac{1}{x}, & \text{if } x \in (0,1]; \\ 0, & \text{if } x = 0. \end{cases}$

Theorem 1.3. *Suppose f and g are of bounded variation on I. Then so are $f+g$ and $f-g$, and in fact*

$$V_{f \pm g} \leqslant V_f + V_g.$$

Moreover, $f \cdot g$ is also of bounded variation on I, and if A bounds f and B bounds g on I, then

$$V_{f \cdot g} \leqslant B \cdot V_f + A \cdot V_g.$$

Consequently, all polynomial functions are of bounded variation on compact intervals.

Lemma 1.4. *If P and Q are partitions of I such that $P \subseteq Q$, then*

$$\sum(P) \leqslant \sum(Q).$$

Theorem 1.5. *Suppose $x \in (a,b)$. Then f is of bounded variation on $[a,b]$ if and only if it is of bounded variation on both $[a,x]$ and $[x,b]$. In either case,*

$$V_f[a,b] = V_f[a,x] + V_f[x,b].$$

Proof. Say f is of bounded variation on $[a,b]$. Let Q be a partition of $[a,x]$, and R be a partition of $[x,b]$. Then $Q \cup R$ is a partition of $[a,b]$. Hence

$$\sum(Q) + \sum(R) = \sum(Q \cup R) \leqslant V_f[a,b].$$

Therefore f is of bounded variation on $[a,x]$ and on $[x,b]$, and

$$V_f[a,x] + V_f[x,b] \leqslant V_f[a,b]. \tag{i}$$

Conversely, say f is of bounded variation both on $[a, x]$ and on $[x, b]$. If P is a partition of $[a, b]$, then so is $P \cup \{x\}$, and the latter is $Q \cup R$ for some partitions Q and R of $[a, x]$ and $[x, b]$ respectively. Hence (by Lemma 1.4)

$$\sum(P) \leqslant \sum(P \cup \{x\}) = \sum(Q) + \sum(R) \leqslant V_f[a, x] + V_f[x, b].$$

Therefore f is of bounded variation on $[a, b]$, and

$$V_f[a, b] \leqslant V_f[a, x] + V_f[x, b]. \tag{ii}$$

Either of the foregoing hypotheses now yields the other and hence yields both (i) and (ii). □

Theorem 1.5 (with Theorem 1.1) gives another proof that polynomial functions are of bounded variation on compact intervals.

Theorem 1.6. *Suppose that f is continuous on $[a, b]$ and differentiable on (a, b) and that f' is bounded on (a, b). Then f is of bounded variation on $[a, b]$.*

Proof. For every partition $\{x_0, \ldots, x_n\}$ in $\mathcal{P}[a, b]$, for every k in $\{1, \ldots, n\}$, by the Mean-Value Theorem there is t_k in (x_{k-1}, x_k) such that

$$\frac{\Delta f_k}{\Delta x_k} = f'(t_k).$$

Hence, if A is an upper bound of $\{f'(x) : x \in (a, b)\}$, then

$$\sum_{k=1}^{n} |\Delta f_k| = \sum_{k=1}^{n} |f'(t_k)| \, \Delta x_k \leqslant \sum_{k=1}^{n} A \Delta x_k = A(b - a),$$

so f is of bounded variation by definition. □

Hence for example the function

$$x \mapsto \begin{cases} x \sin \log x, & \text{if } x \in (0, 1]; \\ 0, & \text{if } x = 0 \end{cases}$$

is of bounded variation. The converse of Theorem 1.6 fails: consider $x \mapsto \sqrt{x} : [0, 1] \to \mathbb{R}$.

Theorem 1.7. *Suppose f is of bounded variation on I. Then the functions $x \mapsto V_f[a, x]$ and $x \mapsto V_f[a, x] - f(x)$ are increasing on I.*

Proof. Suppose $a \leqslant x \leqslant y \leqslant b$. By Theorem 1.5, we have

$$V_f[a, x] \leqslant V_f[a, y].$$

Hence $x \mapsto V_f[a, x]$ is increasing. We also have

$$V_f[a, x] + f(y) - f(x) \leqslant V_f[a, x] + |f(y) - f(x)| \leqslant V_f[a, x] + V_f[x, y] = V_f[a, y],$$

whence $V_f[a, x] - f(x) \leqslant V_f[a, y] - f(y)$. Therefore $x \mapsto V_f[a, x] - f(x)$ is increasing. □

In the proof, note that if, for example, $x = y$, then $[x, y]$ is not in fact an interval, although we can still understand the total variation $V_f[x, y]$ to be 0.

Theorem 1.8. *A function is of bounded variation if and only if it is the difference of two increasing functions.*

Lemma 1.9. *Suppose f is of bounded variation on $[a, b]$, and $c \in [a, b]$. Then f is continuous at c if and only if $x \mapsto V_f[a, x]$ is continuous at c.*

Proof (not given in class). Let V be the increasing function $x \mapsto V_f[a, x]$. If $a \leqslant x \leqslant y \leqslant b$, then

$$0 \leqslant |f(y) - f(x)| \leqslant V_f[x, y] = V(y) - V(x).$$

Hence continuity of V implies that of f.

Suppose conversely that f is continuous at c. Let $\varepsilon > 0$. It is enough to find positive numbers δ_ℓ and δ_r such that, for all x in $[a, b]$, we have the two implications

$$\begin{aligned} c - \delta_\ell < x \leqslant c &\implies V(c) - V(x) < \varepsilon, \\ c \leqslant x < c + \delta_\mathrm{r} &\implies V(x) - V(c) < \varepsilon. \end{aligned}$$

Now, by continuity of f at c, we can pick a positive δ so that, for all x in $[a, b]$,

$$c - \delta < x < c + \delta' \implies |f(c) - f(x)| < \frac{\varepsilon}{2}.$$

There is a partition P of $[a, c]$ such that, for all finer partitions $\{u_0, \dots, u_m\}$, we have

$$V(c) \leqslant \sum_{k=1}^{m} |\Delta f_k| + \varepsilon/2.$$

Let y be the greatest element of $P \smallsetminus \{c\}$, and let

$$\delta_\ell = \min(c - y, \delta).$$

Suppose $c - \delta_\ell < x < c$. We may assume $u_{m-1} = x$; hence

$$V(c) \leqslant \sum_{k=1}^{m-1} |\Delta f_k| + \varepsilon \leqslant V(x) + \varepsilon.$$

Thus we have the first implication above; the second is obtained similarly. $\square$

Theorem 1.10. *A continuous function is of bounded variation if and only if it is the difference of two continuous increasing functions.*

2 The Riemann–Stieltjes Integral

Let us continue to use the notation of the previous section, letting also g be a real-valued function on I. If g is in fact differentiable, and if one writes $g'(x)$ as

$\mathrm{d}\,g/\mathrm{d}\,x$, then one might write $g'(x)\,\mathrm{d}\,x$ as $\mathrm{d}\,g$, and $\int_a^b f(x)g'(x)\,\mathrm{d}\,x$ as $\int_a^b f\,\mathrm{d}\,g$. If f is also differentiable, then the rule of integration by parts is

$$\int_a^b f\,\mathrm{d}\,g = f(b)g(b) - f(a)g(a) - \int_a^b g\,\mathrm{d}\,f. \tag{iii}$$

We shall show that this equation is meaningful and true even if f and g are not differentiable.

If $P, Q \in \mathcal{P}I$, and $P \subseteq Q$, let us say that Q is **finer** than P (also, P is **coarser** than Q).

Theorem 2.1. *There is at most one number A such that, for every positive ε, there is a partition P of I such that, for all partitions $\{u_0, \dots, u_m\}$ of I that are finer than P, and for all t_k in $[u_{k-1}, u_k]$ (where $k \in \{1, 2, \dots, m\}$),*

$$\left|\sum_{k=1}^{m} f(t_k)\triangle g_k - A\right| < \varepsilon.$$

The sum in the theorem is a **Riemann–Stieltjes sum** of f with respect to g for the partition P. If the number A exists, it is denoted

$$\int_a^b f\,\mathrm{d}\,g.$$

This is a **Riemann–Stieltjes integral**, and in particular it is the Riemann integral of f on I *with respect to* g. The set of all functions f on I whose Riemann integrals with respect to g exist is denoted

$$R(g).$$

If g is differentiable, with constant derivative 1, then $\int_a^b f\,\mathrm{d}\,g$ is just $\int_a^b f$, the ordinary Riemann integral.

Theorem 2.2. *If $f \in R(g)$, then $g \in R(f)$, and also equation (iii) holds.*

Proof. Let $\{x_0, \dots, x_n\}$ be a partition P as in Theorem 2.1. Choose some t_k in $[x_{k-1}, x_k]$. Then

$$f(b)g(b) - f(a)g(a) - \sum_{k=1}^{n} g(t_k)\triangle f_k =$$
$$\sum_{k=1}^{n} f(x_k)g(x_k) - \sum_{k=1}^{n} f(x_{k-1})g(x_{k-1}) - \sum_{k=1}^{n} g(t_k)f(x_k) + \sum_{k=1}^{n} g(t_k)f(x_{k-1}) =$$
$$\sum_{k=1}^{n} f(x_k)(g(x_k) - g(t_k)) + \sum_{k=1}^{n} f(x_{k-1})(g(t_k) - g(x_{k-1})).$$

This is a Riemann–Stieltjes sum for f with respect to g corresponding to the partition $\{x_0, t_1, x_1, \dots, x_{n-1}, t_n, x_n\}$. This partition is finer than P, so the sum is within ε of $\int_a^b f\,\mathrm{d}\,g$. Therefore the Riemann–Stieltjes sum $\sum_{k=1}^{n} g(t_k)\triangle f_k$ is within ε of $f(b)g(b) - f(a)g(a) - \int_a^b f\,\mathrm{d}\,g$. □

Theorem 2.3. *$R(g)$ is a vector-space, that is, it contains the 0-function and is also closed under the taking of linear combinations. In fact, if $f_i \in R(g)$ and $a_i \in \mathbb{R}$ (where $i < n$), then*

$$\int_a^b \sum_{i<n} a_i f_i \,\mathrm{d}\, g = \sum_{i<n} a_i \int_a^b f_i \,\mathrm{d}\, g.$$

Hence also

$$\bigcap_{i<n} R(f_i) \subseteq R\bigl(\sum_{i<n} a_i f_i\bigr),$$

and if $g \in \bigcap_{i<n} R(f_i)$, then

$$\int_a^b g \,\mathrm{d} \sum_{i<n} a_i f_i = \sum_{i<n} a_i \int_a^b g \,\mathrm{d}\, f_i.$$

If $\int_a^b f \,\mathrm{d}\, g$ exists, then we may define

$$\int_b^a f \,\mathrm{d}\, g = -\int_a^b f \,\mathrm{d}\, g.$$

Also, $\int_a^a f \,\mathrm{d}\, g = 0$.

Theorem 2.4. *If $x \in I$, and any two of the three integrals*

$$\int_a^x f \,\mathrm{d}\, g, \qquad \int_x^b f \,\mathrm{d}\, g, \qquad \int_b^a f \,\mathrm{d}\, g$$

exist, then they all exist, and their sum is 0.

Theorem 2.5. *Let h be an increasing bijection from an interval $[c, d]$ onto $[a, b]$. If $f \in R(g)$, then $f \circ h \in R(g \circ h)$, and*

$$\int_c^d f \circ h \,\mathrm{d}(g \circ h) = \int_{h(c)}^{h(d)} f \,\mathrm{d}\, g.$$

Proof. Pick P as in Theorem 2.1. This determines a partition $h^{-1}(P)$ of $[c, d]$. A Riemann–Stieltjes sum of $f \circ h$ on $[c, d]$ with respect to $g \circ h$ for a partition finer than $h^{-1}(P)$ can be understood as a Riemann–Stieltjes sum of f on $[a, b]$ with respect to g for a partition finer than P; hence it is within ε of $\int_a^b f \,\mathrm{d}\, g$. □

Theorem 2.6. *If f is bounded on I, and $f \in R(g)$, and g has a continuous derivative on I, then fg' is Riemann-integrable on I, and*

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

Proof. For any partition $\{x_0, \dots, x_n\}$ of I, by the Mean-Value Theorem, there are u_k in (x_{k-1}, x_k) such that $\Delta g_k = g'(u_k)\Delta x_k$. Hence, for any t_k in $[x_{k-1}, x_k]$, the Riemann–Stieltjes sum

$$\sum_{k=1}^{n} f(t_k)\Delta g_k$$

is precisely

$$\sum_{k=1}^{n} f(t_k)g'(u_k)\Delta x_k.$$

We have

$$\left|\sum_{k=1}^{n} f(t_k)g'(u_k)\Delta x_k - \sum_{k=1}^{n} f(t_k)g'(t_k)\Delta x_k\right| = \sum_{k=1}^{n} f(t_k)\left|g'(u_k) - g'(t_k)\right|\Delta x_k.$$

Since f is bounded on I, and g'—being continuous on a compact interval—is uniformly continuous on I, we conclude that, in the last equation, the right member can be made smaller than $\varepsilon/2$, provided that the partition $\{x_0, \dots, x_n\}$ is sufficiently fine. This partition can also be fine enough that the sum $\sum_{k=1}^{n} f(t_k)\Delta g_k$ is within $\varepsilon/2$ of $\int_a^b f \,\mathrm{d}\, g$. In this case, we have

$$\left|\int_a^b f \,\mathrm{d}\, g - \sum_{k=1}^{n} f(t_k)g'(t_k)\Delta x_k\right| < \varepsilon,$$

which establishes the claim. □

Lemma 2.7. *Suppose $u \in (a, b]$, and f is continuous from the left at u, that is, for every positive ε, there is a positive δ such that for all t in I,*

$$u - \delta < t \leqslant u \implies |f(u) - f(t)| < \varepsilon.$$

Let g be defined so that $g(x) = \begin{cases} 0, & \text{if } a \leqslant x < u; \\ 1, & \text{if } u \leqslant x \leqslant b. \end{cases}$ *Then $f \in R(g)$ and*

$$\int_a^b f \,\mathrm{d}\, g = f(u).$$

Proof. Suppose $s \in (a, u)$. For every partition that is finer than $\{a, s, u, b\}$, every Riemann–Stieltjes sum for f with respect to g is

$$f(t)(g(u) - g(s)),$$

that is, $f(t)$, for some t in $[s, u]$. With ε and δ as in the statement, if we choose s in $(u - \delta, u)$, then $|f(u) - f(t)| < \varepsilon$. □

Theorem 2.8. *Every finite sum can be expressed as a Riemann–Stieltjes integral. In fact, if f is continuous on $[a, b]$, and $a < x_1 < \cdots < x_n = b$, then*

$$\sum_{k=1}^{n} f(x_k) = \int_a^b f \,\mathrm{d}\, g$$

for some function g.

