Analysis
Course C9

T. W. Körner

June 2, 2001

Small print The syllabus for the course is defined by the Faculty Board Schedules (which are minimal for lecturing and maximal for examining). I should **very much** appreciate being told of any corrections or possible improvements and might even part with a small reward to the first finder of particular errors. This document is written in LaTeX2e and stored in the file labelled ~twk/1B/Anal.tex on emu in (I hope) read permitted form. My e-mail address is twk@dpmms.

Contents

1   Why do we bother?

It is surprising how many people think that that analysis consists in the difficult proofs of obvious theorems. All we need know, they say, is what a limit is, the definition of continuity and the definition of the derivative. All the rest is 'intuitively clear'.

If pressed they will agree that these definitions apply as much to the rationals $\mathbb{Q}$ as to the real numbers $\mathbb{R}$. They then have to explain the following interesting example.

Example 1.1. *If $f : \mathbb{Q} \to \mathbb{Q}$ is given by*

$$f(x) = -1 \qquad \text{if } x^2 < 2,$$
$$f(x) = 1 \qquad \text{otherwise,}$$

then
(i) f is continuous function with $f(0) = -1$, $f(2) = 1$ yet there does not exist a c with $f(c) = 0$,
(ii) f is a differentiable function with $f'(x) = 0$ for all x yet f is not constant.

What is the difference between $\mathbb{R}$ and $\mathbb{Q}$ which makes calculus work on one even though it fails on the other. Both are 'ordered fields', that is, both

support operations of 'addition' and 'multiplication' together with a relation 'greater than' ('order') with the properties that we expect. If the reader is interested she will find a complete list of the appropriate axioms in texts like the altogether excellent book of Spivak [8] and its many rather less excellent competitors, but, interesting as such things may be, they are irrelevant to our purpose which is not to consider the shared properties of $\mathbb{R}$ and $\mathbb{Q}$ but to identify a *difference* between the two systems which will enable us to exclude the possibility of a function like that of Example 1.1 for functions from $\mathbb{R}$ to $\mathbb{R}$.

To state the difference we need only recall a definition from the beginning of the course C5/6.

Definition 1.2. *If $a_n \in \mathbb{R}$ for each $n \geq 1$ and $a \in \mathbb{R}$ then we say that $a_n \to a$ if given $\epsilon > 0$ we can find an $n_0(\epsilon)$ such that*

$$|a_n - a| < \epsilon \text{ for all } n \geq n_0(\epsilon).$$

The key property of the reals, the *fundamental axiom* which makes everything work was also stated in the course C5/6.

Axiom 1.3 (The fundamental axiom of analysis). *If $a_n \in \mathbb{R}$ for each $n \geq 1$, $A \in \mathbb{R}$ and $a_1 \leq a_2 \leq a_3 \leq \ldots$ and $a_n < A$ for each n then there exists an $a \in \mathbb{R}$ such that $a_n \to a$ as $n \to \infty$.*

Less ponderously, and just as rigorously, the fundamental axiom for the real numbers says *every increasing sequence bounded above tends to a limit.*

Everything which depends on the fundamental axiom is analysis, everything else is mere algebra.

2 The axiom of Archimedes

In the course C5/6 you proved the following results on the limits of sequences.

Lemma 2.1. *(i) The limit is unique. That is, if $a_n \to a$ and $a_n \to b$ as $n \to \infty$ then $a = b$.*

(ii) If $a_n \to a$ as $n \to \infty$ and $n(1) < n(2) < n(3) \ldots$ then $a_{n(j)} \to a$ as $j \to \infty$.

(iii) If $a_n = c$ for all n then $a_n \to c$ as $n \to \infty$.

(iv) If $a_n \to a$ and $b_n \to b$ as $n \to \infty$ then $a_n + b_n \to a + b$.

(v) If $a_n \to a$ and $b_n \to b$ as $n \to \infty$ then $a_n b_n \to ab$.

(vi) If $a_n \to a$ as $n \to \infty$ and $a_n \neq 0$ for each n, $a \neq 0$ then $a_n^{-1} \to a^{-1}$.

(vii) If $a_n \leq A$ for each n and $a_n \to a$ as $n \to \infty$ then $a \leq A$.

Since these results were proved in the earlier course I shall not prove them. However, it is *essential* for further progress that the reader should be able to prove them for herself. If you can not prove these results *consult your supervisor at once*. You should do the same if you cannot prove the following variation on the fundamental axiom.

Lemma 2.2. *A decreasing sequence of real numbers bounded below tends to a limit.*

[Hint. If $a \leq b$ then $-b \leq -a$.]

Useful as the results of Lemma 2.1 are, they are also true of sequences in $\mathbb{Q}$. They are therefore mere, if important, algebra. Our first truly 'analysis' result may strike the reader as rather odd.

Theorem 2.3 (Axiom of Archimedes).

$$\frac{1}{n} \to 0 \ as \ n \to \infty$$

Theorem 2.3 shows that there is no 'exotic' real number ⌐ say (to choose an exotic symbol) with the property that $1/n > $ ⌐ for all integers $n \geq 1$ yet ⌐ > 0 (that is, ⌐ is strictly positive and yet smaller than all strictly positive rationals). There exist number systems with such exotic numbers (the famous 'non-standard analysis' of Abraham Robinson and the 'surreal numbers' of Conway constitute two such systems) but, just as the rationals are, in some sense, too small a system for the standard theorems of analysis to hold so these non-Archimedean systems are, in some sense, too big. Archimedes and Eudoxus realised the need for an axiom to show that there is no exotic number ⌐ bigger than any integer[1] (i.e. ⌐ $> n$ for all integers $n \geq 1$; to see the connection with our form of the axiom consider ⌐ $= 1/$⌐). However, in spite of its name, what was an axiom for Archimedes is a theorem for us.

Theorem 2.4. *Given any real number K we can find an integer n with $n > K$.*

3   The Bolzano–Weierstrass theorem

It is all very well knowing that the fundamental axiom is the foundation of mathematics but how can we find a proof technique which will allow us to apply it. One technique uses the Bolzano–Weierstrass theorem.

---

[1] Footnote for passing historians, this is a course in mathematics.

Theorem 3.1 (Bolzano–Weierstrass). *If $x_n \in \mathbb{R}$ and there exists a K such that $|x_n| \leq K$ for all n then we can find $n(1) < n(2) < \ldots$ and $x \in \mathbb{R}$ such that $x_{n(j)} \to x$ as $j \to \infty$.*

Recall from C5/6 that we say that a sequence converges if it tends to a limit. The Bolzano–Weierstrass theorem thus says that every bounded sequence of reals has a convergent subsequence. Notice that we say nothing about uniqueness, if $x_n = (-1)^n$ then $x_{2n} \to 1$ but $x_{2n+1} \to -1$ as $n \to \infty$.

There is a proof of the Bolzano–Weierstrass theorem by repeated dissection along the lines of the proof of the intermediate value theorem given in course C5/6. We shall give another proof based on the following elegant combinatorial lemma.

Lemma 3.2. *If $x_n \in \mathbb{R}$ then at least one of the following two statements must be true.*
(A) There exist $n(1) < n(2) < \ldots$ such that $x_{n(j)} \leq x_{n(j+1)}$ for each $j \geq 1$.
(B) There exist $m(1) < m(2) < \ldots$ such that $x_{m(j)} \geq x_{m(j+1)}$ for each $j \geq 1$.

There are a variety of techniques for using the fundamental axiom:- successive dissection, using Heine–Borel, looking for the supremum and so on. When the reader is confident in her control of the material in this course (and when she knows the various techniques enumerated in the last sentence) she should test her skill by proving the main results of this course by using each technique in turn. For the moment it seems sensible to concentrate on one proof technique until it is fully mastered. The technique chosen for this exposition is that of Bolzano–Weierstrass. We conclude this section with two important examples, Theorem 3.6 on the existence of suprema and Theorem 3.13 on the general principle of convergence.

It is an unfortunate fact of life that many, otherwise perfectly well behaved sets of numbers do not have greatest members (maxima). A simple example is given by $\{x \in \mathbb{R} : 0 < x < 1\}$. However, we shall see in Theorem 3.6 any non-empty bounded set of real numbers does have a least upper bound (supremum).

Definition 3.3. *Consider a non-empty set A of real numbers. We say that α is a least upper bound (or supremum) for A if the following two conditions hold.*
(i) $\alpha \geq a$ for all $a \in A$ (that is, α is an upper bound).
(ii) If $\beta \geq a$ for all $a \in A$ then $\beta \geq \alpha$ (that is, α is no greater than any possible upper bound).

Lemma 3.4. *The supremum is unique if it exists.*

It is convenient to have the following alternative characterisation of the supremum.

Lemma 3.5. *Consider a non-empty set A of real numbers; α is a* least upper bound *for A if and only if the following two conditions hold.*
(i) $\alpha \geq a$ for all $a \in A$.
(ii) Given $\epsilon > 0$ there exists an $a \in A$ such that $a + \epsilon \geq \alpha$.

We write $\sup A$ or $\sup_{a \in A} a$ for the supremum of A if it exists.
Here is the promised theorem.

Theorem 3.6 (Existence of the supremum). *If A is a non empty set of real numbers which is bounded above (that is, there exists a K such that $a \leq K$ for all $a \in A$) then A has a supremum.*

That this is a genuine theorem of analysis is shown by the fact that, if we work in $\mathbb{Q}$ the set $\{x \in \mathbb{Q} : x^2 < 2\}$ has no least upper bound. (See Question 18.13 for a detailed discussion.)
We leave it to the reader to define the *greatest lower bound* also called the *infimum* of a set A and written $\inf A$ or $\inf_{a \in A} a$. She should prove that

$$\inf_{a \in A} a = -\sup_{a \in A}(-a)$$

provided that either side of the equation exists.
As an example of its use, recall the following lemma from course C5/6.

Lemma 3.7. *We work in $\mathbb{C}$. If $\sum_{n=0}^{\infty} a_n z^n$ converges for $z = z_0$ then it converges for all z with $|z| < |z_0|$.*

The use of Theorem 3.6 gives us a rather transparent proof of the existence of the radius of convergence.

Theorem 3.8. *We work in $\mathbb{C}$. Consider the sum $\sum_{n=0}^{\infty} a_n z^n$. Either $\sum_{n=0}^{\infty} a_n z^n$ converges for all z (in this case we say that the series has infinite radius of convergence) or there exists a real number $R \geq 0$ such that*
(i) $\sum_{n=0}^{\infty} a_n z^n$ converges for all $|z| < R$,
(ii) $\sum_{n=0}^{\infty} a_n z^n$ diverges for all $|z| > R$,
(in this case we say that the series has radius of convergence R).

The nature of the supremum is illustrated by the following example which we leave to the reader as an exercise in the methods of the course C5/6. (As usual contact your supervisor if you can not do it.)

Example 3.9. *(i) The sum $\sum_{n=1}^{\infty} n^{-2} z^n$ has radius of convergence 1 and converges for all $|z| = 1$.*

(ii) The sum $\sum_{n=0}^{\infty} z^n$ has radius of convergence 1 and diverges for all $|z| = 1$.

(iii) The sum $\sum_{n=1}^{\infty} n^{-1} z^n$ has radius of convergence 1, diverges for $z = 1$ and converges for $z = -1$.

[When revising this part of the course you should also bear in mind Theorem 14.20 discussed later.]

We turn now to the general principle of convergence.

Definition 3.10. *We say that a sequence of real numbers x_n is a Cauchy sequence if given any $\epsilon > 0$ we can find $n_0(\epsilon)$ such that $|x_p - x_q| < \epsilon$ whenever $p, q \geq n_0(\epsilon)$.*

Our first lemma is merely algebraic.

Lemma 3.11. *Any convergent sequence forms a Cauchy sequence.*

The converse is a powerful theorem of analysis.

Theorem 3.12. *Any Cauchy sequence of real numbers converges.*

Combining the two results we get the general principle of convergence.

Theorem 3.13 (General principle of convergence). *A sequence of real numbers converges if and only if it is a Cauchy sequence.*

The general principle of convergence is usually too general to be used in the same way as the convergence tests of Course C5/6 but has great theoretical importance as we shall see.

For those who can not wait here is a proof of the uncountability of $\mathbb{R}$. The argument is much closer to Cantor's original proof than the standard one given in course C1. I present it as a heavily starred exercise.

Exercise 3.14 (Uncountability of the reals). *Let y_1, y_2, ... be any sequence of points in $\mathbb{R}$. Let $x_0 = 0$, $\delta_0 = 1$.*

(i) Show that you can construct inductively a sequence of real numbers $x_1, x_2, \ldots$ and positive numbers δ_j such that

(a) $|x_n - x_{n-1}| < \delta_{n-1}/4$,
(b) $x_n \neq y_n$,
(c) $0 < \delta_n < |x_n - y_n|$,
(d) $\delta_n < \delta_{n-1}/4$.

(ii) Show that $\delta_{n+m} < 4^{-m} \delta_n$ for $m, n \geq 0$ and deduce that the x_n form a Cauchy sequence. Conclude that x_n tends to a limit x.

(iii) Show that $|x_{n+m} - x_n| < \delta_n/3$ for all $m, n \geq 0$. Deduce that $|x - x_n| \leq \delta_n/3$ for all $n \geq 0$. Why does this show that $y_n \neq x$ for all n?

(iv) Deduce that the real numbers are uncountable.

4 Higher dimensions

In 1908, G. H. Hardy wrote a textbook to introduce the new rigorous analysis (or 'continental analysis' as it was known in a Cambridge even more insular than today) to 'first year students at the Universities whose abilities approach something like what is usually described as "scholarship standard" '. Apart from the fact that even the most hardened analyst would now hesitate to call an introduction to analysis *A Course of Pure Mathematics* it is striking how close the book is in both content and feel to a modern first course in analysis. (And where there are changes it is often not clear that the modern course[2] has the advantage.) The only major difference is that Hardy only studies the real line but later advances in mathematics mean that we must now study analysis in $\mathbb{R}^m$ as well.

In the course C1/2 you saw that $\mathbb{R}^m$ has a Euclidean norm

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{m} x_j^2 \right)^{1/2}$$

(taking the positive square root) with the following properties.

Lemma 4.1. *If* $\| \ \|$ *is the Euclidean norm on* $\mathbb{R}^m$ *then*
 (i) $\|\mathbf{x}\| \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^m$,
 (ii) If $\|\mathbf{x}\| = 0$ *then* $\mathbf{x} = \mathbf{0}$,
 (iii) If $\lambda \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{R}^m$ *then* $\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$.
 (iv) (The triangle inequality) If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *then* $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The only non-trivial part of this lemma was the triangle inequality which you proved using the Cauchy-Schwarz inequality.

In the course C5/6 you studied the notion of the limit for sequences in $\mathbb{R}^m$.

Definition 4.2. *If* $\mathbf{a}_n \in \mathbb{R}^m$ *for each* $n \geq 1$ *and* $\mathbf{a} \in \mathbb{R}^m$ *then we say that* $\mathbf{a}_n \to \mathbf{a}$ *if given* $\epsilon > 0$ *we can find an* $n_0(\epsilon)$ *such that*

$$\|\mathbf{a}_n - \mathbf{a}\| < \epsilon \text{ for all } n \geq n_0(\epsilon).$$

Notice that this shows that Definition 1.2 was about the *distance* between two points and not the absolute value of the difference of two numbers.

You proved the following results on sequences in $\mathbb{R}^m$ in exactly the same way as you proved the corresponding results in $\mathbb{R}$.

---

[2]Indeed, anyone who works through the exercises in Hardy as a first course and the exercises in Whittaker and Watson's even older *A Course of Modern Analysis* [11] as a second will have had a splendid education in analysis.

Lemma 4.3. *(i) The limit is unique. That is, if* $\mathbf{a}_n \to \mathbf{a}$ *and* $\mathbf{a}_n \to \mathbf{b}$ *as* $n \to \infty$ *then* $\mathbf{a} = \mathbf{b}$.

(ii) If $\mathbf{a}_n \to \mathbf{a}$ *as* $n \to \infty$ *and* $n(1) < n(2) < n(3) \ldots$ *then* $\mathbf{a}_{n(j)} \to \mathbf{a}$ *as* $j \to \infty$.

(iii) If $\mathbf{a}_n = \mathbf{c}$ *for all* n *then* $\mathbf{a}_n \to \mathbf{c}$ *as* $n \to \infty$.

(iv) If $\mathbf{a}_n \to \mathbf{a}$ *and* $\mathbf{b}_n \to \mathbf{b}$ *as* $n \to \infty$ *then* $\mathbf{a}_n + \mathbf{b}_n \to \mathbf{a} + \mathbf{b}$.

(v) Suppose $\mathbf{a}_n \in \mathbb{R}^m$, $\mathbf{a} \in \mathbb{R}^m$, $\lambda_n \in \mathbb{R}$ *and* $\lambda \in \mathbb{R}$. *If* $\mathbf{a}_n \to \mathbf{a}$ *and* $\lambda_n \to \lambda$ *then* $\lambda_n \mathbf{a}_n \to \lambda \mathbf{a}$.

Once again I leave it to the reader to check that she can indeed prove these results and to consult her supervisor if she can not. She should also make sure that she understands why certain results in Lemma 2.1 which dealt with an ordered field do not appear in Lemma 4.3 which deals with a vector space.