Proof. Let $g = \max\{k : x_k \leqslant x\}$. □

In the theorem and its proof, if $x_k = k$ for each k, then g is the greatest-integer function, $x \mapsto \lfloor x \rfloor$. Hence:

Theorem 2.9 (Euler's summation-formula). *If f is continuously differentiable on $[0,n]$, then*

$$\sum_{k=1}^{n} f(k) = \int_0^n f + \int_0^n (x - \lfloor x \rfloor) f'(x) \,\mathrm{d}\, x;$$
$$\sum_{k=0}^{n} f(k) = \int_0^n f + \int_0^n \left(x - \lfloor x \rfloor - \frac{1}{2}\right) f'(x) \,\mathrm{d}\, x + \frac{1}{2}(f(n) + f(0)).$$

Proof. Just calculate:

$$\sum_{k=1}^{n} f(k) = \int_0^n f(x) \,\mathrm{d}\lfloor x \rfloor = \int_0^n f - \int_0^n f(x) \,\mathrm{d}(x - \lfloor x \rfloor) =$$
$$\int_0^n f - \int_0^n (x - \lfloor x \rfloor) \,\mathrm{d}\, f(x) = \int_0^n f + \int_0^n (x - \lfloor x \rfloor) f'(x) \,\mathrm{d}\, x,$$

by Theorems 2.8, 2.3, 2.2 and 2.6 respectively. Note however that, to have the second equation, we must know that $\int_a^b f$ does in fact exist. Since f is continuous, its Riemann-integrability will be a consequence of Theorem 2.18. For the second equation, add $f(0)$ to both sides, and use

$$\int_0^n f' = f(n) - f(0),$$

which will be Theorem 2.22. □

As noted in the last proof, we need to be able to say *which* functions are Riemann-integrable with respect to a given function. Towards this end, we make some definitions—but we must assume that f is bounded:

$$\mathrm{m}_k(f) = \inf\{f(x) : x \in [x_{k-1}, x_k]\}, \qquad \mathrm{L}(P, f, g) = \sum_{k=1}^{n} \mathrm{m}_k(f) \Delta g_k,$$
$$\mathrm{M}_k(f) = \sup\{f(x) : x \in [x_{k-1}, x_k]\}, \qquad \mathrm{U}(P, f, g) = \sum_{k=1}^{n} \mathrm{M}_k(f) \Delta g_k.$$

Lemma 2.10. *Assume g is increasing. For all partitions P and Q of I,*

$$\mathrm{L}(P, f, g) \leqslant \mathrm{L}(P \cup Q, f, g) \leqslant \mathrm{U}(P \cup Q, f, g) \leqslant \mathrm{U}(Q, f, g).$$

If g is increasing, then by the lemma we can define

$$\underline{\int_a^b} f \,\mathrm{d}\, g = \sup\{\mathrm{L}(P, f, g) : P \in \mathcal{P}I\}, \qquad \overline{\int_a^b} f \,\mathrm{d}\, g = \inf\{\mathrm{U}(P, f, g) : P \in \mathcal{P}I\}.$$

Moreover:

Lemma 2.11. *Assume g is increasing. Then*

$$\mathrm{L}(P, f, g) \leqslant \underline{\int_a^b} f \,\mathrm{d}\, g \leqslant \overline{\int_a^b} f \,\mathrm{d}\, g \leqslant \mathrm{U}(P, f, g)$$

for all partitions P of I.

Suppose g is increasing. Let us say that f satisfies **Riemann's condition** with respect to g on I if for each positive ε there is a partition P of I such that

$$\mathrm{U}(P, f, g) - \mathrm{L}(P, f, g) < \varepsilon.$$

Lemma 2.12. *Assume g is increasing. Then the following conditions are equivant:*

(0) f satisfies Riemann's condition with respect to g on I.

(1) $\underline{\int_a^b} f \,\mathrm{d}\, g = \overline{\int_a^b} f \,\mathrm{d}\, g$.

(2) $f \in R(g)$.

Proof. If condition (0) holds, then $\overline{\int_a^b} f \,\mathrm{d}\, g - \underline{\int_a^b} f \,\mathrm{d}\, g < \varepsilon$ for all positive ε, so condition (1) holds.

Suppose $\varepsilon > 0$. By Lemma 2.10, we can let P be a partition of I such that $\mathrm{L}(P, f, g)$ is within ε of $\underline{\int_a^b} f \,\mathrm{d}\, g$, *and* $\mathrm{U}(P, f, g)$ is within ε of $\overline{\int_a^b} f \,\mathrm{d}\, g$. But if $P = \{x_0, \dots, x_n\}$, for any t_k in $[x_{k-1}, x_k]$, we have

$$\mathrm{L}(P, f, g) \leqslant \sum_{i=1}^{n} f(t_k) \Delta g_k \leqslant \mathrm{U}(P, f, g);$$

if condition (1) holds, then by Lemma 2.11, the sum $\sum_{i=1}^{n} f(t_k)\Delta g_k$ must be within ε of the common value of $\underline{\int_a^b} f \,\mathrm{d}\, g$ and $\overline{\int_a^b} f \,\mathrm{d}\, g$. Hence this common value is $\int_a^b f \,\mathrm{d}\, g$, and condition (2) holds.

Finally, suppose (2) holds. We can choose P so that

$$\left| \sum_{i=1}^{n} f(t_k) \Delta g_k - \int_a^b f \,\mathrm{d}\, g \right| < \frac{\varepsilon}{4}$$

for all t_k in $[x_{k-1}, x_k]$. But also, since g is increasing, we can choose the t_k so that, in addition,

$$\left| \sum_{i=1}^{n} f(t_k) \Delta g_k - \mathrm{L}(P, f, g) \right| < \frac{\varepsilon}{4}.$$

Then $\left|\mathrm{L}(P, f, g) - \int_a^b f \,\mathrm{d}\, g\right| < \varepsilon/2$. Likewise, $\left|\mathrm{U}(P, f, g) - \int_a^b f \,\mathrm{d}\, g\right| < \varepsilon/2$. So condition (0) holds. □

Theorem 2.13. *Assume g is increasing. If $f_0, f_1 \in R(g)$, and $f_0 \leqslant f_1$ (that is, $f_0(x) \leqslant f_1(x)$ for all x in I), then*

$$\int_a^b f_0 \,\mathrm{d}\, g \leqslant \int_a^b f_1 \,\mathrm{d}\, g.$$

Lemma 2.14. *Assume g is increasing. If $f \in R(g)$, then $|f| \in R(g)$ and*

$$\left| \int_a^b f \,\mathrm{d}\, g \right| \leqslant \int_a^b |f| \,\mathrm{d}\, g.$$

Proof. Since $\big||f(x)| - |f(y)|\big| \leqslant |f(x) - f(y)|$, we have

$$\mathrm{M}_k(|f|) - \mathrm{m}_k(|f|) \leqslant \mathrm{M}_k(f) - \mathrm{m}_k(f),$$

hence $\mathrm{U}(P, |f|, g) - \mathrm{L}(P, |f|, g) \leqslant \mathrm{U}(P, f, g) - \mathrm{L}(P, f, g)$. Now use Lemma 2.12. □

Lemma 2.15. *Assume g is increasing. If $f \in R(g)$, then $f^2 \in R(g)$.*

Proof. Suppose M bounds $|f|$. Then

$$\begin{aligned}\mathrm{M}_k(f^2) - \mathrm{m}_k(f^2) = {}&\\ &(\mathrm{M}_k(|f|) + \mathrm{m}_k(|f|))(\mathrm{M}_k(|f|) - \mathrm{m}_k(|f|)) \leqslant\\ &\qquad 2M(\mathrm{M}_k(|f|) - \mathrm{m}_k(|f|)).\end{aligned}$$

Now use Lemma 2.12. □

If g is merely of bounded variation, then g is the difference $g_0 - g_1$ of increasing functions, by Theorem 1.8, and $R(g_0) \cap R(g_1) \subseteq R(g)$, by Theorem 2.3. But what about the reverse inclusion?

Lemma 2.16. *Assume g is of bounded variation. Let V be the function $x \mapsto V_g[a, x]$. If f is bounded, and $f \in R(g)$, then $f \in R(V)$.*

Proof (not given in class). Since V is increasing, we can use Lemma 2.12. We have

$$\begin{aligned}\mathrm{U}(P, f, V) - \mathrm{L}(P, f, V) = \sum_{k=1}^{n}(\mathrm{M}_k(f) - \mathrm{m}_k(f))\Delta V_k = {}&\\ \sum_{k=1}^{n}(\mathrm{M}_k(f) - \mathrm{m}_k(f))(\Delta V_k - |\Delta g_k|) + \sum_{k=1}^{n}(\mathrm{M}_k(f) - \mathrm{m}_k(f))\,|\Delta g_k|\,,&\end{aligned}$$

and we can make each of the last summations less than $\varepsilon/2$. Indeed, if $|f| \leqslant M$, then

$$\sum_{k=1}^{n}(\mathrm{M}_k(f) - \mathrm{m}_k(f))(\Delta V_k - |\Delta g_k|) \leqslant 2M(V(b) - \sum_{k=1}^{n}|\Delta g_k|),$$

which is less than $\varepsilon/2$ if P is fine enough. We can also choose t_k and t'_k in $[x_{k-1}, x_k]$ so that

$$\mathrm{M}_k(f) - \mathrm{m}_k(f) < |f(t_k) - f(t'_k)| + \frac{\varepsilon}{4V(b)}.$$

We may assume also that $f(t_k) - f(t'_k)$ has the same sign as Δg_k. Then

$$\sum_{k=1}^{n}(\mathrm{M}_k(f) - \mathrm{m}_k(f))\,|\Delta g_k| < \sum_{k=1}^{n}(f(t_k) - f(t'_k))\Delta g_k + \frac{\varepsilon}{4}.$$

If P is fine enough, then $\sum_{k=1}^{n}(f(t_k) - f(t'_k))\Delta g_k < \varepsilon/4$. □

Theorem 2.17. *Assume g is of bounded variation. Then there are increasing functions g_0 and g_1 such that $g = g_0 - g_1$ and every bounded function in $R(g)$ is also in $R(g) = R(g_0) \cap R(g_1)$.*

Theorem 2.18. *If g is of bounded variation, then $R(g)$ contains all continuous functions.*

Proof. It is enough to prove the theorem in case g is increasing and $g(a) < g(b)$. Suppose f is continuous (on I); then f is uniformly continuous. Let $\varepsilon > 0$. Then there is a positive δ such that for all x and y (in I),

$$|x - y| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{g(b) - b(a)}.$$

Let P be a partition fine enough that $\Delta x_k < \delta$ in each case. Since f is continuous, we have

$$\mathrm{M}_k(f) - \mathrm{m}_k(f) = |f(t_k) - f(t'_k)| < \frac{\varepsilon}{g(b) - b(a)}$$

for some t_k and t'_k in $[x_{k-1}, x_k]$, for each k; this means

$$\mathrm{U}(P, f, g) - \mathrm{L}(P, f, g) < \sum_{k=1}^{n} \frac{\varepsilon}{g(b) - g(a)} \Delta g_k = \varepsilon.$$

Thus, Riemann's condition is met. □

Corollary 2.19. *All functions of bounded variation are Riemann-integrable.*

Lemma 2.20 (Mean-Value Theorem for Integrals). *Suppose f is continuous, and g is increasing. Then*

$$\int_a^b f \,\mathrm{d}\, g = f(t)(g(b) - g(a)) = f(t) \int_a^b \mathrm{d}\, g$$

for some t in (a, b).

Proof. The second equation is clear from the definition of the Riemann–Stieltjes integral, or from Theorem 2.2. For the first equation, work with the partition $\{a, b\}$ of I. We have

$$\mathrm{m}_1(f)(g(b) - g(a)) \leqslant \int_a^b f \,\mathrm{d}\, g \leqslant \mathrm{M}_1(f)(g(b) - g(a)),$$

so $(\int_a^b f \,\mathrm{d}\, g)/(g(b) - g(a))$ is between $\mathrm{m}_1(f)$ and $\mathrm{M}_1(f)$. Now use the intermediate-value theorem for continuous functions. □

Theorem 2.21 (First Fundamental, of Calculus). *If g is increasing, then the function*

$$x \mapsto \int_a^x f \,\mathrm{d}\, g$$

is well-defined on I, and is differentiable, with derivative $f(u)g'(u)$, at every point u of I where f is continuous and g is differentiable.

Proof. If u is such a point, and $u+h \in I$, then

$$\frac{\int_a^{u+h} f \,\mathrm{d}\, g - \int_a^u f \,\mathrm{d}\, g}{h} = \frac{\int_u^{u+h} f \,\mathrm{d}\, g}{h} = f(t)\frac{g(u+h)-g(u)}{h}$$

for some t between u and $u+h$ inclusive. $\square$

Theorem 2.22 (Second Fundamental, of Calculus). *For any function g that is continuous on I and differentiable on (a,b),*

$$\int_a^b g' = g'(b) - g'(a),$$

provided the integral exists.

Proof. For any partition $\{x_0,\dots,x_n\}$ of I, we can pick t_k in (x_{k-1},x_k) such that $g'(t_k)\Delta x_k = \Delta g'_k$, whence the Riemann sum $\sum_{k=1}^n g'(t_k)\Delta x_k$ is just $g'(b) - g'(a)$. $\square$

3 Infinite series and products

An infinite sequence is a function with domain $\{n \in \mathbb{Z} : k \leqslant n\}$ for some k in $\mathbb{Z}$. Usually the domain is $\mathbb{N}$ (the set $\{0,1,2,\dots\}$ of natural numbers) or $\mathbb{Z}_+$ (the set $\{1,2,3,\dots\}$ of positive integers).

Informally, a sequence (a_n) has a real number b as a limit, provided that a_n is close to b whenever n is sufficiently large. We can understand 'sufficiently large' to mean 'close to ∞'. Then we can allow the limit b to be ∞ or $-\infty$ as well.

To be more precise: Recall that a neighborhood of a real number x is a set of real numbers that has, as a subset, an open interval that contains x. We can define a **neighborhood of** ∞ to be a set of real numbers that has, as a subset, an interval of the form (a,∞). Similarly, neighborhoods of $-\infty$ have subsets (∞, a).

Now we can say that the sequence (a_n) of real numbers has the **limit** b (where $b \in \{-\infty\}\cup\mathbb{R}\cup\{\infty\}$) if, for every neighborhood U of b, there is a neighborhood U' of ∞ such that, for all n in the domain of the sequence,

$$n \in U' \implies b_n \in U.$$

In this case, we write

$$\lim_{n\to\infty} a_n = b.$$

If b is **finite** (that is, $b \in \mathbb{R}$), then the sequence is said to **converge to** b. If b is infinite, then the sequence **diverges to** b. If the sequence has no limit, then the sequence simply **diverges**.

The **extended real number system**, $\mathbb{R}^*$, consists of the elements of the set

$$\{-\infty\}\cup\mathbb{R}\cup\{\infty\}.$$

This can be ordered in the obvious way, so that $-\infty < x < \infty$ for all x in $\mathbb{R}$. Defining open and closed intervals of $\mathbb{R}^*$ is somewhat easier than for $\mathbb{R}$. If $a, b \in \mathbb{R}^*$, then

$$(a,b) = \{x \in \mathbb{R}^* : a < x < b\}; \qquad [a,b] = \{x \in \mathbb{R}^* : a \leqslant x \leqslant b\}.$$

An **open** interval of $\mathbb{R}^*$ is an interval (a,b); a **closed**, $[a,b]$. Note that $\mathbb{R}$ is an open, but not a closed, interval of $\mathbb{R}^*$.

Let f be the function from $[-1,1]$ to $\mathbb{R}^*$ given by

$$f(x) = \begin{cases} -\infty, & \text{if } x = -1; \\ \frac{x}{1-x^2}, & \text{if } -1 < x < 1; \\ \infty, & \text{if } x = 1. \end{cases}$$

Then f is a bijective, order-preserving function. This means that both f^{-1} and f send open sets to open sets; that is, both f and f^{-1} are continuous. However, we cannot extend the *usual* metric on $\mathbb{R}$ so as to make $\mathbb{R}^*$ into a metric space; $\mathbb{R}^*$ is simply a **topological space**.

If a subset A of $\mathbb{R}$ has no upper bound in $\mathbb{R}$, then we may write

$$\sup A = \infty;$$

if no lower bound, then $\inf A = -\infty$. Also, $\sup \varnothing = \inf \mathbb{R} = -\infty$, and $\inf \varnothing = \sup \mathbb{R} = \infty$. Thus, every subset of $\mathbb{R}$ has a supremum and in infimum in $\mathbb{R}^*$.

Every monotone sequence has a limit in $\mathbb{R}^*$.

A sequence (a_n) has the **limit superior** b in $\mathbb{R}$ if for all positive ε:

$$\exists M\ \forall n\ (M < n \implies a_n < b + \varepsilon);$$

and

$$\forall M\ \exists n\ (M < n \ \&\ b - \varepsilon < a_n).$$

In this case, we write

$$\limsup_{n\to\infty} a_n = b.$$

Theorem 3.1. *The following are equivalent:*

(1) $\limsup_{n\to\infty} a_n = b$.

(2) $\lim_{n\to\infty} \sup\{a_m : n < m\} = b$.

(3) (a_n) *has a subsequence converging to* b*, but not a subsequence converging to a greater limit.*

If $\{a_n\}$ has no upper bound, then

$$\limsup_{n\to\infty} a_n = \infty;$$

if there *is* an upper bound, but no finite limit superior, then

$$\limsup_{n\to\infty} a_n = -\infty.$$

So every sequence has a limit superior in $\mathbb{R}^*$. The **limit inferior** is defined similarly; in fact,

$$\liminf_{n\to\infty} a_n = -\limsup_{n\to\infty} -a_n.$$

Examples 3.2.

a_n	$\limsup_{n\to\infty} a_n$	$\liminf_{n\to\infty} a_n$
$(-1)^n$	1	-1
$\frac{n+1}{n}(1+(-1)^n)$	2	0
$n(1+(-1)^n)$	∞	0
$\log n$	∞	∞

For a sequence $(a_p, a_{p+1}, \dots)$, the sequence of its **partial sums** is

$$\left(\sum_{i=p}^{n} a_i : n = p, p+1, \dots\right).$$

Considered with respect to its sequence of partial sums, a sequence is a **series**; the series **converges** to a limit b if the sequence of partial sums does; in this case, we write

$$\sum_{i=p}^{\infty} a_i = b.$$

A sequence (a_n), as a series, can be written

$$\sum a_n.$$

Theorem 3.3. *The set of convergent series is a vector-space: if $\sum a_n$ and $\sum b_n$ converge, and $c, d \in \mathbb{R}$, then $\sum(ca_n + db_n)$ converges, and*

$$\sum_{k=p}^{\infty}(ca_n + db_n) = c\sum_{k=p}^{\infty} a_n + d\sum_{k=p}^{\infty} b_n.$$

Lemma 3.4 (Cauchy condition). *The series $\sum a_n$ converges if and only if, for every positive ε, there is M such that, for all k in $\mathbb{N}$,*

$$n > M \implies \left|\sum_{m=n}^{n+k} a_m\right| < \varepsilon.$$

Theorem 3.5. *If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.*

Proof. Use the Cauchy condition when $k = 1$. □

It is clear that, if $\sum a_n$ converges, then so does $\sum(a_{2n} + a_{2n+1})$.

Lemma 3.6. *If $\sum(a_{2n} + a_{2n+1})$ converges, and $\lim_{n\to\infty} a_n = 0$, then $\sum a_n$ converges.*

Proof. Let $\sum_{k=p}^{\infty}(a_{2k} + a_{2k+1}) = b$. Let M be such that

$$n > M \implies \left|\sum_{k=p}^{p+n}(a_{2k} + a_{2k+1}) - b\right|, \left|a_{2(p+n)+1}\right| < \frac{\varepsilon}{2}.$$

Then also $n > M \implies \left|\sum_{k=p}^{2(p+n)} a_k - b\right|, \left|\sum_{k=p}^{2(p+n)+1} a_k - b\right| < \varepsilon$. □

A series $\sum a_n$ is **alternating** if $(-1)^n a_n \geqslant 0$ for each n, or $(-1)^n a_n \leqslant 0$ for each n.

Theorem 3.7. *An alternating series $\sum a_n$ converges if $\lim_{n\to\infty} a_n = 0$.*

Proof. Assume $(-1)^n a_n \geqslant 0$ in each case. Then $a_{2n} - a_{2n+1} \geqslant 0$, and

$$\sum_{k=0}^{n}(a_{2k} - a_{2k+1}) = a_0 - \sum_{k=0}^{n-1}(a_{2k+1} - a_{2k+2}) - a_{2n+1},$$

so partial sums of $\sum(a_{2n} - a_{2n+1})$ are bounded. Therefore the latter series converges. Now use Lemma 3.6. $\square$

Example 3.8. $\sum(-1)^n/n$ converges.

Theorem 3.9. *If $|x| < 1$, then*

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

Proof. $\sum_{k=0}^{n} x^k = (1-x^{n+1})/(1-x)$. If $|x| < 1$, then the partial sums converge to $1/(1-x)$. $\square$

Lemma 3.10 (Comparison Test). *If $0 \leqslant a_n \leqslant cb_n$ when n is large enough, for some non-zero c, and $\sum b_n$ converges, then $\sum a_n$ converges.*

Theorem 3.11 (Limit Comparison Test). *If $\lim_{n\to\infty} a_n/b_n$ is finite and non-zero, then*

$$\sum a_n \text{ converges} \iff \sum b_n \text{ converges.}$$

Theorem 3.12 (Integral Test). *If f is monotone on $[0,\infty)$, then*

$$\sum f(n) \text{ converges} \iff \lim_{x\to\infty}\int_0^x f \text{ is finite.}$$

Proof. Assume f is decreasing. If $f(x) < 0$ for some x, then each member of the equivalence fails. Suppose $f(x) \geqslant 0$ for all x. Since $f(k+1) \leqslant f(x) \leqslant f(k)$ when $k \leqslant x \leqslant k+1$, we have

$$\sum_{k=1}^{n} f(k) \leqslant \int_0^n f \leqslant \sum_{k=0}^{n-1} f(k).$$

If $\int_0^\infty f$ exists, then so does $\sum_{k=1}^{\infty} a_k$; if not, not. $\square$

Examples 3.13. We have

$$\int_1^x \frac{\mathrm{d}\,t}{t^s} = \begin{cases} \log x, & \text{if } s = 1; \\ \dfrac{1}{s-1}\left(1 - \dfrac{1}{x^{s-1}}\right), & \text{if } s \neq 1. \end{cases}$$

Hence $\sum 1/n^s$ converges if and only if $s > 1$.

A series $\sum a_n$ is **absolutely convergent** if $\sum |a_n|$ converges. Absolute convergence implies convergence, by the Cauchy criterion.

Theorem 3.14 (Ratio Test). *Let*

$$r = \liminf_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| \qquad \textit{and} \qquad R = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|.$$

Then

- $R < 1 \implies \sum a_n$ *converges absolutely;*
- $1 < r \implies \sum a_n$ *diverges.*

In case $r \leqslant 1 \leqslant R$, no conclusion is possible.

Proof. Suppose $R < x < 1$. Then, for some M,

$$n \geqslant M \implies \left|\frac{a_{n+1}}{a_n}\right| < x,$$

so $n \geqslant M \implies |a_{n+1}| < |a_M| \, x^{n-M}$. By the Comparison Test, $\sum a_n$ converges.
Suppose $1 < r$. Then, for some M,

$$n \geqslant M \implies \left|\frac{a_{n+1}}{a_n}\right| > 1,$$

which means a_n cannot converge to 0.
Finally, $\lim_{n\to\infty} \dfrac{1/(n+1)}{1/n} = 1 = \dfrac{1/(n+1)^2}{1/n^2}$. □

Theorem 3.15 (Root Test). *Let $\rho = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$. Then*

- $\rho < 1 \implies \sum a_n$ *converges absolutely;*
- $1 < \rho \implies \sum a_n$ *diverges.*

In case $\rho = 1$, no conclusion is possible.