Lemma 4.3 is, of course, merely algebra and applies to $\mathbb{Q}^m$ as much as to $\mathbb{R}^m$. In order to do analysis we need a more powerful tool and, in keeping with the spirit of this course, we extend the Bolzano–Weierstrass theorem to $\mathbb{R}^m$. The extension is easy since the real work has already been done in the one dimensional case.

Theorem 4.4 (Bolzano–Weierstrass). *If* $\mathbf{x}_n \in \mathbb{R}^m$ *and there exists a* K *such that* $\|\mathbf{x}_n\| \leq K$ *for all* n *then we can find* $n(1) < n(2) < \ldots$ *and* $\mathbf{x} \in \mathbb{R}$ *such that* $\mathbf{x}_{n(j)} \to \mathbf{x}$ *as* $j \to \infty$.

Once again 'any bounded sequence has a convergent subsequence'.

The proof of Theorem 4.4 involves extending a one dimensional result to several dimensions. This is more or less inevitable because we stated the fundamental axiom of analysis in a one dimensional form. However the Bolzano–Weierstrass theorem itself contains no reference as to whether we are working in $\mathbb{R}$ or $\mathbb{R}^m$. Thus *exactly the same proofs* as we used in the one dimensional case give us a general principle of convergence in $\mathbb{R}^m$. (Compare Definition 3.10 and Theorem 3.13 with what follows.)

Definition 4.5. *We say that a sequence of points* $\mathbf{x}_n$ *in* $\mathbb{R}^m$ *is a* Cauchy sequence *if given any* $\epsilon > 0$ *we can find* $n_0(\epsilon)$ *such that* $\|\mathbf{x}_p - \mathbf{x}_q\| < \epsilon$ *for all* $p, q \geq n_0(\epsilon)$.

Theorem 4.6 (General Principle of Convergence). *A sequence in* $\mathbb{R}^m$ *converges if and only if it is a Cauchy sequence.*

Here is an example of the use of the general principle of convergence. (Following the conventions of course C5/6 we say that $\sum_{j=1}^{\infty} \mathbf{a}_j$ converges if the sequence of partial sums $\sum_{j=1}^{N} \mathbf{a}_j$ tends to a limit $\mathbf{s}$ say. We write $\sum_{j=1}^{\infty} \mathbf{a}_j = \mathbf{s}$.)

9

Theorem 4.7. *Let* $\mathbf{a}_n \in \mathbb{R}^m$ *for each* n. *If* $\sum_{j=1}^{\infty} \|\mathbf{a}_j\|$ *converges then so does* $\sum_{j=1}^{\infty} \mathbf{a}_j$.

This result substantially generalises the result in course C5/6 which says that an absolutely convergent series converges.

5 Open and closed sets

When we work in $\mathbb{R}$ the intervals are, in some sense, the 'natural' sets to consider. One of the problems that we face when we try to do analysis in many dimensions is that the types of sets with which we have to deal are much more diverse. It turns out that the so called closed and open sets are both sufficiently diverse and sufficiently well behaved to be useful. This short section is devoted to deriving some of their simpler properties. Novices frequently find the topic hard but eventually the reader will appreciate that this section is a rather trivial interlude in a deeper discussion.

The definition of a closed set is a natural one.

Definition 5.1. *A set* $F \subseteq \mathbb{R}^m$ *is closed if whenever* $\mathbf{x}_n \in F$ *for each* n *and* $\mathbf{x}_n \to \mathbf{x}$ *as* $n \to \infty$ *then* $\mathbf{x} \in F$.

Thus a set is closed in the sense of this course if it is 'closed under the operation of taking limits'. An indication of why this is good definition is given by the following version of the Bolzano–Weierstrass theorem.

Theorem 5.2. *If* K *is closed bounded set in* $\mathbb{R}^m$ *then every sequence in* K *has a subsequence converging to a point of* K.

When working in $\mathbb{R}^m$ the words 'closed and bounded' should always elicit the response 'Bolzano–Weierstrass'. We shall see important examples of this slogan in action in the next section (Theorem 6.3 and Theorem 6.12).

We turn now to the definition of an open set.

Definition 5.3. *A set* $U \subseteq \mathbb{R}^m$ *is open if whenever* $\mathbf{x} \in U$ *there exists an* $\epsilon(\mathbf{x}) > 0$ *such that whenever* $\|\mathbf{x} - \mathbf{y}\| < \epsilon(\mathbf{x})$ *we have* $\mathbf{y} \in U$.

Thus every point of an open set lies 'well inside the set'.

Example 5.4. *Consider sets in* $\mathbb{R}$. *The interval* $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ *is closed, the interval* $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ *is open, the interval* $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ *is neither open nor closed,* $\mathbb{R}$ *is both open and closed.*

Example 5.5. *Consider sets in $\mathbb{R}^m$. Let $\mathbf{x} \in \mathbb{R}^m$ and $r > 0$.*
(i) The set $B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^m : \|\mathbf{x} - \mathbf{y}\| < r\}$ is open.
(ii) The set $\bar{B}(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^m : \|\mathbf{x} - \mathbf{y}\| \leq r\}$ is closed.

We call $B(\mathbf{x}, r)$ the open ball of radius r and centre $\mathbf{x}$. We call $\bar{B}(\mathbf{x}, r)$ the closed ball of radius r and centre $\mathbf{x}$. Observe that the closed and and open balls of $\mathbb{R}$ are precisely the closed and open intervals.

The following restatement of the definition helps us picture an open set.

Lemma 5.6. *A subset A of $\mathbb{R}^m$ is open if and only if each point of A is the centre of an open ball lying entirely within A.*

Thus every point of an open set is surrounded by a ball consisting only of points of the set.

The topics of this section are often treated using the idea of *neighbourhoods*. We shall not use neighbourhoods very much but they come in useful from time to time.

Definition 5.7. *The set N is a neighbourhood of the point $\mathbf{x}$ if we can find an $r(\mathbf{x}) > 0$ such that $B(\mathbf{x}, r(\mathbf{x})) \subseteq N$.*

Thus a set is open if and only if it is neighbourhood of every point that it contains.

Returning to the main theme we note the following remarkable fact.

Lemma 5.8. *A subset A of $\mathbb{R}^m$ is open if and only if its complement $\mathbb{R}^m \setminus A$ is closed.*

We observe the following basic results on open and closed sets.

Lemma 5.9. *Consider the collection τ of open sets in $\mathbb{R}^m$.*
(i) $\emptyset \in \tau$, $\mathbb{R}^m \in \tau$.
(ii) If $U_\alpha \in \tau$ for all $\alpha \in A$ then $\bigcup_{\alpha \in A} U_\alpha \in \tau$.
(iii) If $U_1, U_2, \ldots, U_n \in \tau$ then $\bigcap_{j=1}^n U_j \in \tau$.

Lemma 5.10. *Consider the collection $\mathcal{F}$ of closed sets in $\mathbb{R}^m$.*
(i) $\emptyset \in \mathcal{F}$, $\mathbb{R}^m \in \mathcal{F}$.
(ii) If $F_\alpha \in \mathcal{F}$ for all $\alpha \in A$ then $\bigcap_{\alpha \in A} F_\alpha \in \mathcal{F}$.
(iii) If $F_1, F_2, \ldots, F_n \in \mathcal{F}$ then $\bigcup_{j=1}^n F_j \in \mathcal{F}$.

In this context you should note the following examples.

Example 5.11. *(i) $\bigcap_{j=1}^\infty (-2 - j^{-1}, 2 + j^{-1}) = [-2, 2]$, $\bigcap_{j=1}^\infty (-2 + j^{-1}, 2 - j^{-1}) = (-1, 1)$, $\bigcap_{j=1}^\infty (-2 + j^{-1}, 2 + j^{-1}) = (-1, 2]$.*
(ii) $\bigcup_{j=1}^\infty [-1 - j^{-1}, 1 + j^{-1}] = [-2, 2]$, $\bigcup_{j=1}^\infty [-1 + j^{-1}, 1 - j^{-1}] = (-1, 1)$, $\bigcup_{j=1}^\infty [-1 + j^{-1}, 1 + j^{-1}] = (-1, 2]$.

11

Thus the restriction to finite collections in part (iii) of Lemmas 5.9 and 5.10 can not be removed.

So far in this course we have only dealt with subsets and sequences. But analysis deals with functions, or rather with 'well behaved functions'. The simplest such class was defined in course C5/6.

Definition 5.12. *Let $E \subseteq \mathbb{R}^m$ we say that a function $f : E \to \mathbb{R}^p$ is continuous at some point $\mathbf{x} \in E$ if given $\epsilon > 0$ we can find a $\delta(\epsilon, \mathbf{x}) > 0$ such that whenever $\mathbf{y} \in E$ and $\|\mathbf{x} - \mathbf{y}\| < \delta(\epsilon, \mathbf{x})$ we have*

$$\|f(\mathbf{x}) - f(\mathbf{y})\| < \epsilon.$$

If f is continuous at every point $\mathbf{x} \in E$ we say that f is a continuous function on E.

In other words f is continuous at $\mathbf{x}$ if $f(\mathbf{y})$ can be made as close as we wish to $f(\mathbf{x})$ simply by taking $\mathbf{y}$ sufficiently close to $\mathbf{x}$. In other words f is locally approximately constant. In less precise terms 'small changes produce small effects'.

Exercise 5.13. *After looking at parts (iii) to (v) of Lemma 4.3 state the corresponding results for continuous functions. (Thus part (v) corresponds to the statement that if $\lambda : E \to \mathbb{R}$ and $f : E \to \mathbb{R}^p$ are continuous at $\mathbf{x} \in E$ then so is λf.) Prove your statements directly from Definition 5.12. Consult your supervisor if you have any problems.*

In this course we shall mainly use the following simple consequence of the definition.

Lemma 5.14. *Let $E \subseteq \mathbb{R}^m$ and suppose that the function $f : E \to \mathbb{R}^p$ is continuous at $\mathbf{x} \in E$. Then if $\mathbf{x}_n \in E$ and $\mathbf{x}_n \to \mathbf{x}$ it follows that $f(\mathbf{x}_n) \to f(\mathbf{x})$.*

However, more advanced work uses generalisations of the following lemma. (Remember that if $f : X \to Y$ is a function and $A \subseteq Y$ we write

$$f^{-1}(A) = \{x \in X \: : \: f(x) \in A\}$$

and this notation does not require f to be invertible.)

Lemma 5.15. *The function $f : \mathbb{R}^m \to \mathbb{R}^p$ is continuous if and only if $f^{-1}(O)$ is open whenever O is open.*

As a simple example of how Lemma 5.15 can be used contrast the 'ϵ, δ proof' of the next lemma using Definition 5.12 with a proof using Lemma 5.15.

Lemma 5.16. *If $f : \mathbb{R}^m \to \mathbb{R}^p$ and $g : \mathbb{R}^p \to \mathbb{R}^g$ are continuous the so is their composition $g \circ f$.*

(Recall that we write $g \circ f(\mathbf{x}) = g(f(\mathbf{x}))$.)

Although we shall not make much use of it, the following is an important consequence of the Bolzano–Weierstrass theorem.

Theorem 5.17. *Suppose that F_1, F_2, ... are non-empty bounded closed sets in $\mathbb{R}^m$ such that $F_1 \supseteq F_2 \supseteq \dots$. Then $\bigcap_{j=1}^\infty F_j \neq \emptyset$.*

That is, the intersection of bounded, closed, nested non-empty sets is itself non-empty. The example $F_j = [j, \infty)$ shows that 'bounded' can not be dropped from the hypothesis. The example $F_j = (0, j^{-1})$ shows that 'closed' can not be dropped from the hypothesis.

Finally, it is worth emphasising that openess is not an intrinsic property of a set but depends on the space in which the set finds itself.

Example 5.18. *The set $\{(x, 0) \in \mathbb{R}^2 : 0 < x < 1\}$ is not open in $\mathbb{R}^2$.*

(If you have a sufficiently logical mind Example 5.18 is otiose. Few of us have that kind of mind.)

6   The central theorems of analysis

This section contains the core of a first course in analysis. For completeness I shall include some theorems already proved in course C5/6. The first such theorem is the intermediate value theorem proved in C5/6 by successive bisection. That proof is entirely satisfactory but, as an exercise, we shall prove the result using the method of Bolzano–Weierstrass.

Theorem 6.1 (Intermediate value theorem). *If $f : [a, b] \to \mathbb{R}$ is continuous and $f(a) \leq 0 \leq f(b)$ then there exists a $c \in [a, b]$ such that $f(c) = 0$.*

You should compare this result with part (i) of Example 1.1. Note also that the result depends on f taking values in $\mathbb{R}$ rather than $\mathbb{R}^2$ or $\mathbb{C}$.

Example 6.2. *If $f : [0, 1] \to \mathbb{C}$ is given by $f(t) = \exp \pi i t$ then f is continuous, $f(0) = 1$, $f(1) = -1$ but $f(t) \neq 0$ for all t.*

Our next result looks a little abstract at first.

Theorem 6.3. *Let K be a closed bounded subset of $\mathbb{R}^m$ and $f : K \to \mathbb{R}^p$ a continuous function. Then $f(K)$ is closed and bounded.*

Thus the continuous image of a closed bounded set is closed and bounded. (The example $m = 2$, $p = 1$,

$$K = \{(x, y) : y \geq x^{-1}, x > 0\}$$

$f(x, y) = x$ shows that the continuous image of a closed set need not be closed.

The example The example $m = 1$, $p = 1$, K the open interval $(0, 1)$, $f(x) = x^{-1}$, shows that the continuous image of a bounded set need not be bounded.)

We derive a much more concrete corollary.

Theorem 6.4. *Let K be a non-empty closed bounded subset of $\mathbb{R}^m$ and $f : K \to \mathbb{R}$ a continuous function. Then we can find $\mathbf{k}_1$ and $\mathbf{k}_2$ in K such that*

$$f(\mathbf{k}_1) \leq f(\mathbf{k}) \leq f(\mathbf{k}_2)$$

for all $\mathbf{k} \in K$.

In other words a real valued continuous function on a closed bounded set is bounded and attains its bounds. Less usefully we may say that, in this case f actually has a maximum and a minimum. Notice that there is no analogous result for vector valued functions. Much of economics consists in an attempt to disguise this fact (there is unlikely to be a state of the economy in which *everybody* is best off[3]).

Theorem 6.4 has an even more concrete consequence.

Theorem 6.5. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then we can find $k_1, k_2 \in [a, b]$ such that*

$$f(k_1) \leq f(k) \leq f(k_2)$$

for all $k \in [a, b]$

In course C5/6 you used this result to prove Rolle's theorem.

Theorem 6.6 (Rolle's theorem). *If $f : [a, b] \to \mathbb{R}$ is a continuous function with f differentiable on (a, b) and $f(a) = f(b)$ then we can find a $c \in (a, b)$ such that $f'(c) = 0$.*

By 'tilting Rolle's theorem' we obtain the mean value (in one dimension).

---

[3]Of course economists know this but few of their public pronouncements seem to mention it.

Theorem 6.7 (The one dimensional mean value theorem). *If $f : [a, b] \to \mathbb{R}$ is a continuous function with f differentiable on (a, b) then we can find a $c \in (a, b)$ such that*

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

Historically, mathematicians were slow to grasp the importance of the mean value theorem and the subtlety of its proof. Fallacious proofs still appeared in textbooks (British textbooks, admittedly) in the early years of the 20th century. (The proofs were fallacious in the strictest sense. They 'proved' the impossibility of the result given in part (ii) of Example 1.1.)

As you saw in course C5/6, the mean value theorem Theorem 6.7 has the following key corollaries.

Theorem 6.8. *Suppose that $f : (\alpha, \beta) \to \mathbb{R}$ is a differentiable function.*
   (i) If $f'(x) \geq 0$ for all $x \in (\alpha, \beta)$ then f is an increasing function on (α, β) (that is $f(x) \leq f(y)$ whenever $\alpha < x < y < \beta$).
   (ii) If $f'(x) > 0$ for all $x \in (\alpha, \beta)$ then f is a strictly increasing function on (α, β) (that is $f(x) < f(y)$ whenever $\alpha < x < y < \beta$).
   (iii) If $f'(x) = 0$ for all $x \in (\alpha, \beta)$ then f is constant.
   (iv) If $|f'(x)| \leq K$ for all $x \in (\alpha, \beta)$ then $|f(u) - f(v)| \leq K|u - v|$ for all $u, v \in (\alpha, \beta)$

Part (iii) of of Theorem 6.8 is the result that tells us that the solutions (if any) of $f'(x) = g(x)$ on (α, β) differ only by the addition of a constant (the so called 'constant of integration').

Part (iv) of of Theorem 6.8 is an excellent example of the way that the theorems of this section (and many other theorems of analysis) convert *local* information into *global* information. We know that $|f'(x)| \leq K$ so that *locally* the rate of change of f is no greater than K. We deduce that $|f(u) - f(v)| \leq K|x - y|$ so that *globally* the rate of change of f is no greater than K.

The final theorem of this section (Theorem 6.12 on uniform continuity) is another excellent example of the conversion of local information to global. We need a couple of definitions and examples. Recall first from Definition 5.12 what it means for a function to be continuous on a set

Definition 6.9. *Let $E \subseteq \mathbb{R}^m$ we say that a function $f : E \to \mathbb{R}^p$ is continuous on E if given any point $\mathbf{x} \in E$ and any $\epsilon > 0$ we can find a $\delta(\epsilon, \mathbf{x}) > 0$ such that whenever $\mathbf{y} \in E$ and $\|\mathbf{x} - \mathbf{y}\| < \delta(\epsilon, \mathbf{x})$ we have*

$$\|f(\mathbf{x}) - f(\mathbf{y})\| < \epsilon.$$

Now compare that definition with our definition of *uniform* continuity.