Proof. If $\rho < x < 1$, and n is large enough, then

$$\sqrt[n]{|a_n|} < x,$$

whence $|a_n| < x^n$.
If $1 < \rho$, then a_n does not converge to 0.
Finally, $\lim_{n\to\infty} n^{1/n} = \exp \lim_{n\to\infty} \log n/n = e^0 = 1$; hence $\lim_{n\to\infty} (n^2)^{1/n} = 1$. □

Theorem 3.16. *If $\sum a_n$ is absolutely convergent, and $f : \mathbb{N} \to \mathbb{N}$ is a bijection, then $\sum a_{f(n)}$ converges, and*

$$\sum_{k=0}^{\infty} \left|a_{f(k)}\right| = \sum_{k=0}^{\infty} |a_k| .$$

Proof. $\sum_{k=0}^{n} |a_{f(k)}| \leqslant \sum_{k=0}^{\infty} |a_k|$, so $\sum a_{f(n)}$ converges absolutely.

We have

$$\left|\sum_{k=0}^{n} a_{f(k)} - \sum_{k=0}^{\infty} a_k\right| \leqslant \left|\sum_{k=0}^{n} a_{f(k)} - \sum_{k=0}^{m} a_k\right| + \left|\sum_{k=m+1}^{\infty} a_k\right|.$$

Let $\varepsilon > 0$. Let m be large enough that $\left|\sum_{k=m+1}^{\infty} a_k\right| \leqslant \sum_{k=m+1}^{\infty} |a_k| \leqslant \varepsilon/2$. Then n can be large enough that

$$\{0, 1, \ldots, m\} \subseteq \{f(0), f(1), \ldots, f(n)\}.$$

Then $\left|\sum_{k=0}^{n} a_{f(k)} - \sum_{k=0}^{m} a_k\right| < \varepsilon/2$, so $\left|\sum_{k=0}^{n} a_{f(k)} - \sum_{k=0}^{\infty} a_k\right| < \varepsilon$. □

Theorem 3.17. *If $\sum a_n$ is convergent, but not absolutely convergent, and $-\infty \leqslant b \leqslant c \leqslant \infty$, then there is a bijection $f : \mathbb{N} \to \mathbb{N}$ such that*

$$\liminf_{n\to\infty} \sum_{k=0}^{n} a_{f(k)} = b \quad \textit{and} \quad \limsup_{n\to\infty} \sum_{k=0}^{n} a_{f(k)} = c.$$

Proof. For any x in $\mathbb{R}$, we have

$$x = \max(0, x) + \min(0, x) \quad \text{and} \quad |x| = \max(0, x) - \min(0, x).$$

Hence (by Theorem 3.3), if $\sum a_n$ and one of $\sum \max(0, a_n)$ and $\sum \min(0, a_n)$ converge, then $\sum a_n$ converges absolutely.

Suppose $\sum a_n$ converges, but not absolutely, and $-\infty < b \leqslant c < \infty$. Then both $\sum \max(0, a_n)$ and $\sum \min(0, a_n)$ diverge, to ∞ and $-\infty$ respectively. So it is possible to find a strictly increasing sequence $(g(n) : n \in \mathbb{N})$ and a bijection $f : \mathbb{N} \to \mathbb{N}$ such that:

- $g(0) = 0$;
- f is increasing on $\{k \in \mathbb{N} : (\exists n \in \mathbb{N})\ g(2n) \leqslant k < g(2n+1)\}$ and on its complement;
- $a_{f(k)} \geqslant 0 \iff (\exists n \in \mathbb{N})\ g(2n) \leqslant k < g(2n+1)$;
- $\sum_{k=0}^{g(2n+1)-2} a_{f(k)} \leqslant c < \sum_{k=0}^{g(2n+1)-1} a_{f(k)}$;
- $\sum_{k=0}^{g(2n+2)-1} a_{f(k)} < b \leqslant \sum_{k=0}^{g(2n+2)-2} a_{f(k)}$.

Since $\lim_{n\to\infty} a_{f(n)} = 0$, we have

$$\liminf_{n\to\infty} \sum_{k=0}^{n} a_{f(k)} = b \quad \text{and} \quad \limsup_{n\to\infty} \sum_{k=0}^{n} a_{f(k)} = c.$$

If possibly one or both of b and c are infinite, then we modify the construction of f by choosing sequences (b_n) and (c_n) that have limits b and c respectively and that satisfy $b_n < c_n$. Then we require

$$\sum_{k=0}^{g(2n+1)-2} a_{f(k)} \leqslant c_n < \sum_{k=0}^{g(2n+1)-1} a_{f(k)}$$

and $\sum_{k=0}^{g(2n+1)-1} a_{f(k)} < b_n \leqslant \sum_{k=0}^{g(2n+1)-2} a_{f(k)}$. □

Suppose $(p, q) \mapsto a_{p\,q}$ is a function from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{R}$ (that is, a **double sequence**). Then we say

$$\lim_{(p,q)\to\infty} a_{p\,q} = b$$

if for all positive ε there is M such that

$$p, q > M \implies |a_{p\,q} - b| < \varepsilon.$$

Theorem 3.18. *Suppose*

$$\lim_{(p,q)\to\infty} a_{p\,q} = b.$$

If $\lim_{q\to\infty} a_{p\,q}$ *exists for each* p *in* $\mathbb{N}$, *then*

$$\lim_{p\to\infty} \lim_{q\to\infty} a_{p\,q} = b.$$

Proof. Let N be such that

$$p, q > N \implies |a_{p\,q} - b| < \frac{\varepsilon}{2}.$$

Suppose $p > N$ and $\lim_{q\to\infty} a_{p\,q} = c_p$. There is M such that

$$q > M \implies |a_{p\,q} - c_p| < \frac{\varepsilon}{2}.$$

Hence, if $q > \max(M, N)$, then $|c_p - b| \leqslant |c_p - a_{p\,q}| + |a_{p\,q} - b| < \varepsilon$. □

Example 3.19. We have

$$\lim_{p\to\infty} \lim_{q\to\infty} \frac{pq}{p^2 + q^2} = \lim_{p\to\infty} 0 = 0;$$

but $pq/(p^2 + q^2) = 1/2$ if $p = q$; so the double sequence

$$(p, q) \mapsto \frac{pq}{p^2 + q^2}$$

has no limit.

Suppose $(a_n : n \in \mathbb{N})$ is a sequence of real numbers. We shall say that $\prod a_n$ **converges** if, for some p in $\mathbb{N}$, the sequence of products $\prod_{k=p}^{n} a_k$ converges to a *finite* and *non-zero* limit b. In this case, we write

$$\prod_{k=p}^{\infty} a_n = b \quad \text{and} \quad \prod_{k=0}^{\infty} a_k = b \cdot \prod_{k=0}^{p-1} a_k.$$

(If $p = 0$, then $\prod_{k=0}^{p-1} a_k = 1$.) The definition has two immediate consequences:

- If $\prod a_n$ converges, then $a_n \neq 0$ for all but finitely many n.
- $\prod a_n$ converges if and only if $\prod 1/a_n$ converges.

Examples 3.20. $\prod_{k=1}^{n}(1+1/k) = \prod_{k=1}^{n}((k+1)/k) = n+1$, and

$$\prod_{k=2}^{n+1}\left(1-\frac{1}{k}\right) = \prod_{k=2}^{n+1}\frac{k-1}{k} = \frac{1}{n+1};$$

so $\prod(1 \pm 1/n)$ diverges in each case.

Theorem 3.21 (Cauchy condition for products). *The product $\prod a_n$ converges if and only if, for every positive ε, there is M such that, for all n and k in $\mathbb{N}$,*

$$n > M \implies \left|\prod_{\ell=n}^{n+k} a_\ell - 1\right| < \varepsilon.$$

Proof. Suppose $\prod a_n$ converges. Then for some p there is a positive δ such that

$$\left|\prod_{\ell=p}^{p+n} a_\ell\right| > \delta$$

for all n in $\mathbb{N}$. Also, there is M such that

$$n > M \implies \left|\prod_{\ell=p}^{p+n+k+1} a_\ell - \prod_{\ell=p}^{p+n} a_\ell\right| < \varepsilon\delta$$

for all k in $\mathbb{N}$. Division yields

$$n > M \implies \left|\prod_{\ell=p+n+1}^{p+n+k+1} a_\ell - 1\right| < \varepsilon.$$

Suppose conversely that for all positive ε there is M such that

$$n > M \implies \left|\prod_{\ell=n}^{n+k} a_\ell - 1\right| < \varepsilon.$$

for all k in $\mathbb{N}$. Then, in particular, there is p such that

$$\frac{1}{2} < \prod_{\ell=p}^{p+k} a_\ell < \frac{3}{2}$$

for all k in $\mathbb{N}$. Hence if $\lim_{n\to\infty}\prod_{\ell=p}^{p+n} a_\ell$ exists, then it is not zero, so $\prod a_n$ converges. We can show that this limit exists by the Cauchy criterion. Indeed, we have

$$\left|\prod_{\ell=p}^{p+n+k+1} a_\ell - \prod_{\ell=p}^{p+n} a_\ell\right| = \left|\prod_{\ell=p+n+1}^{p+n+k+1} a_\ell - 1\right| \cdot \left|\prod_{\ell=p}^{p+n} a_\ell\right| < \frac{3}{2}\varepsilon$$

if $n > M$. □

Theorem 3.22. *If* (a_n) *is a sequence of* positive *terms, then*

$$\prod(1+a_n) \text{ converges} \iff \sum a_n \text{ converges.}$$

Proof. If $1 < x$, then, by the Mean-Value Theorem, $1 \leqslant e^t = (e^x - 1)/x$ for some t in $(1, x)$, so $1 + x \leqslant e^x$. Hence

$$\sum_{k=0}^{n} a_k < 1 + \sum_{k=0}^{n} a_k \leqslant \prod_{k=0}^{n}(1+a_k) \leqslant \prod_{k=0}^{n} \exp a_k = \exp \sum_{k=0}^{n} a_k.$$

Thus, the increasing sequence of partial sums is bounded if and only if the increasing sequence of partial products is bounded. □

The **Riemann Zeta-function** is defined on the interval $(1, \infty)$ by the formula

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

(See Examples 3.13.)

Theorem 3.23 (Euler product). *Let the set of prime numbers be the range of the increasing sequence* $(p_n : n \in \mathbb{N})$. *Then the Riemann Zeta-function is given by*

$$\zeta(s) = \prod_{k=0}^{\infty} \frac{1}{1 - p_k^{-s}} = \prod_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{1}{p_k^{ns}}.$$

Proof. Every partial product $\prod_{k=0}^{m} \frac{1}{1-p_k^{-s}}$ is the limit of the products

$$\prod_{k=0}^{m} \sum_{\ell=0}^{n} \frac{1}{p_k^{\ell s}},$$

which can be written as the sums $\sum_a 1/a^s$, where a ranges over the positive integers whose only prime factors are among the p_k such that $k \leqslant m$, and which are divisible by any one of these p_k at most n times. Hence

$$\left| \zeta(s) - \prod_{k=0}^{m} \frac{1}{1 - p_k^{-s}} \right| \leqslant \sum_b \frac{1}{b^s} \leqslant \sum_{n=p_m}^{\infty} \frac{1}{n^s},$$

where b ranges over those positive integers divisible by a prime greater than p_m. Since $\lim_{n\to\infty} \sum_{n=p_m}^{\infty} 1/n^s = 0$, the claim follows. □

4 Sequences of functions

Suppose I is a sub-interval of $\mathbb{R}$, and $(f_n : n \in \mathbb{N})$ is a sequence of functions from I to $\mathbb{R}$. If f is another such sequence, then (f_n) **converges pointwise** to f if, for all x in I,

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

Nice properties need not be preserved under the taking of pointwise limits:

Examples 4.1. (0) The sequence of continuous functions $x \mapsto x^n : [0,1] \to \mathbb{R}$ converges pointwise to the non-continuous function

$$x \mapsto \begin{cases} 0, & \text{if } 0 \leqslant x < 1; \\ 1, & \text{if } x = 1. \end{cases}$$

(1) The sequence of integrable functions $x \mapsto n^2 x^n (1-x) : [0,1] \to \mathbb{R}$ converges pointwise to 0, but

$$\int_0^1 n^2 x^n (1-x)\,\mathrm{d}\,x = n^2 \left(\frac{x^{n+1}}{n+1} - \frac{x^{n+2}}{n+2} \right) \Bigg|_0^1 = \frac{n^2}{(n+1)(n+2)},$$

so the sequence of integrals converges to 1.

(2) The sequence of differentiable functions $x \mapsto (\sin n^2 x)/n$ converges to 0, but the sequence of derivatives $x \mapsto 2n \cos nx$ does not converge.

The sequence (f_n) **converges uniformly** to f if for every positive ε there is M such that for all x in I and all n in $\mathbb{Z}$,

$$n > M \implies |f_n(x) - f(x)| < \varepsilon.$$

We showed in Math 271 that the uniform limit of continuous functions is continuous.