Definition 6.10. *Let $E \subseteq \mathbb{R}^m$ we say that a function $f : E \to \mathbb{R}^p$ is uniformly continuous on E if given any $\epsilon > 0$ we can find a $\delta(\epsilon) > 0$ such that whenever $\mathbf{x}, \mathbf{y} \in E$ and $\|\mathbf{x} - \mathbf{y}\| < \delta(\epsilon)$ we have*

$$\|f(\mathbf{x}) - f(\mathbf{y})\| < \epsilon.$$

Example 6.11. *The following three functions are continuous but not uniformly continuous.*
(i) $f_1 : \mathbb{R} \to \mathbb{R}$ *given by* $f_1(x) = x^2$.
(ii) $f_2 : (0, 1) \to \mathbb{R}$ *given by* $f_2(x) = x^{-1}$.
(iii) $f_3 : (0, 1) \to \mathbb{R}$ *given by* $f_2(x) = \sin(x^{-1})$.

Theorem 6.12. *Let K be a closed bounded subset of $\mathbb{R}^m$. If $f : K \to \mathbb{R}^p$ is continuous on K then f is uniformly continuous on K.*

The results of this section are the foundations of all analysis. It took 200 years to found analysis on the same axiomatic basis as Greek geometry and the successful completion of the project was a triumph of the human intellect. If you can trace a compete path from the fundamental axiom to the one dimensional mean value theorem you will have mastered the essential point of this course.

7   Differentiation

A function is continuous if it is locally approximately constant. A function is differentiable if it is locally approximately linear. More precisely, a function is continuous at a point $\mathbf{x}$ if it is locally approximately constant with an error which decreases to zero as we approach $\mathbf{x}$. A function is differentiable at a point $\mathbf{x}$ if it is locally approximately linear with an error which decreases to zero *faster than linearly* as we approach $\mathbf{x}$.

Definition 7.1. *Suppose that E is a subset of of $\mathbb{R}^m$ and $\mathbf{x}$ a point such that there exists a $\delta > 0$ with $B(\mathbf{x}, \delta) \subseteq E$. We say that $f : E \to \mathbb{R}^p$, is differentiable at $\mathbf{x}$ if we can find a linear map $\alpha : \mathbb{R}^m \to \mathbb{R}^p$ such that, when $\|\mathbf{h}\| < \delta$*

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \alpha\mathbf{h} + \boldsymbol{\epsilon}(\mathbf{x}, \mathbf{h})\|\mathbf{h}\| \qquad\qquad ★$$

where $\|\boldsymbol{\epsilon}(\mathbf{x}, \mathbf{h})\| \to 0$ as $\|\mathbf{h}\| \to 0$. We write $\alpha = Df(\mathbf{x})$ or $\alpha = f'(\mathbf{x})$.
If E is open and f is differentiable at each point of E we say that f is differentiable on E.

Needless to say, the centre of the definition is the formula ★ and the reader should concentrate on understanding the rôle of each term in that formula. The rest of the definition is just supporting waffle. Formula ★ is sometimes rewritten

$$\frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \alpha \mathbf{h}}{\|\mathbf{h}\|} \to 0$$

as $\|\mathbf{h}\| \to 0$.

You have already studied differentiation in course C5/6 and you regularly use partial differentiation in applied courses and elsewhere. I shall therefore concentrate on recalling some of the key points.

Lemma 7.2. *Let f be as in Definition 7.1. If we use standard coordinates then if f is differentiable at $\mathbf{x}$ its partial derivatives $f_{i,j}(\mathbf{x}) = \dfrac{\partial f_i}{\partial x_j}$ exist and the matrix of the derivative $Df(\mathbf{x})$ is the Jacobian matrix $(f_{i,j}(\mathbf{x}))$ of partial derivatives.*

It is customary to point out that the existence of the partial derivatives does not imply the differentiability of the function (see Example 8.5 below) but the main objections to over-reliance on partial derivatives are that this makes formulae cumbersome and stifles geometric intuition. Let your motto be **'coordinates and matrices for calculation, vectors and linear maps for understanding'.**

One key property of differentiation is the chain rule. You have already seen this discussed and proved in course C5/6. Our discussion will run along the same lines but is slightly neater.

Lemma 7.3. *If $\alpha : \mathbb{R}^m \to \mathbb{R}^p$ is a linear map we can find a constant K such that $|\alpha \mathbf{x}| \leq K \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^m$.*

We can now make the following definition.

Definition 7.4. *If $\alpha : \mathbb{R}^m \to \mathbb{R}^p$ is a linear map then*

$$\|\alpha\| = \sup_{\|\mathbf{x}\| \leq 1} \|\alpha \mathbf{x}\|.$$

The following easy exercise provides an equivalent definition.

Exercise 7.5. *Show that if $\alpha : \mathbb{R}^m \to \mathbb{R}^p$ is a linear map then*

$$\|\alpha\| = \sup_{\|\mathbf{x}\| = 1} \|\alpha \mathbf{x}\|.$$

This 'operator norm' has many pleasant properties.

Lemma 7.6. *Let $\alpha, \beta : \mathbb{R}^m \to \mathbb{R}^p$ be linear maps.*
  (i) If $\mathbf{x} \in \mathbb{R}^m$ then $\|\alpha \mathbf{x}\| \leq \|\alpha\| \, \|\mathbf{x}\|$.
  (ii) $\|\alpha\| \geq 0$,
  (iii) If $\|\alpha\| = 0$ then $\alpha = 0$,
  (iv) If $\lambda \in \mathbb{R}$ then $\|\lambda\alpha\| = |\lambda| \|\alpha\|$.
  (v) (The triangle inequality) $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.
  (vi) If $\gamma : \mathbb{R}^p \to \mathbb{R}^q$ then $\|\gamma\alpha\| \leq \|\gamma\| \, \|\alpha\|$.

Lemma 7.7 (The chain rule). *Let U be a neighbourhood of $\mathbf{x}$ in $\mathbb{R}^m$, and V a neighbourhood of $\mathbf{x}$ in $\mathbb{R}^p$. Suppose that $f : U \to V$ is differentiable at $\mathbf{x}$ with derivative α, that $f : V \to \mathbb{R}^q$ is differentiable at $\mathbf{y}$ with derivative β and that $f(\mathbf{x}) = \mathbf{y}$. Then $g \circ f$ is differentiable at $\mathbf{x}$ with derivative $\beta\alpha$.*

In more condensed notation

$$D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x}). \qquad \bigstar\bigstar$$

It is important to see that formula $\bigstar\bigstar$ is exactly what we should expect and that its proof is simple book-keeping.

Remark One of the most troublesome culture clashes between pure mathematics and applied is that to an applied mathematician variables like x and t have meanings such as position and time whereas to a pure mathematician all variables are 'dummy variables' or 'place-holders' to be interchanged at will. To a pure mathematician v is an arbitrary function defined by its effect on a variable so that $v(t) = At^3$ means precisely the same thing as $v(x) = Ax^3$ whereas to an applied mathematician who thinks of v as a velocity the statements $v(t) = At^3$ and $v(x) = Ax^3$ mean very different (indeed incompatible) things. The applied mathematician writes

$$\frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt}$$

but the nearest the pure mathematician can get to this statement is

$$\frac{d}{dt}v(x(t)) = v'(x(t))x'(t).$$

The same problems occur in still more confusing form when we deal with partial derivatives. In particular the formula

$$\frac{\partial f}{\partial x_i} = \sum_{k=1}^{n} \frac{\partial f}{\partial y_k}\frac{\partial y_k}{\partial x_i}$$

makes no sense without using further *unspoken* conventions. The only remedy that I can suggest is 'think like a pure mathematician when doing pure mathematics and like an applied mathematician when doing applied mathematics'.

The following simple result has a simple direct proof but it is instructive to prove it by the chain rule.

Lemma 7.8. *Let U be a neighbourhood of $\mathbf{x}$ in $\mathbb{R}^m$. Suppose that $f, g : U \to \mathbb{R}^p$ are differentiable at $\mathbf{x}$. Then $f + g$ is differentiable at $\mathbf{x}$ and $D(f+g)(\mathbf{x}) = Df(\mathbf{x}) + Dg(\mathbf{x})$.*

So far our study of differentiation in higher dimensions has remained on the level of mere algebra. (The definition of the operator norm used the supremum and so lay deeper but we could have avoided the use of the operator norm as was done in course C5/6.) The next result is a true theorem of analysis.

Theorem 7.9 (The mean value inequality). *Suppose that U is an open set in $\mathbb{R}^m$ and that $f : U \to \mathbb{R}^p$ is differentiable. Consider the straight line segment*

$$L = \{(1 - t)\mathbf{a} + t\mathbf{b} : 0 \le t \le 1\}$$

joining $\mathbf{a}$ and $\mathbf{b}$. If $L \subseteq U$ (i.e. L lies entirely within U) and $\|Df(\mathbf{x})\| \le K$ for all $\mathbf{x} \in L$ then

$$\|f(\mathbf{a}) - f(\mathbf{b})\| \le K\|\mathbf{a} - \mathbf{b}\|.$$

The reader may be disappointed that a theorem involving an equality in one dimension should be replaced by one involving an inequality in many dimensions. However, inspection of the use of the mean value theorem in one dimension shows that its essential content is the statement that 'local growth bounds global growth' and this statement is precisely the mean value inequality.

In any case the following example shows that we can not hope for an exact analogue of the one dimensional theorem.

Example 7.10. *Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f(t) = (\cos t, \sin t)^T$. Then $f(0) = f(2\pi)$ but $Df(t) \ne 0$ for all t.*

It is also clear (though we shall not prove it, and, indeed can not state it without using concepts which we have not formally defined) that the correct generalisation when L is not a straight line will run as follows. 'If L is a well behaved path lying entirely within U and $\|Df(\mathbf{x})\| \le K$ for all $\mathbf{x} \in L$ then $\|f(\mathbf{a}) - f(\mathbf{b})\| \le K \times \text{length } L$'.

Example 7.11. *Let*

$$U = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| > 1\} \setminus \{(x,0)^T : x \le 0\}\}$$

If we take $\theta(\mathbf{x})$ to be the unique solution of

$$\cos(\theta(\mathbf{x})) = x, \ \sin(\theta(\mathbf{x})) = y, \ -\pi < \theta(\mathbf{x}) < \pi$$

for $\mathbf{x} = (x,y)^T \in U$ then $\theta : U \to \mathbb{R}$ is everywhere differentiable with $\|D\theta(\mathbf{x})\| < 1$. However if $\mathbf{a} = (-1, 10^{-1})^T$, $\mathbf{b} = (-1, -10^{-1})^T$ then

$$|\theta(\mathbf{a}) - \theta(\mathbf{b})| > \|\mathbf{a} - \mathbf{b}\|.$$

8 Local Taylor theorems

In this section we exploit the mean value inequality to obtain information about the behaviour of reasonably well behaved functions $f : \mathbb{R}^m \to \mathbb{R}$ in the neighbourhood of a point $\mathbf{x}$. (Since we can always translate we shall sometimes put $\mathbf{x} = \mathbf{0}$ without comment.) It may be worth alerting the reader to the fact that we shall be using the one-dimensional mean value inequality (see Theorem 6.8 (iv)) in forms like the following. 'If $|f_{,1}(x,c)| \le K$ for some fixed c and all $a \le x \le b$ then $|f(a,c) - f(b,c)| \le K|b - a|$.' In this section we shall use row rather than column vectors.

Here is our first example.

Theorem 8.1 (Continuity of partial derivatives implies differentiability). *Suppose $\delta > 0$, $\mathbf{x} = (x,y) \in \mathbb{R}^2$, $B(\mathbf{x},\delta) \subseteq E \subseteq \mathbb{R}^2$ and that $f : E \to \mathbb{R}$. If the partial derivatives $f_{,1}$ and $f_{,2}$ exist in $B(\mathbf{x},\delta)$ and are continuous at $\mathbf{x}$ then writing*

$$f((x+h, y+k)) = f(x,y) + f_{,1}(x,y)h + f_{,2}(x,y)k + \epsilon(h,k)(h^2 + k^2)^{1/2}$$

we have $\epsilon(h,k) \to 0$ as $(h^2 + k^2)^{1/2} \to 0$. (In other words f is differentiable at $\mathbf{x}$.)

Although this is not one of the great theorems of all time (or indeed of this course) it does provide a useful short cut for proving functions differentiable. The following easy extension is left to the reader.

Theorem 8.2. *(i) Suppose $\delta > 0$, $\mathbf{x} \in \mathbb{R}^m$ $B(\mathbf{x},\delta) \subseteq E \subseteq \mathbb{R}^m$ and that $f : E \to \mathbb{R}$. If the partial derivatives $f_{,1}$, $f_{,2}$, ... $f_{,m}$ exist in $B(\mathbf{x},\delta)$ and are continuous at $\mathbf{x}$ then f is differentiable at $\mathbf{x}$.*

(ii) Suppose $\delta > 0$, $\mathbf{x} \in \mathbb{R}^m$ $B(\mathbf{x},\delta) \subseteq E \subseteq \mathbb{R}^m$ and that $f : E \to \mathbb{R}^p$. If the partial derivatives $f_{i,j}$ exist in $B(\mathbf{x},\delta)$ and are continuous at $\mathbf{x}$ $[1 \le i \le p, \ 1 \le j \le m]$ then f is differentiable at $\mathbf{x}$.

The next result, although along similar lines, is much more interesting. We write

$$f_{,ij}(\mathbf{x}) = (f_{,j})_{,i}(\mathbf{x}),$$

or, in more familiar notation

$$f_{,ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Theorem 8.3 (Second order Taylor series). *Suppose $\delta > 0$, $\mathbf{x} = (x, y) \in \mathbb{R}^2$ $B(\mathbf{x}, \delta) \subseteq E \subseteq \mathbb{R}^2$ and that $f : E \to \mathbb{R}$. If the partial derivatives $f_{,1}$, $f_{,2}$, $f_{,11}$, $f_{,12}$, $f_{,22}$ exist in $B(\mathbf{x}, \delta)$ and are continuous at $\mathbf{x}$ then writing*

$$f((x + h, y + k)) = f(x, y) + f_{,1}(x, y)h + f_{,2}(x, y)k$$
$$+ \tfrac{1}{2}(f_{,11}(x, y)h^2 + 2f_{,12}(x, y)hk + f_{,22}(x, y)k^2) + \epsilon(h, k)(h^2 + k^2)$$

we have $\epsilon(h, k) \to 0$ as $(h^2 + k^2)^{1/2} \to 0$.

We have the following important corollary.

Theorem 8.4 (Symmetry of second partial derivative). *Suppose $\delta > 0$, $\mathbf{x} = (x, y) \in \mathbb{R}^2$ $B(\mathbf{x}, \delta) \subseteq E \subseteq \mathbb{R}^2$ and that $f : E \to \mathbb{R}$. If the partial derivatives $f_{,1}$, $f_{,2}$, $f_{,11}$, $f_{,12}$, $f_{,21}$ $f_{,22}$ exist in $B(\mathbf{x}, \delta)$ and are continuous at $\mathbf{x}$ then $f_{,12}(\mathbf{x}) = f_{,21}(\mathbf{x})$.*

Stripped of all the fine (and for practical purposes irrelevant) detail this tells us that we can interchange the order of partial differentiation for well behaved functions.

It is possible to produce plausibility arguments for the symmetry of second partial derivatives. Here are a couple.

(1) If f is a multinomial, i.e. $f(x, y) = \sum_{p=0}^{P} \sum_{q=0}^{Q} a_{p,q} x^p y^q$, then $f_{,12} = f_{,21}$. But smooth functions are very close to being polynomial so we would expect the result to be true in general.

(2) Although we can not interchange limits in general it is plausible that if f is well behaved then

$$f_{,12}(x, y) = \lim_{h \to 0} \lim_{k \to 0} h^{-1}k^{-1}(f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y))$$
$$= \lim_{k \to 0} \lim_{h \to 0} h^{-1}k^{-1}(f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y))$$
$$= f_{,21}(x, y).$$

However, these are merely plausible arguments. They do not make clear the role of the continuity of the second derivative (in Example 8.6 we shall see

21

that the result may fail for discontinuous second partial derivatives). More fundamentally they are *algebraic* arguments and, as the use of the mean value theorem indicates, the result is one of *analysis*.

Theorems 8.1 and 8.4 are inextricably linked in many examiners' minds with the following two counter-examples. The present lecturer would rather you forgot all the course than that you forgot how to prove the theorems and remembered the counter-examples but will not let his personal feelings get in the way of his duty.

Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ has $f(\mathbf{0}) = 0$. We write $g(r, \theta) = f(r \cos \theta, r \sin \theta)$ (i.e. use polar coordinates). If f is once differentiable then, to first order in r,

(i) $g(r, \theta)$ grows like r,
(ii) the contours $f(x, y) = c$ are parallel lines.

If we seek counter-examples it is reasonable to look for an f such that (i) is true but (ii) is not. We might therefore be led to consider $g(r, \theta) = r \sin 2\theta$. (Finding counter-examples is more like tinkering with a meccano kit than anything else.)