Let $B(I)$ be the set of bounded real-valued functions on I. This becomes a metric space when we define

$$d(f,g) = \sup\{|f(x) - g(x)| : x \in I\}.$$

Then a sequence of functions in $B(I)$ converges uniformly if and only if it converges in the metric d.

Lemma 4.2 (Cauchy condition for uniform convergence). *A sequence (f_n) of functions on I converges uniformly if and only if for all positive ε there is M such that for all x in I, all n in $\mathbb{Z}$ and all k in $\mathbb{N}$,*

$$n > M \implies |f_{n+k}(x) - f_n(x)| < \varepsilon.$$

Proof. Suppose (f_n) converges uniformly to f. If $\varepsilon > 0$, let M be such that

$$n > M \implies |f_n(x) - f(x)| < \frac{\varepsilon}{2}. \tag{iv}$$

If $n > M$, then also $|f_{n+k}(x) - f(x)| < \varepsilon/2$, so $|f_{n+k}(x) - f(x)| < \varepsilon$.

Conversely, if (f_n) satisfies the Cauchy condition, then each sequence $(f_n(x))$ of real numbers is Cauchy, so it has a limit, say $f(x)$. Hence, for all n, we have

$$\lim_{k\to\infty} |f_n(x) - f_{n+k}(x)| = |f_n(x) - f(x)|.$$

If $\varepsilon > 0$, let M be such that (iv) holds. Then

$$n > M \implies |f_n(x) - f(x)| \leqslant \frac{\varepsilon}{2} < \varepsilon.$$

So (f_n) converges uniformly to f. ∎

A *series* $\sum f_n$ of functions converges uniformly if the sequence of partial sums $\sum_{k=0}^{n} f_k$ converges uniformly.

Lemma 4.3 (Cauchy condition for uniform convergence of series). *A series $\sum f_n$ of functions converges uniformly if and only if, for all positive ε, there is M such that for all x in I, for all n in $\mathbb{Z}$ and all k in $\mathbb{N}$,*

$$n > M \implies \left|\sum_{\ell=n}^{n+k} f_\ell(x)\right| < \varepsilon.$$

Theorem 4.4 (Weierstraß M-test). *Suppose the sequences (f_n) of functions and (M_n) of numbers are such that, for each x in I,*

$$|f_n(x)| \leqslant M_n.$$

If $\sum M_n$ converges, then $\sum f_n$ converges uniformly.

Proof. If $\sum M_n$ converges, then it satisfies the Cauchy criterion for series; so $\sum f_n$ satisfies the Cauchy criterion for series of functions; so $\sum f_n$ converges uniformly. □

Theorem 4.5. *Suppose each f_n is continuous, and $\sum f_n$ converges uniformly to f. Then f is continuous; in particular,*

$$\lim_{y \to x} \sum_{n=0}^{\infty} f_n(y) = \sum_{n=0}^{\infty} \lim_{y \to x} f_n(y).$$

Theorem 4.6 (A space-filling curve). *There is a continuous surjective function from $[0, 1]$ to $[0, 1] \times [0, 1]$.*

Proof. Let ϕ be a continuous function on $\mathbb{R}$ such that, for all n in $\mathbb{Z}$,

$$\phi(x) = \begin{cases} 0, & \text{if } k - 1/6 \leqslant x \leqslant k + 1/6; \\ 1, & \text{if } k + 1/3 \leqslant x \leqslant k + 2/3. \end{cases}$$

Then ϕ is periodic, with period 1; that is, $\phi(x) = \phi(x - \lfloor x \rfloor)$. For each e in $\{0, 1\}$, let f_e be the function given by

$$f_e(x) = \sum_{k=0}^{\infty} \frac{\phi(3^{2k+e}x)}{2^{k+1}}.$$

The series converge uniformly, by the M-test with $M_k = 1/2^{k+1}$. Since ϕ is continuous, so are the f_e. Our function will be

$$x \mapsto (f_0(x), f_1(x)).$$

Any point of $[0, 1] \times [0, 1]$ can be written, in binary notation, as

$$\left(\sum_{k=1}^{\infty} \frac{a_k}{2^k}, \sum_{k=1}^{\infty} \frac{b_k}{2^k}\right),$$

where a_k and b_k are in $\{0,1\}$. Now define

$$c_n = \begin{cases} a_k, & \text{if } 2k-1=n; \\ b_k, & \text{if } 2k=n; \end{cases}$$

and let

$$c = \sum_{n=1}^{\infty} \frac{c_n}{3^n}.$$

Then

$$3^k c - \lfloor 3^k c \rfloor = \frac{c_{k+1}}{3} + \sum_{n=1}^{\infty} \frac{c_{n+k+1}}{3^{n+1}},$$

so $\phi(3^k c) = c_{k+1}$. □

Theorem 4.7. *Let g be of bounded variation on $[a,b]$. If $f_n \in R(g)$, and (f_n) converges uniformly to f on $[a,b]$, then $f \in R(g)$, and $(\int_a^x f_n \,\mathrm{d}\, g)$ converges uniformly to $\int_a^x f \,\mathrm{d}\, g$.*

Proof. We may assume g is increasing and $g(b) \neq g(a)$. There is M such that, for all x in $[a,b]$,

$$n \geqslant M \implies |f_n(x) - f(x)| < \frac{\varepsilon}{3(g(b)-g(a))}.$$

We use this first to prove $f \in R(g)$. For any partition P of $[a,b]$, we have

$$|\mathrm{U}(P, f - f_M, g)|\,, |\mathrm{L}(P, f - f_M, g)| < \frac{\varepsilon}{3}.$$

Suppose in particular that P is fine enough that

$$\mathrm{U}(P, f_M, g) - \mathrm{L}(P, f_M, g) < \frac{\varepsilon}{3}.$$

Because $\sup\{f(x) : x \in [a,b]\} \leqslant \sup\{f(x) - f_M(x) : x \in [a,b]\} + \sup\{f_M(x) : x \in [a,b]\}$, and likewise for the infima, we have

$$\begin{aligned} &\mathrm{U}(P,f,g) - \mathrm{L}(P,f,g) \leqslant \\ &\qquad \mathrm{U}(P, f-f_M, g) + \mathrm{U}(P, f_M, g) - (\mathrm{L}(P, f - f_M, g) + \mathrm{L}(P, f_M, g)) \leqslant \varepsilon. \end{aligned}$$

So Riemann's condition is satisfied, and $f \in R(g)$. We have also

$$\begin{aligned} \left| \int_a^x f_n \,\mathrm{d}\, g - \int_a^x f \,\mathrm{d}\, g \right| = \left| \int_a^x (f_n - f) \,\mathrm{d}\, g \right| \leqslant \\ \int_a^x |f_n - f| \,\mathrm{d}\, g \leqslant \int_a^b |f_n - f| \,\mathrm{d}\, g \leqslant \frac{\varepsilon}{3}, \end{aligned}$$

so the convergence of the sequence of integrals is uniform. □

Under the conditions of the theorem, we have

$$\lim_{n\to\infty} \int_a^x f_n \,\mathrm{d}\, g = \int_a^x (\lim_{n\to\infty} f_n) \,\mathrm{d}\, g.$$

If also $\sum f_n$ converges uniformly, then

$$\sum_{n=0}^{\infty} \int_a^x f_n \, \mathrm{d}\, g = \int_a^x (\sum_{n=0}^{\infty} f_n) \, \mathrm{d}\, g.$$

A **power series** is a function

$$z \mapsto \sum_{n=0}^{\infty} c_n (z-a)^n.$$

At a, this function has the value c_0, by definition. The function may, but need not, be well-defined at other points. Nice results are obtained if the coefficients c_n and the point a (as well as the argument z) are allowed to be *complex numbers.*

Theorem 4.8. *The vector-space $\mathbb{R}^2$ becomes a field when equipped with the multiplication given by*

$$(a,b)(c,d) = (ac - bd, ad + bc).$$

In this field, the multiplicative identity is $(1,0)$, and

$$(a,b)^{-1} = \frac{1}{a^2+b^2}(a,-b).$$

The field in the theorem is denoted $\mathbb{C}$; its elements are **complex numbers**. For the complex numbers $(1,0)$ and $(0,1)$, we may write

$$1 \quad \text{and} \quad i$$

respectively; then the complex number (a,b)—that is, $a(1,0)+b(0,1)$—can be written $a+bi$. Note that $i^2 = -1$.

Theorem 4.9. *If $z, w \in \mathbb{C}$, then $|zw| = |z|\,|w|$.*

If $z \in \mathbb{C}$, then $|z|$ is called the **absolute value** or **modulus** of z. We have the triangle inequality, $|z+w| \leqslant |z| + |w|$. Much of what we have done with sequences of *real* numbers applies to *complex* numbers as well. In particular,

- the *definition* of absolute convergence is meaningful for series of complex numbers;
- the *tests* for absolute convergence work for such series (since absolute convergence is still convergence of a certain series of real numbers);
- the *proof* that absolute convergence implies convergence is still valid for such series.

Let $\rho = \limsup_{n\to\infty} \sqrt[n]{|c_n|}$. Then

$$\limsup_{n\to\infty} \sqrt[n]{|c_n(z-a)^n|} = \limsup_{n\to\infty} \sqrt[n]{|c_n|\,|(z-a)^n|} =$$
$$\limsup_{n\to\infty}(|z-a| \sqrt[n]{|c_n|}) = |z-a|\,\rho.$$

If $\rho > 0$, then the series $\sum c_n(z-a)^n$:

- converges absolutely, if $|z-a| < 1/\rho$;
- diverges, if $|z-a| > 1/\rho$.

The number $1/\rho$ (if it exists) is the **radius of convergence** of the series; the ball $B(a; 1/\rho)$ is the **disc of convergence**. (If $\rho = 0$, then the radius of convergence is ∞; if $\rho = \infty$, then the radius of convergence is 0.)

Theorem 4.10. *Every power series is continuous on its disc of convergence.*

Proof. Let R be the radius of convergence of $\sum c_n(z-a)^n$, and suppose $w \in B(a; R)$. Let $\delta = (R - |w-a|)/2$, and let $F = \{z \in \mathbb{C} : |z-w| \leqslant \delta\}$. Then F is a neighborhood of w, so it is enough to show that the series is continuous on F. For *this*, since $z \mapsto c_n(z-a)^n$ is continous, it is enough to show that the series converges uniformly on F. But F is compact, so some point b of F is furthest from a; that is, if $z \in F$, then

$$|c_n(z-a)^n| = |c_n|\,|z-a|^n \leqslant |c_n|\,|b-a|^n\,.$$

Since $|b-a| < R$, we know $\sum |c_n|\,|b-a|^n$ converges; hence, by the M-test, the series converges uniformly on F. □

Since $\mathbb{C}$ is a field, the definition of the derivative of a function from $\mathbb{R}$ to $\mathbb{R}$ is meaningful for functions from $\mathbb{C}$ to $\mathbb{C}$. We shall show that any power series is *differentiable* on its disc of convergence, with derivative the sum of the derivatives of the terms.

Lemma 4.11. $\sum_{n=0}^{\infty} c_n(z-a)^n$ *and* $\sum_{n=1}^{\infty} nc_n(z-a)^{n-1}$ *have the same radius of convergence.*

Proof. The second series is convergent if and only if $\sum_{n=1}^{\infty} nc_n(z-a)^n$ is convergent. Since $\lim_{n\to\infty} \sqrt[n]{n} = 1$, this series has the same radius of convergence as the first. □

Lemma 4.12. $|z^n - w^n| \leqslant n\,|z-w|\,(|w| + |z-w|)^{n-1}$ *if* $n > 0$.

Proof. The claim is trivial if $w = 0$. Assume $w \neq 0$, and let $t = (z-w)/w$. Then we have to prove

$$|(1+t)^n - 1| \leqslant n\,|t|\,(1+|t|)^{n-1}.$$

If t is a non-negative real number, then the inequality holds, since then

$$(1+t)^n - 1 = n\int_0^t (1+x)^{n-1}\,\mathrm{d}\,x \leqslant nt(1+t)^{n-1}.$$

For arbitrary t in $\mathbb{C}$, we have

$$|(1+t)^n - 1| = \left|\sum_{k=1}^{n} \binom{n}{k} t^k\right| \leqslant \sum_{k=1}^{n} \binom{n}{k} |t|^k = (1+|t|)^n - 1,$$

which, combined with the previous inequality, yields the claim. □

Note that, by interchanging z and w, we can write the inequality of the lemma as

$$|z^n - w^n| \leqslant n\,|z-w|\,(|z|+|z-w|)^{n-1}.$$

Theorem 4.13. *If $\sum_{n=0}^{\infty} c_n(z-a)^n$ has a positive radius of convergence, then the function is differentiable on its disc of convergence, and its derivative is $\sum_{n=1}^{\infty} nc_n(z-a)^{n-1}$.*

Proof. Write $f(z)$ for $\sum_{n=0}^{\infty} c_n(z-a)^n$. Let ζ be an element of the disc of convergence; we shall compute $f'(\zeta)$. For all elements w of this disc distinct from ζ,

$$\frac{f(\zeta)-f(w)}{\zeta-w} = \sum_{n=1}^{\infty} c_n \frac{(\zeta-a)^n-(w-a)^n}{\zeta-w}.$$

We want to make the difference between this and $\sum_{n=1}^{\infty} nc_n(\zeta-a)^{n-1}$ small. To do this, we break this difference into two pieces, and make each piece small. For any positive integer M,

$$\begin{aligned}
&\left|\frac{f(\zeta)-f(w)}{\zeta-w} - \sum_{n=1}^{\infty} nc_n(\zeta-a)^{n-1}\right| \leqslant \\
&\quad \sum_{n=1}^{M} |c_n| \left|\frac{(\zeta-a)^n-(w-a)^n}{\zeta-w} - n(\zeta-a)^{n-1}\right| + \\
&\quad \sum_{n=M+1}^{\infty} c_n \left(\left|\frac{(\zeta-a)^n-(w-a)^n}{\zeta-w}\right| + n\,|\zeta-a|^{n-1}\right) \leqslant \\
&\quad \sum_{n=1}^{M} |c_n| \left|\frac{(\zeta-a)^n-(w-a)^n}{\zeta-w} - n(\zeta-a)^{n-1}\right| + \\
&\qquad\qquad \sum_{n=M+1}^{\infty} n\,|c_n| \left((|\zeta-a|+|\zeta-w|)^{n-1} + |\zeta-a|^{n-1}\right)
\end{aligned}$$

by Lemma 4.12. Let R be the radius of convergence of f, and let $\delta_0 = (R - |\zeta-a|)/2$. Suppose $0 < |\zeta-w| < \delta_0$. Then

$$\begin{aligned}
&\left|\frac{f(\zeta)-f(w)}{\zeta-w} - \sum_{n=1}^{\infty} nc_n(\zeta-a)^{n-1}\right| \leqslant \\
&\quad \sum_{n=1}^{M} |c_n| \left|\frac{(\zeta-a)^n-(w-a)^n}{\zeta-w} - n(\zeta-a)^{n-1}\right| + \\
&\qquad\qquad 2\sum_{n=M+1}^{\infty} nc_n(|\zeta-a|+\delta_0)^{n-1}.
\end{aligned}$$

Say $\varepsilon > 0$. By Lemma 4.11, we can choose M large enough that

$$2\sum_{n=M+1}^{\infty} nc_n(|\zeta-a|+\delta_0)^{n-1} < \varepsilon/2.$$

But for each positive n, the derivative of $z \mapsto (z-a)^n$ is $z \mapsto n(z-a)^{n-1}$; hence, for the chosen M, we can find δ_1 such that, if $0 < |\zeta - w| < \delta_1$, then

$$\sum_{n=1}^{M} |c_n| \left| \frac{(\zeta-a)^n - (w-a)^n}{\zeta - w} - n(\zeta-a)^{n-1} \right| < \frac{\varepsilon}{2}.$$

Therefore, if $0 < |\zeta - w| < \min\{\delta_0, \delta_1\}$, then

$$\left| \frac{f(\zeta) - f(w)}{\zeta - w} - \sum_{n=1}^{\infty} n c_n (\zeta - a)^{n-1} \right| \leqslant \varepsilon;$$

this proves the claim. □

Corollary 4.14. *If $f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n$, and R is the radius of convergence of the series, then f is infinitely differentiable on $B(a;R)$, and*

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

Proof. Repeated application of the theorem yields

$$f^{(k)}(z) = \sum_{n=k}^{\infty} n(n-1)\dots(n-k+1) c_n (z-a)^{n-k} = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} c_n (z-a)^{n-k},$$

hence $f^{(k)}(a) = k! c_k$. □

For which functions f, defined on a neighborhood of a, can we conclude

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

on some neighborhood of a? If $a \in \mathbb{C}$, then it is enough for f to be once-differentiable (but we won't prove this). If $a \in \mathbb{R}$, then it is not even enough to know that f is infinitely differentiable:

Example 4.15. Let $f(x) = \begin{cases} \exp(-x^{-2}), & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$ Then $f^{(n)}(0) = 0$ for all n; so Equation (v) fails when $a = 0$.

We can come up with a condition under which a function f satisfies Equation (v):

Lemma 4.16 (Generalized Mean-Value Theorem). *Let F and G be real-valued functions continuous on $[a,b]$ and differentiable on (a,b). Then*

$$F'(c)[G(b) - G(a)] = G'(c)[F(b) - F(a)]$$

for some c in (a,b).

Proof. Apply Rolle's Theorem to $x \mapsto F(x)[G(b) - G(a)] - G(x)[F(b) - F(a)]$. □

Theorem 4.17. *Suppose $f^{(n-1)}$ and $g^{(n-1)}$ are continuous on $[a,b]$ and differentiable on (a,b). Then for any distinct c and x in $[a,b]$ there is y, strictly between them, such that*

$$f^{(n)}(y)\left[g(x)-\sum_{k=0}^{n-1}\frac{g^{(k)}(c)}{k!}(x-c)^k\right]=g^{(n)}(y)\left[f(x)-\sum_{k=0}^{n-1}\frac{f^{(k)}(c)}{k!}(x-c)^k\right].$$

Proof. Use the Generalized Mean-Value Theorem, with $[c,x]$ (or $[x,c]$) in place of $[a,b]$, with F as the function

$$t\mapsto\sum_{k=0}^{n-1}\frac{f^{(k)}(t)}{k!}(x-t)^k,$$

and with G as a similar function in terms of g. In particular, $G(x)=g(x)$, so

$$G(x)-G(c)=g(x)-\sum_{k=0}^{n-1}\frac{g^{(k)}(c)}{k!}(x-c)^k;$$

also,

$$F'(t)=f'(t)+\sum_{k=1}^{n-1}\left[\frac{f^{(k+1)}(t)}{k!}(x-t)^k-\frac{f^{(k)}(t)}{(k-1)!}(x-t)^{k-1}\right],$$

a telescoping sum, so $F'(t)=f^{(n)}(t)(x-t)^{n-1}/(n-1)!$. □

Corollary 4.18 (Taylor's Theorem). *Suppose $f^{(n-1)}$ is continuous on $[a,b]$ and differentiable on (a,b). Then for any distinct c and x in $[a,b]$ there is y, strictly between them, such that*

$$f(x)=\sum_{k=0}^{n-1}\frac{f^{(k)}(c)}{k!}(x-c)^k+\frac{f^{(n)}(y)}{k!}(x-c)^n.$$

Proof. In the theorem, let g be $t\mapsto(t-c)^n$. We have

$$g^{(k)}(t)=\frac{n!}{(n-k)!}(t-c)^{n-k},$$

so the theorem gives

$$f^{(n)}(y)(x-c)^n=n!\left[f(x)-\sum_{k=0}^{n-1}\frac{f^{(k)}(c)}{k!}(x-c)^k\right],$$

whence the claim. □

Lemma 4.19. *Suppose f has derivatives of all orders on some neighborhood of c, and for some M, each $f^{(n)}$ is bounded by M^n on that neighborhood; then on that neighborhood, Equation (v) holds.*

Proof. For any positive a, we have $\lim_{n\to\infty}(a^n/n!)=0$, since $\sum(a^n/n!)$ converges by the ratio test. □

5 Differentiation in several dimensions

Suppose in this section that f is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, and $\mathbf{a}$ is a point of $\mathbb{R}^n$. We want to investigate the possibilities for differentiating f at $\mathbf{a}$. We can write f as $(f_0, \ldots, f_{m-1})$, where each f_i is a function from $\mathbb{R}^n$ to $\mathbb{R}$; so if $n = 1$, then we can form $(f'_0, \ldots, f'_{m-1})$.

In the general case, we can take *partial derivatives*: Define a function $\delta : \mathbb{N}\times\mathbb{N} \to \{0, 1\}$ by

$$\delta_{i\,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$$

Let $\mathbf{e}_i$ be the element $(\delta_{i\,0}, \ldots, \delta_{i\,n-1})$ of $\mathbb{R}^n$. Let $D_i f$ be the function defined by

$$D_i f(\mathbf{a}) = \lim_{h\to 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h};$$

this function (where it is well-defined) is the **partial derivative of f with respect to the ith coordinate**. Note that $D_i f_j(\mathbf{a})$ is the ordinary derivative at 0 of the function

$$x \mapsto f_j(\mathbf{a} + x\mathbf{e}_i).$$

Example 5.1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

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

Then $D_0 f(0, 0) = 0 = D_1 f(0, 0)$, but f is not continuous at $(0, 0)$.

For a stronger property than having partial derivatives, we define the generalization called the **directional derivative**: If $\mathbf{u} \in \mathbb{R}^n$, then

$$f'(\mathbf{a}; \mathbf{u}) = \lim_{h\to 0} \frac{f(\mathbf{a} + h\mathbf{u}) - f(\mathbf{a})}{h}.$$

Thus $f'(\mathbf{a}; \mathbf{e}_i) = D_i f(\mathbf{a})$. In Example 5.1, not all directional derivatives at $(0, 0)$ exist.

Example 5.2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = \begin{cases} xy^2/(x^2 + y^4), & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$$

Then all directional derivatives at 0 exist, since

$$\frac{f(hu, hv) - f(0, 0)}{h} = \frac{uv^2}{u^2 + h^2 v^4},$$

which has a limit as h goes to 0, whether or not $u = 0$. However, $f(y^2, y) = 1/2$, unless $y = 0$; so f is not continuous at $(0, 0)$.

If $n = 1$, then f is differentiable at a if and only if there is a number, called $f'(a)$, such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - f'(a)\cdot h}{h} = 0.$$

The function $h \mapsto f'(a)\cdot h : \mathbb{R} \to \mathbb{R}$ is *linear*; call it λ. (A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is **linear** if $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$ in all cases.) Then the function $x \mapsto f(a) + \lambda(x - a)$ is an approximation to f near a.

The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **differentiable at a** if there is a linear function $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})|}{|\mathbf{h}|} = 0.$$

In this case, λ is the **total derivative of f at a** and is denoted $Df(\mathbf{a})$. Note well that this is a *function*; however, we have to check that it is unique:

Theorem 5.3. *If $Df(\mathbf{a})$ exists, then $Df(\mathbf{a})(\mathbf{u}) = f'(\mathbf{a};\mathbf{u})$ for all $\mathbf{u}$ in $\mathbb{R}^n$; so the total derivative is uniquely determined by the directional derivatives.*

Proof. If $Df(\mathbf{a})$ exists, and $\mathbf{u} \neq 0$, then

$$0 = \lim_{h\to 0} \frac{|f(\mathbf{a}+h\mathbf{u}) - f(\mathbf{a}) - Df(\mathbf{a})(h\mathbf{u})|}{|h\mathbf{u}|} = 0;$$

but $Df(\mathbf{a})(h\mathbf{u}) = hDf(\mathbf{a})(\mathbf{u})$; hence the claim. □

Corollary 5.4. $Df(\mathbf{a})(\mathbf{u}) = \sum_{i<n} u_i D_i f(\mathbf{a})$ *for all* $\mathbf{u}$.

Proof. It's true when $\mathbf{u} = \mathbf{e}_i$; hence it is true in general, by linearity. □

Lemma 5.5. *For any linear function $T : \mathbb{R}^n \to \mathbb{R}^m$ there is a number M such that*

$$|T(\mathbf{h})| \leqslant M\,|\mathbf{h}|$$

for all $\mathbf{h}$ in $\mathbb{R}^n$. Hence T is continuous at $\mathbf{0}$.

Proof. $|T(\mathbf{x})| = \left|\sum_{i<n} x_i T(\mathbf{e}_i)\right| \leqslant \sum_{i<n} |x_i|\,|T(\mathbf{e}_i)| \leqslant |\mathbf{x}| \sum_{i<n} |T(\mathbf{e}_i)|$. □

Theorem 5.6. *Differentiability implies continuity.*

Proof. If $Df(\mathbf{a})$ exists, then

$$\lim_{\mathbf{h}\to\mathbf{0}} (f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{h})) = 0;$$

but also $\lim_{\mathbf{h}\to\mathbf{0}} Df(\mathbf{a})(\mathbf{h}) = 0$, by Lemma 5.5; hence

$$\lim_{\mathbf{h}\to\mathbf{0}} f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}),$$

whence the claim. □

The sum of products in Corollary 5.4 can be written as a matrix product:

$$Df(\mathbf{a})(\mathbf{u}) = \begin{bmatrix} D_0 f(\mathbf{a}) & \cdots & D_{n-1}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_{n-1}. \end{bmatrix}$$

As $\mathbf{u}$, so can each partial derivative $D_i f(\mathbf{a})$ be written as a column-vector. Thus $Df(\mathbf{a})(\mathbf{u})$ is the matrix product $\mathbf{u} \cdot f'(\mathbf{a})$, where

$$f'(\mathbf{a}) = \begin{bmatrix} D_0 f_0(\mathbf{a}) & \cdots & D_{n-1} f_0(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ D_0 f_{m-1}(\mathbf{a}) & \cdots & D_{n-1} f_{m-1}(\mathbf{a}) \end{bmatrix}.$$

This matrix $f'(\mathbf{a})$ is the **Jacobian matrix of f at $\mathbf{a}$**.