Example 8.5. *If*

$$f(x, y) = \frac{xy}{(x^2 + y^2)^{1/2}} \qquad for \ (x, y) \neq (0, 0),$$

$$f(0, 0) = 0,$$

then f has first partial derivatives everywhere, f is differentiable except at $(0, 0)$ where it is not.

Emboldened by our success we could well guess immediately a suitable function to look for in the context of Theorem 8.4. If not let us consider f and g as before. Suppose that

$$f(\mathbf{0}) = f_{,1}(\mathbf{0}) = f_{,2}(\mathbf{0}) = 0.$$

If f obeys the conclusions of Theorem 8.3 (i.e. has a second order Taylor series) then, to second order in r,

(i) $g(r, \theta)$ grows like r^2,
(ii) the contours $f(x, y) = c$ are conics.

Once again we look for an f such that (i) is true but (ii) is not. We might, for example, consider $g(r, \theta) = r^2 \sin 4\theta$.

Example 8.6. *If*

$$f(x, y) = \frac{xy(x^2 - y^2)}{(x^2 + y^2)} \qquad for \ (x, y) \neq (0, 0),$$

$$f(0, 0) = 0,$$

then f has first and second partial derivatives everywhere but $f_{,12}(0,0) \neq f_{,21}(0,0)$.

It is profoundly unfortunate that the last contact many mathematicians have with this topic is a couple of complicated counter-examples. Multi-dimensional calculus leads towards differential geometry and infinite dimensional calculus (functional analysis). Both subjects depend on *understanding* objects which we know to be well behaved but which our limited geometric intuition makes it hard for us to comprehend. Counter-examples such as the ones just produced are simply irrelevant

What happens if a function is smooth (has partial derivatives of all orders)? The following theorem completes the discussion of this section. It is not on the syllabus and left to the reader as an exercise (but a very instructive one).

Theorem 8.7 (The local Taylor's theorem). *Suppose $\delta > 0$, $\mathbf{x} \in \mathbb{R}^m$ $B(\mathbf{x}, \delta) \subseteq E \subseteq \mathbb{R}^m$ and that $f : E \to \mathbb{R}^p$. If the all partial derivatives $f_{i,j}$, $f_{i,jk}$, $f_{i,jkl}$, ... exist in $B(\mathbf{x}, \delta)$ and are continuous at $\mathbf{x}$ then writing*

$$f_i(\mathbf{x} + \mathbf{h}) = f_i(\mathbf{x}) + \sum_{j=1}^{m} f_{i,j}(\mathbf{x})h_j + \sum_{j=1}^{m}\sum_{k=1}^{m} f_{i,jk}(\mathbf{x})h_j h_k$$
$$+ \cdots + sum\ up\ to\ nth\ powers + \epsilon_i(\mathbf{h})\|\mathbf{h}\|^n$$

we have $\|\boldsymbol{\epsilon}(\mathbf{h})\| \to 0$ as $\|\mathbf{h}\| \to 0$.

The wide awake reader may observe that I used (gasp, horror!) coordinates in the statement of Theorem 8.7. However, the main formula can be stated in a coordinate free way as

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \alpha_1(\mathbf{h}) + \alpha_2(\mathbf{h}, \mathbf{h}) + + \cdots + \alpha_n(\mathbf{h}, \mathbf{h}, \ldots, \mathbf{h}) + \boldsymbol{\epsilon}(\mathbf{h})\|\mathbf{h}\|^n,$$

where $\alpha_k : \mathbb{R}^m \times \mathbb{R}^m \cdots \times \mathbb{R}^m \to \mathbb{R}^p$ is linear in each variable (i.e. a k-linear function). For more details consult section 12 of chapter VIII of Dieudonné's *Foundations of Modern Analysis* [5] where the higher derivatives are dealt with in a coordinate free way. Like Hardy's book [6], Dieudonné's is a masterpiece but in very different tradition.

Anyone who feels that the higher derivatives are best studied using coordinates should reflect that if $f : \mathbb{R}^3 \to \mathbb{R}^3$ is well behaved then the 'third derivative behaviour' at of f at a single point is apparently given by the $3 \times 3 \times 3 \times 3 = 81$ numbers $f_{i,jkl}(\mathbf{x})$. By symmetry (see Theorem 8.4) only 30 of the numbers are distinct but these 30 numbers are independent (consider polynomials in three variables for which the total degree of each term is 3). How can we understand the information carried by an array of 30 real numbers?

9 Riemann integration

(Much of this section was sketched in the course C5/6 but this treatment will supply some important extra detail.) Everybody knows what area is, but then everybody knows what honey tastes like. But does honey taste the same to you as it does to me? Perhaps the question is unanswerable but for many practical purposes it is sufficient that we agree on what we call honey. In the same way it is important that when two mathematicians talk about area that (1) they should agree on which sets E actually have area, and that (2) when a set E has area they should agree as to what the area is.

At one time some mathematicians may well have hoped that every bounded set in $\mathbb{R}^2$ could be given an area, every bounded set in $\mathbb{R}^3$ could be given a volume, and so on. However Banach and Tarski showed that a ball in $\mathbb{R}^3$ could be split into seven parts which could be reassembled after rotation and translation into a ball of smaller radius. (The proof is non-trivial but there is a good discussion in Wagon's book *The Banach–Tarski Paradox* [10].) If each of these seven parts had a volume then repeated Banach Tarski decompositions would enable us to get a quart into a pint pot. Warned by this we shall not attempt to assign area to every subset of $\mathbb{R}^2$ or to integrate every function.

Turning specifically to the Riemann integral we consider bounded functions $f : [a, b] \to \mathbb{R}$. (Thus there exists a K with $|f(x)| \leq K$ for all $x \in [a, b]$). A dissection $\mathcal{D}$ of $[a, b]$ is a finite subset of $[a, b]$ containing the end points a and b. By convention we write

$$\mathcal{D} = \{x_0, x_1, \ldots, x_n\} \text{ with } a = x_0 \leq x_1 \leq x_2 \leq \cdots \leq x_n = b.$$

We define the *upper sum* and *lower sum* of f associated with $\mathcal{D}$ by

$$S(f, \mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} f(x),$$

$$s(f, \mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{x \in [x_{j-1}, x_j]} f(x)$$

The next lemma is hardly more than an observation but it is the key to the proper treatment of the integral.

Lemma 9.1 (Key integration property). *If* $f : [a, b] \to \mathbb{R}$ *is bounded and* $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *are two dissections then*

$$S(f, \mathcal{D}_1) \geq S(f, \mathcal{D}_1 \cup \mathcal{D}_2) \geq s(f, \mathcal{D}_1 \cup \mathcal{D}_2) \geq s(f, \mathcal{D}_2). \qquad \bigstar$$

The inequality ★ tells us that, whatever dissection you pick and whatever dissection I pick, your lower sum can not exceed my upper sum. There is no way we can put a quart in a pint pot and the Banach-Tarski phenomenon is avoided.

Since $S(f, \mathcal{D}) \geq -(b-a)K$ for all dissections $\mathcal{D}$ we can define the *upper integral* $I^*(f) = \inf_{\mathcal{D}} S(f, \mathcal{D})$. We define the *lower integral* similarly as $I_*(f) = \sup_{\mathcal{D}} s(f, \mathcal{D})$. The inequality ★ tells us that these concepts behave well.

Lemma 9.2. *If $f : [a, b] \to \mathbb{R}$ is bounded then $I^*(f) \geq I_*(f)$.*

If $I^*(f) = I_*(f)$ we say that f is Riemann integrable and we write

$$\int_a^b f(x) \, dx = I^*(f).$$

We write $\mathcal{R}[a, b]$ or sometimes just $\mathcal{R}$ for the set of Riemann integrable functions on $[a, b]$.

The following lemma provides a convenient criterion for Riemann integrability.

Lemma 9.3. *(i) A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if given any $\epsilon > 0$ we can find a dissection $\mathcal{D}$ with*

$$S(f, \mathcal{D}) - s(f, \mathcal{D}) < \epsilon.$$

(ii) A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable with integral I if and only if given any $\epsilon > 0$ we can find a dissection $\mathcal{D}$ with

$$S(f, \mathcal{D}) - s(f, \mathcal{D}) < \epsilon, \ \text{and} \ S(f, \mathcal{D}) \geq I \geq s(f, \mathcal{D}).$$

The reader who is tempted to start her treatment of the Riemann integral with Lemma 9.3 (ii) should reflect that without the inequality ★ she will find it hard to prove the uniqueness of I.

Even though we admit that we can not expect all functions to be integrable, any satisfactory theory of integration must have a large collection of integrable functions.

Theorem 9.4. *(i) Any monotonic (that is any increasing or decreasing) function $f : [a, b] \to \mathbb{R}$ is Riemann integrable.*
(ii) Any continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable.

Part (i) of Theorem 9.4 looks a little odd. However most of the functions f met with in 'every day life' including many discontinuous ones are the difference of two increasing functions (that is $f = f_1 - f_2$ with f_1 and f_2 increasing). Such functions are called 'functions of bounded variation' and have a very pretty, if old fashioned, theory of their own. Lemma 9.7 tells us that every function of bounded variation is Riemann integrable. It is worth noting that Riemann did not have the theorem that every continuous function on a closed bounded interval is uniformly continuous and so was unable to prove that every continuous function is Riemann integrable.

The reader will recall from course C5/6 the standard example of bounded function which is not Riemann integrable.

Example 9.5. *If $f : [0, 1] \to \mathbb{R}$ is given by*

$$f(x) = 0 \qquad \text{when } x \text{ is rational,}$$
$$f(x) = 1 \qquad \text{when } x \text{ is irrational,}$$

then f is not Riemann integrable.

The following results are easy to prove.

Lemma 9.6. *Suppose that $a \le c \le b$ and $f : [a, b] \to \mathbb{R}$. If $f|_{[a,c]} \in \mathcal{R}[a, c]$ and $f|_{[c,b]} \in \mathcal{R}[c, b]$ then $f \in \mathcal{R}[a, b]$ and*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

If we adopt the convention that $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$ then the formula

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

holds independently of the order of a, b and c.

Lemma 9.7. *If $\lambda, \mu \in \mathbb{R}$ and $f, g : [a, b] \to \mathbb{R}$ are Riemann integrable then so is $\lambda f + \mu g$ and*

$$\int_a^b \lambda f(x) + \mu g(x)\,dx = \lambda \int_a^b f(x)\,dx + \mu \int_a^b g(x)\,dx.$$

In the language of linear algebra $\mathcal{R}[a, b]$ is a vector space and the integral is a linear functional (i.e. a linear map from $\mathcal{R}[a, b]$ to $\mathbb{R}$.).

Lemma 9.8. *If $f, g : [a, b] \to \mathbb{R}$ are Riemann integrable and $f(x) \geq g(x)$ for all $x \in [a, b]$ then*

$$\int_a^b f(x) \geq \int_a^b g(x)\, dx.$$

Lemma 9.9. *If $f : [a, b] \to \mathbb{R}$ is Riemann integrable then so is $|f|$.*

The next result requires a small idea.

Lemma 9.10. *(i) If $f : [a, b] \to \mathbb{R}$ is Riemann integrable then so is f^2.*
(ii) If $f, g : [a, b] \to \mathbb{R}$ are Riemann integrable then so is fg.

We are now in a position to prove the fundamental theorem of the calculus.

Theorem 9.11 (Fundamental theorem of the calculus). *Suppose that $u \in (a, b)$ and $f : (a, b) \to \mathbb{R}$ is a Riemann integrable function function which is continuous at some $t \in (a, b)$. If we set*

$$F(s) = \int_u^s f(x)\, dx$$

then F is differentiable at t and $F'(t) = f(t)$.

Sometimes we think of the fundamental theorem in a slightly different way.

Theorem 9.12. *Suppose that $f : (a, b) \to \mathbb{R}$ is continuous, that $u \in (a, b)$ and $c \in \mathbb{R}$. Then there is a unique solution to the differential equation $g'(t) = f(t)$ $[t \in (a, b)]$ such that $g(u) = c$.*

Thus (under appropriate circumstances) integration and differentiation are inverse operations and the the theories of differentiation and integration are subsumed in the greater theory of the calculus. Under appropriate circumstances, if the graph of F has tangent with slope $f(x)$ at x

area under the graph of slope of tangent of F
$$= \text{area under the graph of } f$$
$$= \int_a^b f(x)\, dx = \int_a^b F'(x)\, dx = F(b) - F(a).$$

Although it is not explicitly on the syllabus the examiners seem deeply attached to the next theorem. The result was stated in course C4 and the proof makes use of quite a lot of the methods developed in the present course.

Theorem 9.13 (Differentiation under the integral). *Suppose $g : [a, b] \times [c, d]$ is continuous and that the partial derivative g_2 exists and is continuous. Then writing $G(y) = \int_a^b g(x, y) \, dx$ we have G differentiable on (c, d) with*

$$G'(y) = \int_a^b g_{,2}(x, y) \, dx.$$

In the last few years the examiners have shown an excessive interest in the next lemma probably because the proof neatly combines several of our previous results.

Lemma 9.14. *Suppose that $g : \mathbb{R}^2 \to \mathbb{R}$ is continuous and has continuous partial derivative $g_{,2}$.*
(i) If we set $G(x, y) = \int_0^x g(u, y) \, du$ then G is differentiable with partial derivatives

$$G_{,1}(x, y) = g(x, y) \text{ and } G_{,2}(x, y) = \int_0^x g_{,2}(u, y) \, du.$$

(ii) If we set $F(t) = \int_0^t g(u, t) \, du$ then F is differentiable with derivative

$$F'(t) = g(t, t) + \int_0^t g_{,2}(u, t) \, du.$$

10 Further remarks on integration

We have defined Riemann integration for bounded functions on bounded intervals. However, the reader will already have evaluated, as a matter of routine, integrals in the following manner

$$\int_0^1 x^{-1/2} \, dx = \lim_{\epsilon \to 0+} \int_\epsilon^1 x^{-1/2} \, dx = \lim_{\epsilon \to 0+} [2x^{1/2}]_\epsilon^1 = 2,$$

and

$$\int_1^\infty x^{-2} \, dx = \lim_{R \to \infty} \int_1^R x^{-2} \, dx = \lim_{R \to \infty} [-x^{-1}]_1^R = 1.$$

A full theoretical treatment of such integrals with the tools at our disposal is apt to lead into a howling wilderness of 'infinite integrals of the first kind', 'Cauchy principal values' and so on. Instead I shall give a few typical theorems, definitions and counter-examples from which the reader should be able to reconstruct any parts of the theory that she needs.

Definition 10.1. *If $f : [a, \infty) \to \mathbb{R}$ is such that $f|_{[a,X]} \in \mathcal{R}[a, X]$ for each $X > a$ and $\int_a^X f(x) \, dx \to L$ as $X \to \infty$ then we say that $\int_a^\infty f(x) \, dx$ exists with value L.*

Lemma 10.2. *Suppose $f : [a, \infty) \to \mathbb{R}$ is such that $f|_{[a,X]} \in \mathcal{R}[a, X]$ for each $X > a$. If $f(x) \geq 0$ for all x then $\int_a^\infty f(x) \, dx$ exists if and only if there exists a K such that $\int_a^X f(x) \, dx \leq K$ for all X.*

Lemma 10.3. *Suppose $f : [a, \infty) \to \mathbb{R}$ is such that $f|_{[a,X]} \in \mathcal{R}[a, X]$ for each $X > a$. If $\int_a^\infty |f(x)| \, dx$ exists then $\int_a^\infty f(x) \, dx$ exists.*

It is natural to state Lemma 10.3 as saying that 'absolute convergence of the integral implies conditional convergence'.

Additional problems arise when there are two limits involved.

Example 10.4. *If $\lambda, \mu > 0$ then*

$$\int_{-\mu R}^{\lambda R} \frac{x}{1 + x^2} \, dx \to \log \frac{\lambda}{\mu}$$

as $R \to \infty$.

A pure mathematician gets round this problem by making a definition along these lines.

Definition 10.5. *If $f : \mathbb{R} \to \mathbb{R}$ is such that $f|_{[-X,Y]} \in \mathcal{R}[X, Y]$ for each $X, Y > 0$ $\int_{-\infty}^\infty f(x) \, dx$ exists with value L if and only if the following condition holds. Given $\epsilon > 0$ we can find an $X_0(\epsilon) > 0$ such that*

$$\left| \int_{-X}^Y f(x) \, dx - L \right| < \epsilon.$$

for all $X, Y > X_0(\epsilon)$.

The physicist gets round the problem by ignoring it. If she is a *real physicist* with correct physical intuition this works splendidly[4] but if not, not.

Speaking broadly, infinite integrals $\int_E f(x) \, dx$ work well when they are absolutely convergent, that is $\int_E |f(x)| \, dx < \infty$, but are full of traps for the unwary otherwise.

---

[4]In [2] Boas reports the story of a friend visiting the Princeton common room ' ... where Einstein was talking to another man, who would shake his head and stop him; Einstein then thought for a while, then started talking again; was stopped again; and so on. After a while, ... my friend was introduced to Einstein. He asked Einstein who the other man was. "Oh," said Einstein, "that's my mathematician." '

So far we have dealt only with the integration of functions $f : E \to \mathbb{R}$ with E a 'well behaved' subset of $\mathbb{R}$. It is not difficult to extend our definitions to the full multidimensional case but the main users of multidimensional integrals are either applied mathematicians (in the broadest sense), differential geometers and analysts (including probabilists). The applied mathematicians are happy to take the details on trust and both the differential geometers and the analysts require more sophisticated approaches (though of rather different kinds).