Theorem 5.7 (Chain-Rule). *If the functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ are differentiable at $\mathbf{a}$ and $f(\mathbf{a})$ respectively, then $g \circ f$ is differentiable at $\mathbf{a}$, and*

$$D(g \circ f)(\mathbf{a}) = Dg(f(\mathbf{a})) \circ Df(\mathbf{a}),$$

whence $(g \circ f)'(\mathbf{a}) = g'(f(\mathbf{a})) \cdot f'(\mathbf{a})$.

Proof. We have to prove

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{g(f(\mathbf{a} + \mathbf{h})) - g(f(\mathbf{a})) - Dg(f(\mathbf{a}))(Df(\mathbf{a})(\mathbf{h}))}{|\mathbf{h}|} = 0.$$

The numerator of the fraction can be written as the sum of

$$g(f(\mathbf{a} + \mathbf{h})) - g(f(\mathbf{a})) - Dg(f(\mathbf{a}))(f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}))$$

and

$$Dg(f(\mathbf{a}))(f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{h})).$$

Write these summands respectively as

$$\psi(f(\mathbf{a} + \mathbf{h})) \quad \text{and} \quad Dg(f(\mathbf{a}))(\phi(\mathbf{h})).$$

We shall show:

(0) $\displaystyle\lim_{\mathbf{h} \to \mathbf{0}} \frac{|\psi(f(\mathbf{a} + \mathbf{h}))|}{|\mathbf{h}|} = 0;$

(1) $\displaystyle\lim_{\mathbf{h} \to \mathbf{0}} \frac{|Dg(f(\mathbf{a}))(\phi(\mathbf{h}))|}{|\mathbf{h}|} = 0.$

By Lemma 5.5, there is M such that, for all $\mathbf{h}$,

$$|Dg(f(\mathbf{a}))(\phi(\mathbf{h}))| \leqslant M \, |\phi(\mathbf{h})| \, ;$$

since $\lim_{\mathbf{h} \to \mathbf{0}} |\phi(\mathbf{h})| \, / \, |\mathbf{h}| = 0$, this proves (1). For (0), note that

$$\lim_{\mathbf{k} \to \mathbf{0}} \frac{|\psi(f(\mathbf{a}) + \mathbf{k})|}{|\mathbf{k}|} = 0;$$

also, if $f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a})$, then $\psi(f(\mathbf{a}+\mathbf{h})) = \mathbf{0}$. If $f(\mathbf{a}+\mathbf{h}) \neq f(\mathbf{a})$, then we have

$$\frac{|\psi(f(\mathbf{a}+\mathbf{h}))|}{|\mathbf{h}|} = \frac{|\psi(f(\mathbf{a}+\mathbf{h}))|}{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})|} \cdot \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})|}{|\mathbf{h}|}.$$

The second factor here is

$$\frac{|\phi(\mathbf{h}) + Df(\mathbf{a})(\mathbf{h})|}{\mathbf{h}};$$

hence there is N such that

$$\frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})|}{|\mathbf{h}|} \leqslant \frac{|\phi(\mathbf{h})|}{\mathbf{h}} + N.$$

Suppose $\varepsilon > 0$. There are positive numbers δ_0, δ_1 and δ_2 such that

$$\begin{aligned}
0 < |\mathbf{h}| < \delta_0 &\implies \frac{|\phi(\mathbf{h})|}{|\mathbf{h}|} < 1;\\
0 < |\mathbf{k}| < \delta_1 &\implies \frac{|\psi(f(\mathbf{a})+\mathbf{k})|}{|\mathbf{k}|} < \frac{\varepsilon}{N+1};\\
0 < |\mathbf{h}| < \delta_2 &\implies |f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})| < \delta_2.
\end{aligned}$$

Let $\delta = \min\{\delta_0, \delta_2\}$; then

$$\frac{|\psi(f(\mathbf{a}+\mathbf{h}))|}{|\mathbf{h}|} < \varepsilon$$

whenever $0 < |\mathbf{h}| < \delta$. □

Lemma 5.8. *If the partial derivatives $D_i f$ are defined on a neighborhood of $\mathbf{a}$ and are continuous at $\mathbf{a}$, then f is differentiable at $\mathbf{a}$, that is, the total derivative $Df(\mathbf{a})$ exists.*

Proof. We may assume $m = 1$. The $D_i f$ must be defined on an open *interval* I that contains $\mathbf{a}$. (Note then that I is $I_0 \times \cdots \times I_{n-1}$ for some open intervals I_i of $\mathbb{R}$ such that $a_i \in I_i$.) We shall show that $Df(\mathbf{a})$ exists by showing that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\left|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{i<n} D_i f(\mathbf{a}) \cdot h_i\right|}{|\mathbf{h}|} = 0.$$

Suppose $\mathbf{h} \in I \smallsetminus \{\mathbf{a}\}$. We have

$$\begin{aligned}
&f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\\
&\quad = f(\mathbf{a}+h_0\mathbf{e}_0) - f(\mathbf{a})\\
&\quad + f(\mathbf{a}+h_0\mathbf{e}_0 + h_1\mathbf{e}_1) - f(\mathbf{a}+h_0\mathbf{e}_0)\\
&\quad + \cdots\\
&\quad + f(\mathbf{a}+h_0\mathbf{e}_0 + \cdots + h_{n-1}\mathbf{e}_{n-1}) - f(\mathbf{a}+h_0\mathbf{e}_0 + \cdots + h_{n-2}\mathbf{e}_{n-2})\\
&\qquad = \sum_{i<n}(f(\mathbf{a}+\sum_{j<i+1} h_j\mathbf{e}_j) - f(\mathbf{a}+\sum_{j<i} h_j\mathbf{e}_j)) = \sum_{i<n} D_i f(\mathbf{b}_i) \cdot h_i,
\end{aligned}$$

where $\mathbf{b}_i = \mathbf{a} + \sum_{j<i} h_j\mathbf{e}_j + c_i\mathbf{e}_i$ for some c_i between 0 and h_i, by the Mean-Value Theorem as applied to the function

$$x \mapsto f(\mathbf{a}+\sum_{i<j} h_j\mathbf{e}_j + x\mathbf{e}_i).$$

In particular, $\mathbf{b}_i \in I$. We now have

$$\begin{aligned}
0 &\leqslant \frac{\left|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{i<n} D_i f(\mathbf{a}) \cdot h_i\right|}{|\mathbf{h}|} \\
&= \frac{\left|\sum_{i<n} (D_i f(\mathbf{b}_i) - D_i f(\mathbf{a})) \cdot h_i\right|}{|\mathbf{h}|} \\
&\leqslant \sum_{i<n} \frac{|D_i f(\mathbf{b}_i) - D_i f(\mathbf{a})| \cdot |h_i|}{|\mathbf{h}|} \\
&\leqslant \sum_{i<n} |D_i f(\mathbf{b}_i) - D_i f(\mathbf{a})|.
\end{aligned}$$

If $\varepsilon > 0$, then, by the continuity of the $D_i f$ at $\mathbf{a}$, there is a ball $B(\mathbf{a};\delta)$ around $\mathbf{a}$ such that, if $\mathbf{x}$ is in this ball, then

$$|D_i f(\mathbf{x}) - D_i f(\mathbf{a})| \leqslant \frac{\varepsilon}{n}.$$

If $\mathbf{h}$ is $B(\mathbf{a};\delta)$, then so are the $\mathbf{b}_i$, which means

$$\frac{\left|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{i<n} D_i f(\mathbf{a}) \cdot h_i\right|}{|\mathbf{h}|} < \varepsilon.$$

The claim now follows. □

The sufficient condition for differentiability of f at $\mathbf{a}$ given by the theorem is that f be **continuously differentiable at a**.

To the Chain-Rule, we can now state a:

Corollary 5.9. *If f is continuously differentiable at $\mathbf{a}$, and the function $g : \mathbb{R}^m \to \mathbb{R}$ is differentiable at $f(\mathbf{a})$, then $g \circ f$ is differentiable at $\mathbf{a}$, and*

$$D_i(g \circ f)(\mathbf{a}) = \sum_{j<m} D_j g(f(\mathbf{a})) \cdot D_i f_j(\mathbf{a})$$

for each i less than n.

Proof. By the Chain-Rule,

$$\begin{aligned}
&\begin{bmatrix} D_0(g \circ f)(\mathbf{a}) & \cdots & D_{n-1}(g \circ f)(\mathbf{a}) \end{bmatrix} = (g \circ f)'(\mathbf{a}) = g'(f(\mathbf{a})) \cdot f'(\mathbf{a}) \\
&= \begin{bmatrix} D_0 g(f(\mathbf{a})) & \cdots & D_{m-1} g(f(\mathbf{a})) \end{bmatrix} \cdot \begin{bmatrix} D_0 f_0(\mathbf{a}) & \cdots & D_{n-1} f_0(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ D_0 f_{m-1}(\mathbf{a}) & \cdots & D_{n-1} f_{m-1}(\mathbf{a}) \end{bmatrix}.
\end{aligned}$$

Consideration of the i-th entry of the first and last members of the equation yields the result. □

Theorem 5.10 (Inverse Function). *Suppose $m = n$, and f is continuously differentiable on a neighborhood of $\mathbf{a}$, and $Df(\mathbf{a})$ is invertible, that is, $\det(f'(\mathbf{a})) \neq 0$. Then there are neighborhoods V of $\mathbf{a}$ and W of $f(\mathbf{a})$ such that f is a bijection from V to W with a differentiable inverse f^{-1} satisfying*

$$D(f^{-1})(f(\mathbf{x})) = Df(\mathbf{x})^{-1} \tag{vi}$$

for all $\mathbf{x}$ *in* V*, equivalently,*

$$(f^{-1})'(f(\mathbf{x})) = (f'(\mathbf{x}))^{-1}.$$

Proof. (0) For any function $f : \mathbb{R}^n \to \mathbb{R}^n$, let $\Phi(f)$ be the statement,

> If f is continuously differentiable near $\mathbf{a}$, then $\mathbf{a}$ and $f(\mathbf{a})$ have neighborhoods V and W respectively such that f is a bijection from V to W with differentiable inverse.

Now, any linear function on $\mathbb{R}^n$ is its own derivative everywhere. In particular, let id be the identity on $\mathbb{R}^n$. Then

$$D(\mathsf{id})(\mathbf{x}) = \mathsf{id}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$. If $\Phi(f)$, then $f^{-1} \circ f = \mathsf{id}$ on V, so

$$Df^{-1}(f(\mathbf{x})) \circ Df(\mathbf{x}) = \mathsf{id},$$

whence (vi). It remains for us to show simply that $\Phi(f)$ whenever $Df(\mathbf{a})$ is invertible.

(1) If $\Phi(f)$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible and linear, then $\Phi(T \circ f)$. In particular, if $Df(\mathbf{a})$ is invertible, and $\Phi(Df(\mathbf{a})^{-1} \circ f)$, then $\Phi(f)$. But

$$D(Df(\mathbf{a})^{-1} \circ f)(\mathbf{a}) = \mathsf{id},$$

by the Chain-Rule. So it is enough for us to show that $\Phi(f)$ whenever

$$Df(\mathbf{a}) = \mathsf{id}. \tag{vii}$$

(2) Let us now assume that f is continuously differentiable near $\mathbf{a}$, and that (vii) holds. Then

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

But we also have

$$\frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \mathbf{h}|}{|\mathbf{h}|} \geqslant \left| \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})|}{|\mathbf{h}|} - 1 \right|.$$

Therefore

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})|}{|\mathbf{h}|} = 1,$$

which means in particular that, $\mathbf{a}$ has a neighborhood U such that

$$f(\mathbf{x}) = f(\mathbf{a}) \implies \mathbf{x} = \mathbf{a} \tag{viii}$$

for all $\mathbf{x}$ in U.

(3) We can make certain additional assumptions about U. We may assume that U is a closed ball,

$$U = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| \leqslant r\}.$$

Since f is continuously differentiable near $\mathbf{a}$, the function $\mathbf{x} \mapsto \det(f'(\mathbf{x}))$ is continuous near $\mathbf{a}$, and therefore we may assume that $Df(\mathbf{x})$ is invertible for all $\mathbf{x}$ in U; we may also assume

$$|D_i f_j(\mathbf{x}) - D_i f_j(\mathbf{a})| \leqslant \frac{1}{2n^2} \tag{ix}$$

for each i and j.