However the particular case of a function $f : [a, b] \to \mathbb{R}^m$ is needed in the complex variable courses (with $m = 2$) and elsewhere when line integrals are used. The definition is simple.

Definition 10.6. *If $f : [a, b] \to \mathbb{R}^m$ is such that $f_j : [a, b] \to \mathbb{R}$ is Riemann integrable for each j then $\int_a^b f(x)\, dx = \mathbf{y}$ where $\mathbf{y} \in \mathbb{R}^m$ and*

$$y_j = \int_a^b f_j(x)\, dx.$$

It is easy to obtain the properties of this integral directly from its definition. Here is an example.

Lemma 10.7. *If $\alpha : \mathbb{R}^m \to \mathbb{R}^p$ is linear and $f : [a, b] \to \mathbb{R}^m$ is Riemann integrable then so is $\alpha \circ f$ and*

$$\int_a^b \alpha(f(x))\, dx = \alpha \left(\int_a^b f(x)\, dx \right).$$

Taking α to be any orthogonal transformation of $\mathbb{R}^m$ to itself we see that that our definition is, in fact, coordinate independent. Choosing a particular orthogonal transformation we obtain the following nice result.

Theorem 10.8. *If $f : [a, b] \to \mathbb{R}^m$ is Riemann integrable then*

$$\left\| \int_a^b f(x)\, dx \right\| \le (b - a) \sup_{x \in [a,b]} \| f(x) \|.$$

This result and its extensions are often quoted in the form

$$\text{integral} \le \text{length} \times \sup.$$

Exercise 10.9. *Show that the collection $\mathcal{R}$ of Riemann integrable functions $f : [a, b] \to \mathbb{R}^m$ form a real vector space. If we write*

$$Tf = \int_a^b f(x)\, dx$$

show that $T : \mathcal{R} \to \mathbb{R}$ is a linear map and $|Tf| \le (b - a)\|f\|$.

The Riemann integral is simple to define and easy to use. So long as we only need to integrate continuous functions (or functions continuous except at a few simple discontinuities) it is perfectly satisfactory. The great majority of mathematicians need nothing else. However, the modern theory of probability and many parts of modern analysis need the more powerful integral of Lebesgue (a hint as to why this might be is given in Exercise 13.9) and its relatives. This is just as easy to use but substantially harder to set up[5].

11 Metric spaces

One of the unifying ideas of modern analysis is that of distance. We start with a definition modelled on those properties of Euclidean distance which do not refer to vector space structures. (Thus you should both *compare* and *contrast* Lemma 4.1.)

Definition 11.1. *We say that (X, d) is a metric space if X is a set and $d : X^2 \to \mathbb{R}$ is a function with the following properties:-*
   (i) $d(x, y) \geq 0$ for all $x, y \in X$.
   (ii) $d(x, y) = 0$ if and only if $x = y$.
   (iii) $d(x, y) = d(y, x)$ for all $x, y \in X$.
   (iv) (The triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$

As ought to be the case, we have the following result.

Lemma 11.2. *If $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ then $(\mathbb{R}^m, d)$ is a metric space.*

I shall give later some examples of metric spaces which are both useful and novel (in the context of this course). The metric of Lemma 11.2 is useful but hardly novel. In the interest of balance let me give a metric which will be novel to most of my readers even if it is not useful.

Example 11.3. *If $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\|$ when $\mathbf{x} \neq \mathbf{y}$ and $d(\mathbf{x}, \mathbf{x}) = 0$ then $(\mathbb{R}^m, d)$ is a metric space.*

(This metric is called the British Railway metric.) Here is an even odder railway metric.

Example 11.4. *If $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent (in more geometrical language) if $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{0}$ are on the same straight line) then set*

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

---

[5]I know of two universities which teach undergraduates Lebesgue theory in their first year but of no university where undergraduates learn Lebesgue theory in their first year.

Otherwise, set

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\|.$$

With this definition, $(\mathbb{R}^m, d)$ is a metric space.

Surprising as it may seem the following rather trivial metric is a rich source of intuition for communication theory. (However it is of no analytic interest.)

Lemma 11.5 (Hamming metric). *Let M be the space of sequences*

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

with $x_j = 0$ or $x_j = 1$. We call M the space of messages of length n. If $\mathbf{x}$ and $\mathbf{y}$ are messages we define

$$d(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} |x_j - y_j|$$

to be the Hamming distance *between the two messages. (M, d) is a metric space.*

Much of the 'mere algebra' which we did for the Euclidean metric carries over with hardly any change to the general metric case. Compare the following definition with Definition 4.2.)

Definition 11.6. *Let (X, d) be a metric space. If $a_n \in X$ for each $n \geq 1$ and $a \in X$ then we say that $a_n \to a$ if given $\epsilon > 0$ we can find an $n_0(\epsilon)$ such that*

$$d(a_n, a) < \epsilon \text{ for all } n \geq n_0(\epsilon).$$

However, we are not now dealing with a 'norm on a vector space' so some results do not carry over. The result corresponding to Lemma 4.3 has far fewer parts.

Lemma 11.7. *Let (X, d) be a metric space.*
(i) The limit is unique. That is, if $a_n \to a$ and $a_n \to b$ as $n \to \infty$ then $a = b$.
(ii) If $a_n \to a$ as $n \to \infty$ and $n(1) < n(2) < n(3) \ldots$ then $a_{n(j)} \to a$ as $j \to \infty$.
(iii) If $a_n = c$ for all n then $a_n \to c$ as $n \to \infty$.

The material on open and closed sets from Section 5 goes through essentially unchanged.

Definition 11.8. *Let (X, d) be metric space. A set $F \subseteq X$ is closed if whenever $x_n \in F$ for each n and $x_n \to x$ as $n \to \infty$ then $x \in F$.*

Definition 11.9. *Let (X, d) be metric space A set $U \subseteq X$ is open if whenever $x \in U$ there exists an $\epsilon > 0$ such that whenever $d(x, y) < \epsilon$ we have $y \in U$.*

Example 11.10. *Let (X, d) be metric space. Let $x \in X$ and $r > 0$.*
(i) The set $B(x, r) = \{y \in X : d(x, y) < r\}$ is open.
(ii) The set $\bar{B}(x, r) = \{y \in X : d(x, y) \leq r\}$ is closed.

We call $B(x, r)$ the open ball of radius r and centre x. We call $\bar{B}(x, r)$ the closed ball of radius r and centre x.

Lemma 11.11. *Let (X, d) be a metric space. A subset A of X is open if and only if each point of A is the centre of an open ball lying entirely within A.*

Definition 11.12. *The set N is a neighbourhood of the point x if we can find an $r > 0$ such that $B(x, r) \subseteq N$.*

Lemma 11.13. *Let (X, d) be a metric space. A subset A of X is open if and only if its complement $X \setminus A$ is closed.*

Lemma 11.14. *Let (X, d) be a metric space. Consider the collection τ of open sets in X.*
(i) $\emptyset \in \tau$, $X \in \tau$.
(ii) If $U_\alpha \in \tau$ for all $\alpha \in A$ then $\bigcup_{\alpha \in A} U_\alpha \in \tau$.
(iii) If $U_1, U_2, \ldots, U_n \in \tau$ then $\bigcap_{j=1}^n U_j \in \tau$.

Lemma 11.15. *Let (X, d) be a metric space. Consider the collection $\mathcal{F}$ of closed sets in X.*
(i) $\emptyset \in \mathcal{F}$, $X \in \mathcal{F}$.
(ii) If $F_\alpha \in \mathcal{F}$ for all $\alpha \in A$ then $\bigcap_{\alpha \in A} F_\alpha \in \mathcal{F}$.
(iii) If $F_1, F_2, \ldots, F_n \in \mathcal{F}$ then $\bigcup_{j=1}^n F_j \in \mathcal{F}$.

The new definition of continuity only breaks the chain of translations slightly because it involves *two* metric spaces.

Definition 11.16. *Let (X, d) and (Z, ρ) be metric spaces. We say that a function $f : X \to Z$ is continuous at some point $x \in X$ if given $\epsilon > 0$ we can find a $\delta(\epsilon, x) > 0$ such that if $y \in X$ and $d(x, y) < \delta(\epsilon, x)$ we have*

$$\rho[(f(x), f(y)) < \epsilon.$$

If f is continuous at every point $x \in X$ we say that f is a continuous function on X.

The reader may feel that Definition 5.12 is more general in than Definition 11.16 because it involves a set E.

The following remark shows that this is not so.

Lemma 11.17. *Let (X, d) be a metric space and $E \subseteq X$. Let $d_E : E^2 \to \mathbb{R}$ be given by $d_E(u, v) = d(u, v)$ whenever $u, v \in E$. Then (E, d_E) is a metric space.*

We conclude this section with more results taken directly from Section 5

Lemma 11.18. *Let (X, d) and (Z, ρ) be metric spaces and suppose that the function $f : X \to Z$ is continuous at $x \in X$. Then if $x_n \in X$ and $x_n \to x$ it follows that $f(x_n) \to f(x)$.*

Lemma 11.19. *Let (X, d) and (Z, ρ) be metric spaces. The function $f : X \to Z$ is continuous if and only if $f^{-1}(O)$ is open whenever O is open.*

Lemma 11.20. *Let (X, d) (Y, θ) and (Z, ρ) be metric spaces. If $f : X \to Y$ and $g : Y \to Z$ are continuous then so is their composition $g \circ f$.*

Question 21.1 which asks you to look at open balls and open sets in the two railway metrics given in Examples 11.3 and 11.4 emphasises the fact that open sets in different metrics may be very different.

12   A look to the future

In this short starred section I look at a very important class of metrics. The discussion will be informal, as indicated by the replacement of the word 'Theorem' by 'Statement'.

Suppose we wish to build a road joining two points of the plane. The cost of a building will depend on the nature of the terrain. If the cost of building a short stretch of length δs near a point (x, y) is (to first order) $g(x, y)\delta s$ then subject to everything being well behaved the cost of road Γ will be

$$\int_\Gamma g(x, y) ds.$$

'But' cries the reader 'you have not defined line integrals yet!' To which I reply 'Mathematical discourse takes place on many levels and we must learn to move freely between them'[6]. It is tempting to define the distance $d(A, B)$ between two points A and B in the plane as

$$d(A, B) = \inf \left\{ \int_\Gamma g(x, y) ds : \Gamma \text{ joining } A \text{ and } B \right\}.$$

If we include amongst our conditions that g is continuous and $g(x, y) > 0$ everywhere we see that d is a metric.

Statement 12.1. *(i) $d(A, B) \geq 0$ for all A and B.*
(ii) $d(A, B) = 0$ if and only if $A = B$.
(iii) $d(A, B) = d(B, A)$ for all A and B.
(iv) $d(A, C) \leq d(A, B) + d(B, C)$ for all A, B and C.

I have rather carefully kept to 'inf' rather than to 'min'. One problem is that I have not specified the kind of Γ that is permissible. (Notice that the argument we used to prove Statement 12.1 means that we could not just use 'smooth'.) However, it can be shown that if we choose a suitable class for the possible Γ and a suitably well behaved g then the minimum is attained. If Γ_0 is a path from A to B such that

$$\int_{\Gamma_0} g(x, y) ds = d(A, B)$$

we call Γ_0 a geodesic. (The geodesic need not be unique, consider road building for two towns at diametrically opposite points of a circular marsh.) If any two points are joined by a geodesic then $d(A, B)$ is the 'length of the geodesic path joining A and B' where length refers to the metric d and not to the Euclidean metric.

Let us try to use these ideas to find a metric on the upper half-plane

$$H = \{z \in \mathbb{C} : \Im z > 0\}$$

which is invariant under Möbius transformation. (See course C1/2 for background). More precisely we want a metric d such that if T is a Möbius map mapping H to itself bijectively, then $d(Tz_1, Tz_2) = d(z_1, z_2)$ for all $z_1, z_2 \in H$. Of course no such map may exist but we shall see where the question leads.

Statement 12.2. *The set $\mathcal{H}$ of Möbius maps T such that $T|_H : H \to H$ is bijective is a subgroup of the group $\mathcal{M}$ of all Möbius transformations. The*

---

[6]'Shut up!' he explained.

subgroup $\mathcal{H}$ is generated by the transformations T_a with $a \in \mathbb{R}$, D_λ with $\lambda > 0$ and J where

$$
\begin{aligned}
T_a &= a + z \\
D_\lambda(z) &= \lambda z \\
J(z) &= - z^{-1}
\end{aligned}
$$

By looking at the effect of elements of $\mathcal{H}$ on a small arc, we are led to the following. (Throughout we write write $z = x + iy$ and identify the plane $\mathbb{R}^2$ with $\mathbb{C}$ in the usual manner.)

Statement 12.3. *Suppose $g : H \to \mathbb{R}$ is well behaved. Then*

$$
\int_{T(\Gamma)} g(x, y) ds = \int_\Gamma g(x, y) ds
$$

for all nice paths Γ in H and all $T \in \mathcal{H}$ if and only if $g(x, y) = Ay^{-1}$ for some $A > 0$.

Let us now fix $g(x, y) = y^{-1}$ and let d be the derived metric.

Statement 12.4. *The metric just described is invariant under Möbius transformation.*

We can use the calculus of variations (course C10, Mathematical Methods) to find geodesics.

Statement 12.5. *The geodesics for the metric just described are arcs of circles with centres on the real axis and straight lines which cut the real axis at right angles.*

Those who are interested in the axiom of parallels for Euclidean geometry will note that, for (H, d), if we call circles with centres on the real axis and straight lines which cut the real axis at right angles geodesic lines for (H, d), then given any geodesic line Γ and any point A in H not on Γ we can find many geodesic lines through A which do not intersect Γ. More details are given in course D1 (Geometry).

Of course the methods of C10 are indicative rather than conclusive. (We have used a *necessary* condition and not a *sufficient* one.) However, in this particular case, it is not hard, once the answer is known, to state and prove everything rigorously. If we want general underpinning theorems they can be found in Differential Geometry but require a lot of hard work.

There is an interesting analogue of this kind of metric for groups.

Lemma 12.6. *Let G be a group and A a subset of G generating G. Write $A^{-1} = \{a^{-1} : a \in A\}$. If we set*

$$d(x,y) = \min\{n : \text{we can find } a_1, a_2, \ldots, a_n \in A \cup A^{-1} \text{ with } a_1 a_2 \ldots a_n = xy^{-1}\},$$

then d is a metric on G with $d(xz, yz) = d(x,y)$ for all $x, y, z \in G$. (Thus d is a right translation invariant metric.)

If $d(x,y)$ is bounded we call $\max d(x,y)$ the *diameter* of the group. The notion of diameter has an obvious relevance to the problem 'How many shuffles do we need to produce a well shuffled pack?'.

13 Completeness

If we examine the arguments of Section 11 we see that they are all mere algebra. What do we need to introduce to do genuine analysis on metric spaces. We can not use a variant of the fundamental axiom because there is no order on our spaces[7]. The correct variant of the Bolzano–Weierstrass method is, it is generally agreed, the notion of compactness which will be studied in the context of topological spaces (a concept more general than metric spaces) in course C 12. Instead we use a generalisation of the general principle of convergence.

Definition 13.1. *If (X, d) is a metric space we say that a sequence of points $x_n \in X$ is Cauchy if given any $\epsilon > 0$ we can find $n_0(\epsilon)$ such that $d(x_p, x_q) < \epsilon$ for all $p, q \geq n_0(\epsilon)$.*

Definition 13.2. *A metric space (X, d) is complete if every Cauchy sequence converges.*

Of course, $\mathbb{R}^n$ with the Euclidean metric is complete.

To show that the notion of completeness is distinct from 'Bolzano–Weierstrass notions' we introduce a dull but useful metric space.

Lemma 13.3. *Let X be any set. If we define $d : X^2 \to \mathbb{R}$ by*

$$
\begin{aligned}
d(x,y) &= 1 \qquad \text{if } x \neq y \\
d(x,x) &= 0
\end{aligned}
$$

then (X, d) is a metric space.

---

[7]There is an appropriate theory for objects with order (lattices) but we shall not pursue it here.

We call the metric d of the previous lemma the *discrete metric*.

Lemma 13.4. *Let (X, d) be a space with the discrete metric. Then (X, d) is complete.*

Suppose X is infinite. Then although X is bounded (if $x_0 \in X$ then $\bar{B}(x_0, 1) = X$) we can find a sequence x_n such that no subsequence converges.

Here is another property of the discrete metric.

Lemma 13.5. *If (X, d) is a space with the discrete metric then every subset of X is both open and closed.*

The contraction mapping theorem (Theorem 15.1) and its applications will provide a striking example of the utility of the concept of completeness. However this section and the next are devoted more to examples of spaces which are and are not complete. Students who want to see completeness in action immediately should do the next, starred, example.

Exercise 13.6. *We say that a metric space (X, d) has no isolated points if given $y \in X$ and $\epsilon > 0$ we can find an $x \in X$ such that $0 < d(x, y) < \epsilon$. Show by the methods of Exercise 3.14 that a non-empty metric space with no isolated points is uncountable.*

Give an example of an infinite countable metric space. Give an example of an uncountable metric space all of whose points are isolated.

The next lemma gives a good supply of metric spaces which are complete and of metric spaces which are not complete.

Lemma 13.7. *Let (X, d) be a complete metric space. If E is a subset of X and we define $d_E : E^2 \to \mathbb{R}$ by $d_E(u, v) = d(u, v)$ whenever $u, v \in E$ then (E, d_E) is complete if and only if E is closed in (X, d).*

Thus, for example, the closed interval $[a, b]$ is complete for the usual metric but the open interval (a, b) is not.

Here is a more interesting example of a metric which is not complete.

Example 13.8. *Consider $C([-1, 1])$ the set of continuous functions $f : [-1, 1] \to \mathbb{R}$. If we set*

$$d(f, g) = \|f - g\|_1 = \int_{-1}^{1} |f(x) - g(x)| \, dx$$

then d is a metric. Let

$$\begin{aligned}
f_n(x) &= -1 \qquad \text{for } -1 \le x \le -1/n \\
f_n(x) &= nx \qquad \text{for } -1/n \le x \le 1/n \\
f_n(x) &= 1 \qquad \text{for } 1/n \le x \le 1.
\end{aligned}$$

The sequence f_n is Cauchy but has no limit. Thus $(C([-1,1]), d)$ is not complete.

The example just given leaves open the possibility that the same type of metric might be be complete when applied to Riemann integrable functions. The following example which is long, heavily starred, and only for the ambitious shows that this is not so.

Exercise 13.9. *Consider $\mathcal{R}([-1,1])$ the set of Riemann integrable functions functions $f : [-1,1] \to \mathbb{R}$. If we set*

$$d(f,g) = \|f - g\|_1 = \int_{-1}^{1} |f(x) - g(x)| \, dx$$

show that d satisfies conditions (i), (iii) and (iv) of the definition of a metric (see Definition 11.1) but give an example to show that condition (ii) may fail. We could now stop since d is not a metric but this would be to miss the interesting part.

In many ways, condition (ii) is the least important part of the definition of a metric. We call something satisfying all the other conditions a pseudo-metric. *We can define limits for a pseudo-metrics in the same way as for metrics.*

If (X, ρ) is a pseudo-metric space show that if $x_n \to x$ then $x_n \to y$ if and only if $\rho(x,y) = 0$.

In the same way we can define Cauchy sequences. A pseudo-metric space is called complete if every Cauchy sequence converges. Our object is to show that $(\mathcal{R}([-1,1]), d)$ is not complete.

(i) Set

$$\Delta_n(x) = 0 \qquad \text{if } 2^{-3n} \le |x|$$
$$\Delta_n(x) = 1 - 2^{3n}|x| \qquad \text{if } |x| \le 2^{-3n}.$$

and $g_n(t) = \sum_{r=-2^n}^{2^n} \Delta_n(t - r2^{-n})$ for all $-1 \le t \le 1$. Sketch g_n and show that $\|g_n\|_1 \le 2^{-2n+1}$.

(ii) If $f_n(x) = \max\{f_1(x), f_2(x), \ldots, f_n(x)\}$ explain briefly why f_n is a continuous function on $[-1,1]$ with $0 \le f_n(x) \le 1$. Show that $\|f_{n+1} - f_n\| \le 2^{-2n+3}$ and deduce that f_n is a Cauchy sequence in $(\mathcal{R}([-1,1]), d)$.

(iii) Suppose if possible that there exists an $f \in \mathcal{R}([-1,1])$ such that $\|f_n - f\| \to 0$ as $n \to \infty$. Show that $\|f\| \le 2/3$.

(iv) We now use the notation of Section 9. Explain why there must exist a dissection $\mathcal{D}$ of $[-1,1]$ such that $S(f, \mathcal{D}) \le 1$. Deduce that we can find a and b with $-1 \le a < b \le 1$ such that $f(x) \le 1/2$ for all $x \in [a,b]$.

(v) Show that $\int_a^b |f(x) - f_n(x)|\, dx \nrightarrow 0$ and deduce that $\|f - f_n\|_1 \nrightarrow$ 0. This contradiction shows that the sequence f_n does not converge and $(\mathcal{R}([-1,1]), d)$ is not complete.

The reader who has got this far can feel well pleased with herself but by working still harder we can extract a little more juice from this example and show that f_n converges pointwise to a function which is not Riemann integrable.

(v) Show that, for each x, $f_n(x)$ is a bounded increasing function. Deduce that $f_n(x)$ tends to limit $F(x)$, say, for each x. Show that $I^*(F) = 2$.

(vi) Let $Z_n = f_n^{-1}(0)$. Explain why Z_n is closed and $Z_1 \supseteq Z_2 \supseteq \dots$. If $\mathcal{D}$ is a fixed dissection of $[-1,1]$ show that for each n we can find a collection of $\mathcal{I}_n$ of intervals $I = [x_j, x_{j-1}]$ with end points successive points of the dissection and of total length greater than 1 such that $I \cap Z_n \neq \emptyset$.

(vii) Let $\mathcal{I} = \bigcap_{n=1}^{\infty} \mathcal{I}_n$. Show that $\mathcal{I}$ is a collection of intervals $I = [x_j, x_{j-1}]$ with end points successive points of the dissection $\mathcal{D}$ and with total length greater than 1 such that $I \cap Z_n \neq \emptyset$ for each n.

(viii) Use Theorem 5.17 to show that $I \cap \bigcap_{n=1}^{\infty} Z_n \neq \emptyset$ whenever $I \in$ mathcalI. Deduce that $s(F, \mathcal{D}) \leq 1$.

(ix) Conclude that F is not Riemann integrable.

If we seek to embed $(\mathcal{R}([-1,1]), d)$ in some larger space of functions which is complete we are led to Lebesgue theory.

14   The uniform metric

This section is devoted to one of the most important metrics on functions. We shall write $\mathbb{F}$ to mean either $\mathbb{R}$ or $\mathbb{C}$.

Definition 14.1. *If E is a non-empty set we write $\mathcal{B}(E)$ for the set of bounded functions $f : E \to \mathbb{F}$. The uniform norm $\|\ \|_\infty$ on $\mathcal{B}$ is defined by $\|f\|_\infty = \sup_{x \in E} |f(x)|$.*

Lemma 14.2. *If we use the standard operations $\mathcal{B}(E)$ is a vector space over $\mathbb{F}$. If $f, g, h \in \mathcal{B}(E)$ the following results are true.*
    (i) $\|f\|_\infty \geq 0$,
    (ii) If $\|f\|_\infty = 0$ then $f = 0$,
    (iii) If $\lambda \in \mathbb{F}$ then $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty$.
    (iv) (The triangle inequality) $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.
    (v) The pointwise product $fg \in \mathcal{B}(E)$ (where $fg(x) = f(x)g(x)$) and $\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$.

We call the metric d given by $d(f, g) = \|f - g\|_\infty$ the uniform metric.

Theorem 14.3. *The uniform metric on $\mathcal{B}(E)$ is complete.*

The space $\mathcal{B}(E)$ is not very interesting in itself but if E is a metric space it has a very interesting subspace.

Definition 14.4. *If (E,d) is a non-empty metric space we write $\mathcal{C}(E)$ for the set of bounded continuous functions $f : E \to \mathbb{F}$.*

The next remark merely restates what we already know.

Lemma 14.5. *If (E,d) is a non-empty metric space then $\mathcal{C}(E)$ is a vector subspace of $\mathcal{B}(E)$. Further if $f, g \in \mathcal{C}(E)$ the pointwise product $fg \in \mathcal{C}(E)$.*

However the next result is new and crucial.

Theorem 14.6. *If (E,d) is a non-empty metric space then $\mathcal{C}(E)$ is a closed subset of $\mathcal{B}(E)$ under the uniform metric.*

This is the famous '$\epsilon/3$ Theorem'[8]. Traditionally, it has been presented in a different but essentially equivalent manner.

Definition 14.7 (Uniform convergence). *If E is a non-empty set and $f_n : E \to \mathbb{F}$ and $f : E \to \mathbb{F}$ are functions we say that f_n converges uniformly to f as $n \to \infty$ if, given any $\epsilon > 0$ we can find an $n_0(\epsilon)$ such that $|f_n(x) - f(x)| < \epsilon$ for all $x \in E$ and all $n \geq n_0(\epsilon)$.*

Theorem 14.6 now takes the following form.

Theorem 14.8. *If (E,d) is a non-empty metric space and $f_n : E \to \mathbb{F}$ form a sequence of continuous functions tending uniformly to f then f is continuous.*

More briefly, the uniform limit of continuous functions is continuous. (To see the exact equivalence observe that if $f_n \to f$ uniformly then we can find an N such that $f_n - f_N$ is bounded for all $n \geq N$.)

Since $\mathcal{C}(E)$ is a closed subset of $\mathcal{B}(E)$, Theorem 14.3 and Lemma 13.7 gives us another important theorem.

Theorem 14.9. *If (E,d) is a non-empty metric space then the the uniform metric on $\mathcal{C}(E)$ is complete.*

This result, also, has a more traditional form.

---

[8]Or according to a rival school of thought the '3ϵ Theorem'.

Theorem 14.10 (General principle of uniform convergence). *Suppose that (E, d) is a non-empty metric space and $f_n : E \to \mathbb{F}$ is a continuous function $[n \geq 1]$. The sequence f_n converge uniformly to a continuous function f if and only if given any $\epsilon > 0$ we can find an $n_0(\epsilon)$ such that $|f_n(x) - f_m(x)| < \epsilon$ for all $n, m \geq n_0(\epsilon)$.*

This theorem is also known as the GPUC by those who do not object to theorems which sound as though they were a branch of the secret police.

If E is a closed bounded subset of $\mathbb{R}^m$ with the Euclidean metric then by Theorem 6.4 all continuous functions $f : E \to \mathbb{F}$ are bounded. In these circumstances we shall write $C(E) = \mathcal{C}$. For the rest of the course we will only consider this special case. We start with a couple of important examples.

Example 14.11 (The witch's hat). *Define $f_n : [0, 2] \to \mathbb{R}$ by*

$$
\begin{aligned}
f_n(x) &= 1 - n|x - n^{-1}| \qquad & for\ |x - n^{-1}| \leq n^{-1}, \\
f_n(x) &= 0 & otherwise.
\end{aligned}
$$

Then the f_n are continuous functions such that $f_n(x) \to 0$ as $n \to \infty$ for each x but $f_n \not\to 0$ uniformly.

More briefly, pointwise convergence does not imply uniform convergence.

Example 14.12. *Define $f_n : [0, 1] \to \mathbb{R}$ by $f_n(x) = x^n$. Then $f_n(x) \to f(x)$ as $n \to \infty$ where $f(x) = 0$ for $0 \leq x < 1$ but $f(1) = 1$.*

Thus the pointwise limit of continuous functions need not be continuous. Uniform convergence is a very useful tool when dealing with integration.

Theorem 14.13. *Let $f_n \in C([a, b])$. If $f_n \to f$ uniformly then $f \in C([a, b])$ and*

$$
\int_a^b f_n(x)\, dx \to \int_a^b f(x)\, dx.
$$

Students often miss the full force of this theorem because the formula is so easy to prove. We also need to prove that the second integral actually exists. The second half of Example 13.9 shows that the pointwise limit of continuous functions need not be integrable.

Theorem 14.13 should be considered in the context of the following two examples.

Example 14.14 (The tall witch's hat). *Define* $f_n : [0, 2] \to \mathbb{R}$ *by*

$$f_n(x) = n(1 - n|x - n^{-1}|) \qquad for \ |x - n^{-1}| \leq n^{-1},$$
$$f_n(x) = 0 \qquad\qquad\qquad\qquad otherwise.$$

Then the f_n are continuous functions such that $f_n(x) \to 0$ as $n \to \infty$ but

$$\int_0^2 f_n(x) \, dx \nrightarrow 0$$

as $n \to \infty$.

Example 14.15 (Escape to infinity). *Define* $f_n : \mathbb{R} \to \mathbb{R}$ *by*

$$f_n(x) = 1 - n^{-1}|x| \qquad for \ |x| \leq n,$$
$$f_n(x) = 0 \qquad\qquad\qquad otherwise.$$

Then the f_n are continuous functions such that $f_n(x) \to 0$ uniformly as $n \to \infty$ but

$$\int_{-\infty}^{\infty} f_n(x) \, dx \nrightarrow 0$$

as $n \to \infty$.

Example 14.15 must be well understood since it represents a very common phenomenon which we have to learn to live with.

Traditionally, Theorem 14.13 is always paired with the following result which is really a disguised theorem on integration!

Theorem 14.16. *Suppose that $f_n : [a, b] \to \mathbb{R}$ is differentiable on $[a, b]$ with continuous derivative f_n' (we take the one-sided derivative at end points if necessary). Suppose that $f_n(x) \to f(x)$ as $n \to \infty$ for each $x \in [a, b]$ and suppose that f_n' converges uniformly to a limit F on $[a, b]$. Then f is differentiable with derivative F.*

The reader knows how to turn results on the limits of sequences into results on infinite sums and vice versa. Applied to Theorems 14.13 and 14.16 the techniques produces the following results.

Theorem 14.17 (Term by term integration). *Let $g_j : [a, b] \to \mathbb{R}$ be continuous. If $\sum_{j=1}^{n} g_j$ converges uniformly as $n \to \infty$ then*

$$\int_a^b \sum_{j=1}^{\infty} g_j(x) \, dx = \sum_{j=1}^{\infty} \int_a^b g_j(x) \, dx.$$

Theorem 14.18 (Term by term differentiation). *Let $g_j : [a, b] \to \mathbb{R}$ be differentiable with continuous derivative. If $\sum_{j=1}^{n} g_j(x)$ converges for each x and $\sum_{j=1}^{n} g_j'$ converges uniformly as $n \to \infty$ then $\sum_{j=1}^{\infty} g_j$ is differentiable and*

$$\frac{d}{dx}\left(\sum_{j=1}^{\infty} g_j(x)\right) = \sum_{j=1}^{\infty} g_j(x).$$

Here is a starred application of Theorem 14.16. The reader may, of course omit it but I include it partly because it is often more useful than the theorem (Theorem 9.13) that it extends and partly because it provides an excellent example of how Theorem 14.16 is used in practice.

Theorem 14.19 (Differentiation under an infinite integral). *Suppose $g : [0, \infty] \times [c, d]$ is continuous and that the partial derivative $g_{,2}$ exists and is continuous. Suppose further that that there exists a continuous function $h : [0, \infty] \times [c, d]$ with $|g_{,2}(x, y)| \leq h(x)$ for all (x, y) and such that $\int_0^\infty h(x)\, dx$ exists and is finite. Then, if $G(y) = \int_0^\infty g(x, y)\, dx$ exists for all $y \in (c, d)$ we have G differentiable on (c, d) with*

$$G'(y) = \int_0^\infty g_{,2}(x, y)\, dx.$$

We conclude this section with a remark which may appear rather isolated but will become more important in course C12 and later work on complex variable.

Theorem 14.20. *We work in $\mathbb{C}$. Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series with radius of convergence R. If $0 \leq r < R$ then $\sum_{n=0}^{N} a_n z^n$ converges uniformly on $\bar{B}(0, r)$ as $N \to \infty$. However, it need not be true that $\sum_{n=0}^{N} a_n z^n$ converges uniformly on $B(0, R)$.*

15   The contraction mapping theorem

The next result is a beautiful example of the power of abstraction. In it Banach transformed a 'folk-technique' into a theorem.

Theorem 15.1 (The contraction mapping theorem). *Let (X, d) be a non-empty complete metric space and $T : X \to X$ a mapping such that there exists a $K < 1$ with $d(Tx, Ty) \leq K d(x, y)$ for all $x, y \in X$. Then there exists a unique $x_0 \in X$ such that $Tx_0 = x_0$.*

More briefly, a contraction mapping on a complete metric space has a unique fixed point.

Wide though the conditions are the reader should exercise caution before attempting to widen them further.

Example 15.2. *(i) If X is the unit circle centre $\mathbf{0}$ in the plane, d Euclidean distance and T a non-trivial rotation then (X, d) is a complete metric space and $d(T\mathbf{x}, T\mathbf{y}) = d(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in X$ but T has no fixed point.*
(ii) If $X = \mathbb{R}$, d is Euclidean distance and

$$Tx = 1 + x + g(x)$$

for some suitably chosen g then (X, d) is a complete metric space and $d(Tx, Ty) < d(x, y)$ for all $x, y \in X$ but T has no fixed point.

We use the contraction mapping theorem to show that a wide class of differential equations actually have a solution. We shall be looking at equations of the form

$$y' = f(x, y).$$

Our first, simple but important, result is that this problem on differential equations can be turned into a problem on integral equations.

Lemma 15.3. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous, $x_0, y_0 \in \mathbb{R}$ and $\delta > 0$ then the following two statements are equivalent.*
(A) The function $y : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ is differentiable and satisfies the equation $y'(x) = f(x, y(x))$ for all $x \in (x_0 - \delta, x_0 + \delta)$ together with the boundary condition $y(x_0) = y_0$.
(B) The function $y : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ is continuous and satisfies the condition

$$y(x) = y_0 + \int_{x_0}^{x} f(u, y(u)) \, du$$

for all $x \in (x_0 - \delta, x_0 + \delta)$.