By (vii), $D_i f_j(\mathbf{a}) = \delta_{i\,j}$. Let g be the function $\mathbf{x} \mapsto f(\mathbf{x}) - \mathbf{x}$; then

$$D_i g_j(\mathbf{a}) = D_i f_j(\mathbf{x}) - \delta_{i\,j} = D_i f_j(\mathbf{x}) - D_i f_j(\mathbf{a}). \tag{x}$$

For all $\mathbf{x}$ and $\mathbf{u}$ in U, we have

$$\begin{aligned}
|\mathbf{u}-\mathbf{x}| - |f(\mathbf{u}) - f(\mathbf{x})| &\leqslant |g(\mathbf{u}) - g(\mathbf{x})| \leqslant \sum_{i<n} |g_i(\mathbf{u}) - g_i(\mathbf{x})| \\
&\leqslant \sum_{i<n}\sum_{j<n} \left| g(\sum_{k=0}^{j} u_k\mathbf{e}_k + \sum_{k=j+1}^{n-1} x_k\mathbf{e}_k) - g(\sum_{k=0}^{j-1} u_k\mathbf{e}_k + \sum_{k=j}^{n-1} x_k\mathbf{e}_k) \right| \\
&\leqslant \sum_{i<n}\sum_{j<n} \frac{1}{2n^2} \cdot |u_j - x_j| \leqslant \sum_{i,j<n} \frac{1}{2n^2} \cdot |\mathbf{u}-\mathbf{x}| \leqslant \frac{1}{2} \cdot |\mathbf{u}-\mathbf{x}|
\end{aligned}$$

by the Mean-Value Theorem, (ix) and (x), and therefore

$$|\mathbf{u}-\mathbf{x}| \leqslant 2 \cdot |f(\mathbf{u}) - f(\mathbf{x})|\,. \tag{xi}$$

Hence f is injective on U, and f^{-1} is well-defined and continuous on $f(U)$.

It remains to prove that $f(U)$ is in fact a neighborhood of $f(\mathbf{a})$ and that f^{-1} is differentiable on some neighborhood of $f(\mathbf{a})$.

(4) The **boundary** of U can be denoted ∂U and is the compact set

$$\{\mathbf{x} \in U : |\mathbf{x} - \mathbf{a}| = r\}.$$

Note then that

$$U \smallsetminus \partial U = B(\mathbf{a}; r).$$

The continuous function $\mathbf{x} \to |f(\mathbf{x}) - f(\mathbf{a})| : \partial U \to \mathbb{R}$ must attain a minimum value d, which must be positive, by (viii). Now define

$$W = B(f(\mathbf{a}); d/2),$$

the open ball with center $f(\mathbf{a})$ and radius $d/2$. If $\mathbf{y} \in W$, and $\mathbf{u} \in \partial U$, then

$$d \leqslant |f(\mathbf{u}) - f(\mathbf{a})| \leqslant |f(\mathbf{u}) - \mathbf{y}| + |\mathbf{y} - f(\mathbf{a})|\,;$$

since also $|\mathbf{y} - f(\mathbf{a})| < d/2$, we have

$$|\mathbf{y} - f(\mathbf{a})| < |f(\mathbf{u}) - \mathbf{y}|\,. \tag{xii}$$

(5) We shall now show that

$$W \subseteq f(U \smallsetminus \partial U).$$

Then we can define
$$V = (U \smallsetminus \partial U) \cap f^{-1}(W),$$
so that V is an open neighborhood of $\mathbf{a}$, and f is a continuous bijection between this and the open neighborhood W of $f(\mathbf{a})$.

Let $\mathbf{y} \in W$; we shall find $\mathbf{x}$ in $U \smallsetminus \partial U$ such that $f(\mathbf{x}) = \mathbf{y}$. Let h be the continuously differentiable function
$$\mathbf{u} \mapsto |\mathbf{y} - f(\mathbf{u})|^2 : U \to \mathbb{R}.$$
By (xii), we have
$$\mathbf{u} \in \partial U \implies h(\mathbf{a}) < h(\mathbf{u}).$$
Since U is compact, h attains a minimum at some $\mathbf{x}$, which must therefore be in $U \smallsetminus \partial U$. Then $Dh(\mathbf{x})$ is the zero-function. Now, h is the composition of $\mathbf{u} \mapsto \mathbf{y} - f(\mathbf{u})$ and $\mathbf{z} \mapsto \sum_{i<n} z_i^2$. By the Chain-Rule, we find
$$2\begin{bmatrix} y_0 - f_0(\mathbf{x}) & \cdots & y_{n-1} - f_{n-1}(\mathbf{x}) \end{bmatrix} \cdot f'(\mathbf{x}) = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}.$$
The invertibility of $f'(\mathbf{x})$ yields $f(\mathbf{x}) = \mathbf{y}$, as desired.

(6) It remains to show that if $\mathbf{x} \in V$, then $D(f^{-1})(f(\mathbf{x}))$ exists. It will be enough to show
$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{\left|f^{-1}(f(\mathbf{x})+\mathbf{k}) - \mathbf{x} - Df(\mathbf{x})^{-1}(\mathbf{k})\right|}{|\mathbf{k}|} = 0.$$
By Lemma 5.5, it will therefore be enough to show
$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{\left|Df(\mathbf{x})(f^{-1}(f(\mathbf{x})+\mathbf{k}) - \mathbf{x} - Df(\mathbf{x})^{-1}(\mathbf{k}))\right|}{|\mathbf{k}|} = 0. \tag{xiii}$$
We re-write this, first defining
$$h(\mathbf{k}) = f^{-1}(f(\mathbf{x})+\mathbf{k}) - \mathbf{x},$$
so that
$$\mathbf{k} = f(\mathbf{x} + h(\mathbf{k})) - f(\mathbf{x}).$$
Then
$$\begin{aligned}
&\frac{\left|Df(\mathbf{x})(f^{-1}(f(\mathbf{x})+\mathbf{k}) - \mathbf{x} - Df(\mathbf{x})^{-1}(\mathbf{k}))\right|}{|\mathbf{k}|} \\
&\quad = \frac{\left|Df(\mathbf{x})(h(\mathbf{k}) - Df(\mathbf{x})^{-1}(\mathbf{k}))\right|}{|\mathbf{k}|} = \frac{|\mathbf{k} - Df(\mathbf{x})(h(\mathbf{k}))|}{|\mathbf{k}|} \\
&\quad = \frac{|f(\mathbf{x}+h(\mathbf{k})) - f(\mathbf{x}) - Df(\mathbf{x})(h(\mathbf{k}))|}{|\mathbf{k}|} \\
&\quad = \frac{|f(\mathbf{x}+h(\mathbf{k})) - f(\mathbf{x}) - Df(\mathbf{x})(h(\mathbf{k}))|}{|h(\mathbf{k})|} \cdot \frac{|h(\mathbf{k})|}{|\mathbf{k}|}.
\end{aligned}$$
By (ix), the second factor here is bounded by 2; and the first factor has limit 0 at $\mathbf{0}$, since $\lim_{\mathbf{k}\to\mathbf{0}} h(\mathbf{k}) = \mathbf{0}$. Therefore (xiii), which completes the proof. □

Corollary 5.11 (Implicit Function Theorem). *Suppose* $\mathbf{a} \in \mathbb{R}^n$ *and* $\mathbf{b} \in \mathbb{R}^m$ *and* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *is continuously differentiable in a neighborhood of* $(\mathbf{a}, \mathbf{b})$*. Let*

$$f'(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} A & B \end{bmatrix},$$

where B *is a* square *matrix. If* $f(\mathbf{a}, \mathbf{b}) = 0$*, and* $\det B \neq 0$*, then there is a function* $g : \mathbb{R}^n \to \mathbb{R}^m$ *such that* $g(\mathbf{a}) = \mathbf{b}$*, and* $f(\mathbf{x}, g(\mathbf{x})) = 0$ *for all* $\mathbf{x}$ *in a neighborhood of* $\mathbf{a}$*, and* g *is differentiable on this neighborhood.*

Proof. Let F be the function

$$(\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{x}, f(\mathbf{x}, \mathbf{y})) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m.$$

Then

$$F'(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} I & 0 \\ A & B \end{bmatrix},$$

so $\det(F'(\mathbf{a}, \mathbf{b})) = \det B \neq 0$. Hence $(\mathbf{a}, \mathbf{b})$ and $(\mathbf{a}, \mathbf{0})$ have neighborhoods $U \times V$ and W respectively such that F is a bijection from $U \times V$ to W with a differentiable inverse. We can write this inverse as

$$(\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{x}, h(\mathbf{x}, \mathbf{y})),$$

where h is the appropriate function from W to V. Then

$$(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}, h(\mathbf{x}, \mathbf{y})) = (\mathbf{x}, f(\mathbf{x}, h(\mathbf{x}, \mathbf{y}))),$$

which means

$$\mathbf{y} = f(\mathbf{x}, h(\mathbf{x}, \mathbf{y})),$$

when $(\mathbf{x}, \mathbf{y}) \in W$. Hence, in particular,

$$\mathbf{0} = f(\mathbf{x}, h(\mathbf{x}, \mathbf{0})).$$

So we may let g be $\mathbf{x} \mapsto h(\mathbf{x}, \mathbf{0})$. □

6 Integration in several dimensions

A **compact interval** of $\mathbb{R}^n$ is a subset

$$I_0 \times \dots I_{n-1},$$

where each I_j is a closed, bounded interval of $\mathbb{R}$. Every compact interval I has a **measure**, denoted $\mu(I)$ and defined recursively by:

- $\mu([a, b]) = b - a$;
- $\mu(I \times I') = \mu(I) \cdot \mu(I')$.

Let I be the compact interval $I_0 \times \dots I_{n-1}$. A **partition** of I is a product

$$P_0 \times \dots \times P_{n-1},$$

where each P_j is a partition of I_j. Let the partition of I be P. If each P_j has m_j+1 elements, then P determines $m_0 \cdots m_{n-1}$ sub-intervals of I in an obvious way. Let J range over the set of these sub-intervals. If $t_J \in J$ for each J, and f is a function from I to $\mathbb{R}$, then f has a **Riemann sum**

$$\sum_J f(t_J) \cdot \mu(J)$$

for the partition P. Then f is **Riemann-integrable on** I if there is a number A such that, for all positive numbers ε, there is a partition of I such that, for all finer partitions of I into sub-intervals J, for all choices of t_J in J,

$$\left|\sum_J f(t_J) \cdot \mu(J) - A\right| < \varepsilon.$$

In this case, A is unique and can be denoted

$$\int_I f.$$

As in the special case where $I \subset \mathbb{R}$, so in the general case, integrability is equivalent to a condition known as **Riemann's Condition**.

Suppose S is a bounded subset of $\mathbb{R}^n$, and f is a real-valued function on S. Define the **characteristic function** of S to be $\chi_S : \mathbb{R}^n \to \mathbb{R}$, where

$$\chi_S(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in S; \\ 0, & \text{if } \mathbf{x} \notin S. \end{cases}$$

Then $f \cdot \chi_S$ is a well-defined function on $\mathbb{R}^n$, with value 0 on $\mathbb{R}^n \smallsetminus S$. If I and I' are two compact intervals including S, and $\int_I f \cdot \chi_S$ exists, then so does $\int_{I'} f \cdot \chi_S$, and the two intervals are equal; we can then define this common value to be the integral

$$\int_S f.$$

We may ask two questions:

- How can we tell whether integrals $\int_S f$ exist?
- If they exist, how can they be computed?

The set S has **measure zero** if for all positive ε there is a countable set $\{I_j : j \in \mathbb{N}\}$ of intervals such that

$$S \subseteq \bigcup_{j \in \mathbb{N}} I_j \quad \text{and} \quad \sum_{j=0}^{\infty} \mu(I_j) \leqslant \varepsilon.$$

Let us say that $\mathbf{x}$ is a **discontinuity** of f if f is not continuous at $\mathbf{x}$.

Theorem 6.1 (Lebesgue Criterion). *Let f be a bounded real-valued function on a compact interval I of $\mathbb{R}$. Then $\int_I f$ exists if and only if the set of discontinuities of f has measure zero.*

Example 6.2. $\mathbb{Q}$ has measure zero. Indeed, $\mathbb{Q}$ can be written as $\{x_j : j \in \mathbb{N}\}$. If $\varepsilon > 0$, let

$$I_j = \left[x_j - \frac{\varepsilon}{2^{j+2}}, x_j + \frac{\varepsilon}{2^{j+2}}\right].$$

Then

$$\sum_{j=0}^{\infty} \mu(I_j) = \sum_{j=0}^{\infty} \frac{\varepsilon}{2^{j+1}} = \varepsilon,$$

and $x_j \in I_j$. Let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} 0, & \text{if } x \notin \mathbb{Q}; \\ 1/\left|n\right|, & \text{if } x = m/n \text{ in lowest terms.} \end{cases}$$

Then the set of discontinuities of f is precisely $\mathbb{Q}$, so f is integrable on every compact interval, and each of the integrals is 0.

The **boundary** of a subset S of $\mathbb{R}^n$ is the set ∂S of points $\mathbf{x}$ of $\mathbb{R}^n$ such that every ball with center $\mathbf{x}$ contains points of both S and $\mathbb{R}^n \setminus S$. If S is bounded, then S is said to be **Jordan-measurable** if ∂S has measure zero.

Theorem 6.3. *Let S be a bounded, Jordan-measurable subset of $\mathbb{R}^n$, and let f be a real-valued function on S. Then $\int_S f$ exists if and only if the set of discontinuities of f on S has measure zero.*