Theorem 15.4. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous, $x_0, y_0 \in \mathbb{R}$ and $\delta > 0$. Suppose further that there exists a $K > 0$ such that $K\delta < 1$ and*

$$|f(x, u) - f(x, v)| \leq K|u - v| \qquad\qquad \bigstar$$

for all $u, v \in (x_0 - \delta, x_0 + \delta)$. Then there exists a unique $y : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ is continuous and satisfies the condition

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t)) \, dt$$

for all $x \in (x_0 - \delta, x_0 + \delta)$.

Condition ★ is called a Lipschitz condition. It closely related to differentiability (consider the mean value theorem) but does not imply differentiability (consider $f(x,y) = |y|$). In the absence of such a condition differential equations can have unexpected properties.

Example 15.5. *There are an infinity of different solutions to the equation*

$$y' = 3y^{2/3}$$

with $y(0) = 0$.

The lecturer can not bear to let the topic go at this point but the rest of this section is heavily starred. As it stands Theorem 15.4 is not quite powerful enough for practical purposes. Careful thought, without the need for new ideas shows that it can be extended to the following result.

Theorem 15.6. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function satisfying the following condition. There exists a $K : [0, \infty) \to \mathbb{R}$ such that*

$$|f(x, u) - f(x, v)| \leq K(R)|u - v|$$

whenever $|x|, |u|, |v| \leq R$. Then given any $(x_0, y_0) \in \mathbb{R}^2$ we can find a $\delta(x_0, y_0) > 0$ such that there exists a unique differentiable function $y : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ which satisfies the equation $y'(x) = f(x, y(x))$ for all $x \in (x_0 - \delta, x_0 + \delta)$ together with the boundary condition $y(x_0) = y_0$.

Theorem 15.6 tells us that the under very wide conditions the differential equation has a *local solution* through each (x_0, y_0). Does it have a *global solution*, that is, can we find a solution for the equation $y'(x) = f(x, y(x))$ which is defined for all $x \in \mathbb{R}$?

We finish our discussion by showing that the question is non-trivial.

Theorem 15.7. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function satisfying the following condition. There exists a $K : [0, \infty) \to \mathbb{R}$ such that*

$$|f(x, u) - f(x, v)| \leq K(R)|u - v|$$

for all $|x| \leq R$ and all u and v. Then given any $(x_0, y_0) \in \mathbb{R}^2$ we can find a unique $y : \mathbb{R} \to \mathbb{R}$ which is is differentiable and satisfies the equation $y'(x) = f(x, y(x))$ for all $x \in \mathbb{R}$ together with the boundary condition $y(x_0) = y_0$

Example 15.8. *If $f(x, y) = 1 + y^2$ then*

$$|f(x, u) - f(x, v)| \leq 2R|u - v|$$

whenever $|u|, |v| \leq R$. However, there does not exist a differentiable function $y : \mathbb{R} \to \mathbb{R}$ with $y' = f(x, y)$ for all x.

16 Green's function

(The contents of this section are all starred)

In course C12 you saw how to solve the differential equation

$$y'' + a(x)y' + b(x)y = f(x) \qquad (*)$$

subject to the conditions $y(0) = y(1) = 0$ by using the Green's function. There is no doubt in the present author's mind that the 'physicist's approach' using impulses, delta functions and hand-waving is the most important for the student to master. However, it adds to our confidence in the hand-waving approach to see that its results can be confirmed rigourously.

We start by looking at solutions of

$$y'' + a(x)y' + b(x)y = 0 \qquad (**)$$

Observe first that the results of the last section on the existence of solutions of differential equations can be extended to the vectorial case with only the most minor variations in the proof. Thus, for example, Theorem 15.7 extends as follows.

Theorem 16.1. *Suppose $f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ is a continuous function satisfying the following condition. There exists a $K : [0, \infty) \to \mathbb{R}$ such that*

$$\|f(x, \mathbf{w}) - f(x, \mathbf{z})\| \le K(R)\|\mathbf{w} - \mathbf{z}\|$$

for all $|x| \le R$ and all $\mathbf{w}$ and $\mathbf{z}$. Then given any $x_0 \in \mathbb{R}$ and $\mathbf{y}_0 \in \mathbb{R}^m$ we can find a unique $\mathbf{y} : \mathbb{R} \to \mathbb{R}^m$ which is is differentiable and satisfies the equation

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

for all $x \in \mathbb{R}$ together with the boundary condition $\mathbf{y}(x_0) = \mathbf{y}_0$.

As the the reader knows from course C4 (Differential Equations) it is easy to convert equation $(**)$ into a form appropriate for the application of Theorem 16.1. If we set

$$\mathbf{y} = \begin{pmatrix} y \\ y' \end{pmatrix} \text{ and } f(x, \mathbf{w}) = \begin{pmatrix} 0 & 1 \\ -a(x) & -b(x) \end{pmatrix} \mathbf{w} = \begin{pmatrix} w_2 \\ -b(x)w_1 - a(x)w_2 \end{pmatrix},$$

then

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

Now

$$\|f(x, \mathbf{w}) - f(x, \mathbf{z})\| = \left\| \begin{pmatrix} w_2 - z_2 \\ -b(x)(w_1 - z_1) - a(x)(w_2 - z_2) \end{pmatrix} \right\|$$
$$\leq (1 + |a(x)| + |b(x)|)\|\mathbf{w} - \mathbf{z}\|$$

so that writing $K(R) = \sup_{t \in [-R,R]}(1 + |a(t)| + |b(t)|)$ we have

$$\|f(x, \mathbf{w}) - f(x, \mathbf{z})\| \leq K(R)\|\mathbf{w} - \mathbf{z}\|$$

whenever $\|x| \leq K(R)$. Thus there is one and only one solution $y_{[\lambda]}$, say, of $(**)$ with $y(0) = 0$ and $y'(0) = \lambda$. By linearity, $y_{[\lambda]} = \lambda y_{[1]}$.

Let us write y_1 for the solution of $(**)$ with $y_1(0) = 0$, $y_1'(0) = 1$. We make the following

<div align="center">

key assumption $y_1(1) \neq 0$.

</div>

We write y_2 for the solution of $(**)$ with $y_2(0) = 0$, $y_2'(0) = 1$. The next lemmas are familiar from courses C4 and C10. We define

$$W(x) = y_1(x)y_2'(x) - y_2(x)y_1'(x).$$

The function W is called the Wronskian.

Lemma 16.2. *(i) The Wronskian is differentiable with*

$$W'(x) = -a(x)W(x).$$

(ii) The Wronskian is never zero.

We can now define the Green's function $G : \mathbb{R}^2 \to \mathbb{R}$ by

$$\begin{aligned} G(s,t) &= y_1(s)y_2(t)W(t)^{-1} && \text{for } s < t \\ G(s,t) &= y_2(s)y_1(t)W(t)^{-1} && \text{for } t \leq s. \end{aligned}$$

(Why should we do any such thing? Because the intuitive arguments of C10 tell us to.) Using Lemma 9.14 it is easy to verify our main result.

Theorem 16.3. *If $f : \mathbb{R} \to \mathbb{R}$ is continuous then*

$$y(x) = \int_0^1 G(x, w)f(w)\, dw$$

is twice differentiable and satisfies

$$y'' + a(x)y' + b(x)y = f(x) \tag{$*$}$$

together with the conditions $y(0) = y(1) = 0$.

(Our proof is very direct and involves neither differentiation under the integral nor Lemma 9.14. However, it is not hard to see that such ideas would be relevant in more complicated situations.)

From the point of view of more advanced work this shows how differential operators like

$$y \mapsto Sy = y'' + a(x)y + b(x)y$$

can be linked with better behaved integral operators like

$$f \mapsto Tf \text{ with } Tf(x) = \int_0^1 G(x,v)f(v)\,dv.$$

Note that we have shown $STf = f$ for f continuous, but note also that, if f is merely continuous, Sf need not be defined. The Green's function G is an example of an integral kernel[9]. More formally, if we write

$$Tf(x) = \int_0^1 K(x,v)f(v)\,dv,$$

then T is called an integral operator with kernel K. (So far as I know, there is no connection with the 'kernel' that you meet in the linear mathematics course P1.)

We end with an example to show that things really do go awry if our **key assumption** fails.

Example 16.4. *The equation*

$$y'' + \pi^2 y = x$$

has no solution satisfying $y(0) = y(1) = 0$.

---

[9]Most of the new words in this last paragraph are well worth dropping.

> You must lie among the daisies and discourse in novel phrases of your complicated state of mind,
> The meaning doesn't matter if it's only idle chatter of a transcendental kind.
> And everyone will say,
> As you walk your mystic way,
> 'If this young man expresses himself in terms too deep for me
> Why, what a very singularly deep young man this deep young man must be!'

17   Further reading

Since new introductions to analysis pour off the printing presses in an un-ending torrent, it may well be my fault that I can not recommend one book to cover the whole course. Spivak's *Calculus* [8] and J. C. Burkill's *A First Course in Mathematical Analysis* [3] are both excellent introductions to analysis but only deal with the real line and not with $\mathbb{R}^m$. J. C. Burkill and H. Burkill's *A Second Course in Mathematical Analysis* [4] is also good but rather old fashioned. (Of course, this has nothing to do with the authors' knowledge but a great deal to do with what was then thought the appropriate content for a second course.) Dieudonné's *Foundations of Modern Analysis* [5] is superb but written by someone who neither knew nor cared what others thought was the appropriate content for a second course[10]. A suitable compromise is reached in Marsden and Hoffman's *Elementary Classical Analysis* [7]. This covers the material on multidimensional calculus very clearly and is probably the book that most students will find most useful. Make sure that your college library has a copy of the second edition.

The material on metric spaces is covered in Sutherland's *Introduction to Metric and Topological Spaces* [9] which is workmanlike though a bit dull. (Once analytic topology is detached from its applications it takes a very gifted writer to make it exciting.)

I would be glad to hear of other suggestions for suitable books. A completely unsuitable but interesting version of the standard analysis course is given by Berlinski's *A Tour of the Calculus* [1] — Spivak rewritten by Sterne with additional purple passages by the Ankh-Morpork tourist board.

References

[1] D. Berlinski *A Tour of the Calculus* Mandarin Paperbacks 1997.

[2] R. P. Boas *Lion Hunting and Other Mathematical Pursuits* (Editors G. L. Alexander and D. H. Mugler), Dolciani Expositions, Vol 15, MAA, 1995.

[3] J. C. Burkill *A First Course in Mathematical Analysis* CUP, 1962.

[4] J. C. Burkill and H. Burkill *A Second Course in Mathematical Analysis* CUP, 1970.

---

[10]Boas notes that 'There is a test for identifying some of the future professional mathematicians at an early age. These are students who instantly comprehend a sentence beginning "Let X be an ordered quintuple $(a, T, \pi, \sigma, \mathcal{B})$ where … '. They are even more promising if they add, 'I never really understood it before." ' ([2] page231.)

[5] J. Dieudonné *Foundations of Modern Analysis*, Academic Press, 1960.

[6] G. H. Hardy *A Course of Pure Mathematics*, CUP, 1908. (Still available in its 10th edition.)

[7] J. E. Marsden and M. J. Hoffman *Elementary Classical Analysis* (2nd Edition), W. H. Freeman, New York, 1993.

[8] M. Spivak, *Calculus* Addison-Wesley/Benjamin-Cummings, 1967.

[9] W. A. Sutherland *Introduction to Metric and Topological Spaces*, OUP, 1975.

[10] S. Wagon *The Banach–Tarski Paradox*, CUP, 1993.

[11] E. T. Whittaker and G. N. Watson *A Course of Modern Analysis* CUP, 1902 (Still available in its 4th edition.)

18 First Sheet of Exercises

I have tried to produce 12 rather routine questions for each sheet. Any further questions are for interest only. Ambitious students and their supervisors should look at Tripos questions to supplement this meagre fare.

Q 18.1. (We work with the same ideas as in Example 1.1.) (i) Find a differentiable function $f : \mathbb{Q} \to \mathbb{Q}$ such that $f'(x) = 1$ for all $x \in \mathbb{Q}$ but $f(0) > f(1)$.

(ii) If we define $g : \mathbb{Q} \to \mathbb{Q}$ by $g(x) = (x^2 - 2)^{-2}$ show that g is continuous but unbounded on $\{x \in \mathbb{Q} : |x| \le 2\}$.

Q 18.2. In Question 18.3 below you are asked to prove Lemma 2.1. Write a paragraph on each of the following subjects.

(i) Would it have been better for the lecturer to prove Lemma 2.1 rather than leave it to you. Why?

(ii) What benefit, if any, do you expect to derive from doing Lemma 2.1? Give reasons.

(iii) What is the best way of approaching Question 18.1. For example should you look at your notes from last year before attacking the question? Should you look at your notes from last year after attacking the question?

Q 18.3. Prove Lemma 2.1 (restated below for your convenience)

(i) The limit is unique. That is, if $a_n \to a$ and $a_n \to b$ as $n \to \infty$ then $a = b$.

(ii) If $a_n \to a$ as $n \to \infty$ and $n(1) < n(2) < n(3) \ldots$ then $a_{n(j)} \to a$ as $j \to \infty$.

(iii) If $a_n = c$ for all n then $a_n \to c$ as $n \to \infty$.

(iv) If $a_n \to a$ and $b_n \to b$ as $n \to \infty$ then $a_n + b_n \to a + b$.

(v) If $a_n \to a$ and $b_n \to b$ as $n \to \infty$ then $a_n b_n \to ab$.

(vi) If $a_n \to a$ as $n \to \infty$ and $a_n \ne 0$ for each n, $a \ne 0$ then $a_n^{-1} \to a^{-1}$.

(vii) If $a_n \le A$ for each n and $a_n \to a$ as $n \to \infty$ then $a \le A$.

Q 18.4. Prove Lemma 2.2 which states that a decreasing sequence of real numbers bounded below tends to a limit.

Q 18.5. (i) If a_j is a integer with $0 \le a_j \le 9$ show *from the fundamental axiom* that

$$\sum_{j=1}^{\infty} a_j 10^{-j}$$

exists. Show that $0 \leq \sum_{j=1}^{\infty} a_j 10^{-j} \leq 1$ carefully quoting any theorems that you use.

(ii) If $0 \leq x \leq 1$ show that we can find integers x_j with $0 \leq x_j \leq 9$ such that

$$x = \sum_{j=1}^{\infty} x_j 10^{-j}.$$

What important deep result do you use?

(iii) If a_j and b_j are integers with $0 \leq a_j, b_j \leq 9$ and $a_j = b_j$ for $j < N$, $a_N > b_N$ show that

$$\sum_{j=1}^{\infty} a_j 10^{-j} \geq \sum_{j=1}^{\infty} b_j 10^{-j}.$$

Give the precise necessary and sufficient condition for equality and prove it.

Q 18.6. (i) Let us write

$$S_n = \sum_{r=0}^{n} \frac{1}{r!}.$$

Show, from first principles, that S_n converges to a limit (which, with the benefit of extra knowledge, we call e).

(ii) Show that, if $n \geq 2$ and $r \geq 0$ then

$$\frac{n!}{(n+r)!} \leq \frac{1}{3^r}.$$

Deduce carefully that, if $m \geq n \geq 2$,

$$0 \leq n!(S_m - S_n) \leq \frac{1}{2}.$$

and that

$$0 < n!(e - S_n) \leq \frac{1}{2}.$$

Deduce that $n!e$ is not an integer for any n and conclude that e is irrational.

(iii) Show similarly that $\sum_{r=0}^{\infty} \frac{1}{(2r)!}$ is irrational.

53

Q 18.7. In this question you should feel free to use any results you know (provided you quote them correctly).

(i) By using the mean value theorem, show that $(1+x^{-1})^x$ is an increasing function of x for $x \geq 0$.

(ii) Show that $(1 + n^{-1})^n$ tends to limit L as $n \to \infty$.

(iii) Show that, if $n \geq N$,

$$e \geq \left(1 + \frac{1}{n}\right)^n \geq \sum_{r=0}^{N} \binom{n}{r} n^{-r}.$$

Deduce that

$$e \geq L \geq \sum_{r=0}^{N} \frac{1}{r!},$$

and conclude that

$$\left(1 + \frac{1}{n}\right)^n \to e$$

as $n \to \infty$.

(iv) Assuming the Taylor expansion for e^t show that

$$\left(1 + \frac{t}{n}\right)^n \to e^t$$

as $n \to \infty$ for all $t > 0$.

(v) Show that

$$1 - \left(1 + \frac{t^2}{n^2}\right)^n \to 0$$

and deduce that

$$1 - \left(1 - \frac{t^2}{n^2}\right)^n \to 0$$

as $n \to \infty$. Hence deduce that

$$\left(1 - \frac{t}{n}\right)^n \to e^{-t}$$

as $n \to \infty$ for all $t > 0$.

(vi) Conclude that

$$\left(1 + \frac{t}{n}\right)^n \to e^t$$

as $n \to \infty$ for all real t.

54

Q 18.8. (a) Prove the statements made in Example 3.9.

(i) The sum $\sum_{n=1}^{\infty} n^{-2} z^n$ has radius of convergence 1 and converges for all $|z| = 1$.

(ii) The sum $\sum_{n=0}^{\infty} z^n$ has radius of convergence 1 and diverges for all $|z| = 1$.

(iii) The sum $\sum_{n=1}^{\infty} n^{-1} z^n$ has radius of convergence 1, diverges for $z = 1$ and converges for $z = -1$.

(b) Suppose that $\sum_{j=0}^{\infty} a_j z^j$ has radius of convergence R and $\sum_{j=0}^{\infty} b_j z^j$ has radius of convergence S.

(i) Show that if $R \neq S$ then $\sum_{j=0}^{\infty} (a_j + b_j) z^j$ has radius of convergence $\min(R, S)$.

(ii) Show that if $R = S$ then $\sum_{j=0}^{\infty} (a_j + b_j) z^j$ has radius of convergence at least R.

(iii) If $T \geq 1 = R = S$ give an example where $\sum_{j=0}^{\infty} (a_j + b_j) z^j$ has radius of convergence T.

Q 18.9. (i) Suppose that x_n is a bounded sequence of real numbers. Show that $y_n = \sup_{m \geq n} x_m$ is a well defined bounded decreasing sequence and so y_n tends to a limit y say. We write

$$\limsup_{n \to \infty} x_n = y.$$

(ii) Let a_n be a sequence of complex numbers. Show that if the sequence $|a_n|^{1/n}$ is unbounded, then $\sum_{j=0}^{\infty} a_j z^j$ has radius of convergence 0 and if the the sequence $|a_n|^{1/n}$ is bounded and non-zero, then $\sum_{j=0}^{\infty} a_j z^j$ has radius of convergence $(\limsup_{n \to \infty} |a_n|^{1/n})^{-1}$. State, with reasons, what happens when $\limsup_{n \to \infty} |a_n|^{1/n} = 0$

Q 18.10. Suppose that A and B are non-empty bounded subsets of $\mathbb{R}$. Show that

$$\sup\{a + b : a \in A, \ b \in B\} = \sup A + \sup B.$$

The last formula is more frequently written

$$\sup_{a \in A, \ b \in B} a + b = \sup_{a \in A} a + \sup_{b \in B} b.$$

Suppose, further that a_n and b_n are bounded sequences of real numbers. For each of the following statements either give a proof that it is always true or an example to show that it is sometimes false.

(i) $\sup_n (a_n + b_n) = \sup_n a_n + \sup_n b_n$.

(ii) $\sup_{a \in A, \ b \in B} ab = (\sup_{a \in A} a)(\sup_{b \in B} b)$.

(iii) $\inf_{a \in A, \ b \in B} a + b = \inf_{a \in A} a + \inf_{b \in B} b$.

Q 18.11. Prove all the statements made in Lemma 4.3, viz:

(i) The limit is unique. That is, if $\mathbf{a}_n \to \mathbf{a}$ and $\mathbf{a}_n \to \mathbf{b}$ as $n \to \infty$ then $\mathbf{a} = \mathbf{b}$.

(ii) If $\mathbf{a}_n \to \mathbf{a}$ as $n \to \infty$ and $n(1) < n(2) < n(3) \ldots$ then $\mathbf{a}_{n(j)} \to \mathbf{a}$ as $j \to \infty$.

(iii) If $\mathbf{a}_n = \mathbf{c}$ for all n then $\mathbf{a}_n \to \mathbf{c}$ as $n \to \infty$.

(iv) If $\mathbf{a}_n \to \mathbf{a}$ and $\mathbf{b}_n \to \mathbf{b}$ as $n \to \infty$ then $\mathbf{a}_n + \mathbf{b}_n \to \mathbf{a} + \mathbf{b}$.

(v) Suppose $\mathbf{a}_n \in \mathbb{R}^m$, $\mathbf{a} \in \mathbb{R}^m$ $\lambda_n \in \mathbb{R}$, and $\lambda \in \mathbb{R}$. If $\mathbf{a}_n \to \mathbf{a}$ and $\lambda_n \to \lambda$ then $\lambda_n \mathbf{a}_n \to \lambda \mathbf{a}$.

Q 18.12. Use the Bolzano–Weierstrass theorem to prove the general principle of convergence in $\mathbb{R}^m$ (Theorem 4.6).

The remaining questions are included for general interest and not for relevance to the syllabus or to passing Tripos exams.

Q 18.13. By comparing

$$\frac{a}{b} \text{ with } \frac{3a + 4b}{2a + 3b}$$

show that given any $x \in \mathbb{Q}$ with $x^2 \leq 2$ we can find a $y \in \mathbb{Q}$ with $y > x$ and $y^2 \leq 2$. Deduce that the set $\{x \in \mathbb{Q} : x^2 < 2\}$ has no supremum in $\mathbb{Q}$. (The idea of comparing a/b with $(3a + 4b)/(2a + 3b)$ goes back as far as Euclid.)

Q 18.14. If $a < c < b$ and E is a subset of $\mathbb{R}$ explain why if $E \cap [a, c]$ and $E \cap [c, b]$ are finite then so is $E \cap [a, b]$. Use this to give a proof by bisection of Bolzano's theorem.

Q 18.15. Do Exercise 3.14 showing the uncountability of the reals.

Let y_1, y_2, ... be any sequence of points in $\mathbb{R}$. Let $x_0 = 0$, $\delta_0 = 1$.

(i) Show that you can construct inductively a sequence of real numbers x_1, x_2, ... and positive numbers δ_j such that

(a) $|x_n - x_{n-1}| < \delta_{n-1}/4$,

(b) $x_n \neq y_n$,

(c) $0 < \delta_n < |x_n - y_n|$,

(d) $\delta_n < \delta_{n-1}/4$.

(ii) Show that $\delta_{n+m} < 4^{-m}\delta_n$ for $m, n \geq 0$ and deduce that the x_n form a Cauchy sequence. Conclude that x_n tends to a limit x.

(iii) Show that $|x_{n+m} - x_n| < \delta_n/3$ for all $m, n \geq 0$. Deduce that $|x - x_n| \leq \delta_n/3$ for all $n \geq 0$. Why does this show that $y_n \neq x$ for all n.

(iv) Prove that the real numbers are uncountable.

Q 18.16. Suppose we consider the collection $\mathcal{Q}$ of rational functions

$$f(X) = \frac{a_0 + a_1 X + \cdots + a_n X^n}{b_0 + b_1 X + \cdots + b_m X^m}$$

(where the a_j and b_j are real and where $b_m \neq 0$ and $a_n \neq 0$ if $n \geq 1$) with the usual rules for addition and multiplication. Suppose that we say that $f(X) > 0$ if $a_n b_m > 0$ and that $f > g$ if $f - g > 0$. Convince yourself that $\mathcal{Q}$ has all the standard properties of arithmetic and order we expect. (Some students will only be able to convince themselves by looking up the axioms for an ordered field in, for example, [8] and then verifying each axiom in turn. Good luck to them, I say.)

Show that $1/n > 1/X > 0$ for all integers n and so Archimedes' Axiom is false for $\mathcal{Q}$.

19 Second Sheet of Exercises

I have tried to produce 12 rather routine questions for each sheet. Any further questions are for interest only. Ambitious students and their supervisors should look at Tripos questions to supplement this meagre fare.

Q 19.1. Here is an alternative proof of Theorem 6.8
(i) Let $K \geq 0$ and $\epsilon > 0$. If $f : [a, b] \to \mathbb{R}$ has the property that

$$f(b) - f(a) \geq (K + \epsilon)(b - a)$$

and N is strictly positive integer show that there exists an integer r with $1 \leq r \leq N$ and

$$f(a + r(b - a)/N) - f(a + (r - 1)(b - a)/N) \geq (K + \epsilon)(b - a)/N$$

(ii) Under the conditions of part (i) deduce that there exists a sequence of closed intervals $[x_n, y_n]$ with

$$[a, b] = [x_0, y_0] \supseteq [x_1, y_1] \supseteq [x_2, y_2] \supseteq \ldots$$

such that $f(y_n) - f(x_n) \geq (K + \epsilon)(y_n - x_n)$ and $y_n - x_n \to 0$ as $n \to \infty$.
(iii) Continuing with the same notation, show that $x_n \to c$ for some $c \in [a, b]$. Show that $c \in [x_n, y_n]$ for all n and that, if $c \neq x_n, y_n$ then

$$\max \left(\frac{f(y_n) - f(c)}{y_n - c}, \frac{f(c) - f(x_n)}{c - x_n} \right) \geq K + \epsilon.$$

Conclude that if f is differentiable at c then $f'(c) \geq K + \epsilon$.
(iv) Deduce all the results of Theorem 6.8.

Q 19.2. (i) If E is a set show that the set

$$E^\circ = \bigcup \{U : U \text{ is open and } U \subseteq E\}$$

has the following properties.
(a) E° is open.
(b) $E^\circ \subseteq E$.
(c) If A is open and $A \subseteq E$ then $A \subseteq E^\circ$.
(ii) Suppose that E' is such that
(a) E' is open.
(b) $E' \subseteq E$.
(c) If A is open and $A \subseteq E$ then $A \subseteq E'$.

Show that $E' = E^\circ$.

(iii) Explain why (i) and (ii) make it possible to define the *interior* of E as the largest open set contained in E.

(iv) Show why it is possible to define the *closure* of E as the smallest closed set containing E.

(v) Find the interior and closure of (a, b), $[a, b)$, $(a, b]$ and $[a, b]$ where $a < b$.

Q 19.3. Do the Exercise 5.13 restated below.

After looking at parts (iii) to (v) of Lemma 4.3 state the corresponding results for continuous functions. (Thus part (v) corresponds to the statement that if $\lambda : E \to \mathbb{R}$ and $f : E \to \mathbb{R}^p$ are continuous at $\mathbf{x} \in E$ then so is λf.) Prove your statements directly from Definition 5.12.

Q 19.4. If $f, g : \mathbb{R}^p \to \mathbb{R}^m$ are continuous show that

$$\begin{pmatrix} f \\ g \end{pmatrix} : \mathbb{R}^p \to \mathbb{R}^{2m} = \mathbb{R}^m \times \mathbb{R}^m$$

defined by

$$\begin{pmatrix} f \\ g \end{pmatrix} (\mathbf{x}) = \begin{pmatrix} f(\mathbf{x}) \\ g(\mathbf{x}) \end{pmatrix}$$

is continuous.

Show that the map $A : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ defined by

$$A \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{x} + \mathbf{y}$$

whenever $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ is continuous. By applying the continuity of compositions of continuous functions (Lemma 5.16) to $A \circ \begin{pmatrix} f \\ g \end{pmatrix}$ deduce that $f + g$ is continuous.

Obtain the other results of Question 19.3 similarly.

Q 19.5. Let $f : \mathbb{R} \to \mathbb{R}$. State which of the following statements are always true and which may be false giving a proof or a counter example as appropriate.

(i) If $f^{-1}(E)$ is closed whenever E is closed then f is continuous.

(ii) If f is continuous then $f(U)$ is open whenever U is open.

(iii) If f is continuous then $f(E)$ is bounded whenever E is bounded.

(iv) If $f(E)$ is bounded whenever E is bounded then f is continuous.

Q 19.6. Suppose that E is a subset of $\mathbb{R}^n$. Show that E is closed and bounded if and only if every continuous function $f : E \to \mathbb{R}^n$ is bounded.

Q 19.7. If A and B are subsets of $\mathbb{R}^m$ we write

$$A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \ \mathbf{b} \in B\}$$

By considering the case $m = 1$,

$$A = \{-n : n \in \mathbb{N}\} \text{ and } B = \{n + 1/n : n \in \mathbb{N}, \ n \geq 2\}$$

show that if A and B are closed it does not follow that $A + B$ is closed.

Show however that if A is closed and bounded and B is closed then $A + B$ is closed.

If A and B are open, does it follow that $A + B$ is open? Give reasons.

Q 19.8. Consider the map $\Omega : \mathbb{R}^6 \to \mathbb{R}^3$ given by

$$\Omega \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{x} \wedge \mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$. (Here $\wedge$ is the 'vector product' of mathematical methods.) Show directly from the definition that Ω is differentiable.

Having done this, find the Jacobian matrix of Ω.

Q 19.9 (Very traditional). Consider the map $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = xy|x - y|.$$

At which points is f differentiable? Prove your statements.
[Deal quickly with the easy part. Do not be surprised if the hard part reveals a special case that must be dealt with separately.]

Q 19.10 (Also very traditional). (i) Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(t) = t^2 \sin 1/t$ for $t \neq 0$, $f(0) = 0$. Show that f is everywhere differentiable and find its derivative. Show that f' is not continuous. [Deal quickly with the easy part and then go back to the definition to deal with $t = 0$. There are wide selection of counter-examples obtained by looking at $t^\beta \sin t^\alpha$ for various values of α and β.]

(ii) Find an infinitely differentiable function $g : (1, \infty) \to \mathbb{R}$ such that $g(t) \to 0$ but $g'(t) \not\to 0$ as $t \to \infty$.

(iii) Consider the function $F : \mathbb{R}^2 \to \mathbb{R}$ given by $F(s, t) = f(s) + f(t)$ where f is the function defined in (i). Show that F is everywhere differentiable but $F_{,1}$ and $F_{,2}$ are not continuous at $(0, 0)$.

Q 19.11. Let **u** and **v** be linearly independent vectors in $\mathbb{R}^2$. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}$ has the property that

$$\frac{F(\mathbf{x} + h\mathbf{u}) - F(\mathbf{x})}{h} \to a(\mathbf{x}), \text{ and } \frac{F(\mathbf{x} + h\mathbf{v}) - F(\mathbf{x})}{h} \to b(\mathbf{x})$$

as $h \to 0$ for all $\mathbf{x} \in \mathbb{R}^2$. Suppose further that $a : \mathbb{R}^2 \to \mathbb{R}$ is continuous. Show that F is everywhere differentiable.

Q 19.12. Show that Definition 7.4 can be modified to give the definition of an operator norm $\|\alpha\|$ for a linear map $\alpha : \mathbb{C}^m \to \mathbb{C}^p$ and that the results of Lemma 7.6 continue hold *mutatis mutandis*[11].

If α has matrix (a_{ij}) with respect to the standard bases show that

$$\max_{i,j} |a_{ij}| \le \|\alpha\| \le mp \max_{i,j} |a_{ij}|.$$

For the rest of this question we take $m = p$ (so we are dealing with *endomorphisms* of the vector space $\mathbb{C}^m$).

(i) If α has a lower triangular matrix (a_{ij}) with respect to the standard basis show that the set of its eigenvalues is precisely the set of the diagonal entries a_{ii}. (Pure algebra.)

(ii) If α has a lower triangular matrix (a_{ij}) and $\epsilon > 0$ show that we can find β such that β has m distinct non-zero eigenvalues and $\|\beta - \alpha\| < \epsilon$.

(iii) In algebra you show that given any endomorphism α we can find an invertible endomorphism θ such that $\theta\alpha\theta^{-1}$ has a lower triangular matrix with respect to the standard basis. (N.B. This is true for complex vector spaces but not for real ones.) Use this fact to show that given any endomorphism α and any $\eta > 0$ we can an endomorphism γ such that $\|\gamma - \alpha\| < \eta$ and γ has m distinct non-zero eigenvalues.

(iv) Write $P_\alpha(t) = \det(t\iota - \alpha)$ (where ι is the identity endomorphism). Show that if α has m distinct eigenvalues then $P_\alpha(\alpha) = 0$. (Pure algebra.)

(v) Let α be any endomorphism. By considering a sequence α_n of endomorphisms with m distinct eigenvalues such that $\|\alpha_n - \alpha\| \to 0$ as $n \to \infty$ show that $P_\alpha(\alpha) = 0$.

[This procedure is substantially longer than that used to prove the Cayley-Hamilton theorem in your algebra course but it is an instructive way of looking at things.]

---

[11]Changing those things that need to be changed.

20   Third Sheet of Exercises

I have tried to produce 12 rather routine questions for each sheet. Any remaining questions are for interest only. Ambitious students and their supervisors should look at Tripos questions to supplement this meagre fare.

Q 20.1. At the beginning of the 19th century it was hoped that the calculus could be reduced to *mere algebra* by using Taylor's theorem. The following example due to Cauchy shows why such hopes were misplaced.

Let $E : \mathbb{R} \to \mathbb{R}$ be defined by $E(t) = \exp(-1/t^2)$ for $t \neq 0$ and $E(0) = 0$. Use induction to show that E is infinitely differentiable with

$$E^{(n)}(t) = Q_n(1/t)E(t) \qquad \text{for } t \neq 0$$
$$E^{(n)}(0) = 0.$$

where Q_n is a polynomial. For which values of t is it true that

$$E(t) = \sum_{n=0}^{\infty} \frac{E^n(0)t^n}{n!}?$$

Explain why the behaviour of E is consistent with Lemma 8.7 (taking $m = 1$).

Q 20.2. In the first year you saw how the find stationary points for a well behaved function $f : \mathbb{R}^2 \to \mathbb{R}$ and how to decide (except in certain specified cases) whether they are maxima, minima or saddle-points. Use Theorem 8.3 to state and prove your results rigorously.

In the course on quadratic mathematics you will see that given any real symmetric $m \times m$ matrix A you can find a special orthogonal matrix P such that PAP^T is diagonal. Explain the relevance of this result to the problem of deciding (except in certain *specified* cases) when a stationary point of a well behaved function $f : \mathbb{R}^m \to \mathbb{R}$ is a maximum, a minimum or a saddle point.

Q 20.3. Define $f : [0,1] \to \mathbb{R}$ by $f(p/q) = 1/q$ when p and q are coprime integers with $1 \leq p < q$ and $f(x) = 0$ otherwise.
(i) Show that f is Riemann integrable and find $\int_0^1 f(x)\,dx$.
(ii) At which points is f continuous? Prove your answer.

Q 20.4. We say that a function $f : [0,1] \to \mathbb{R}$ is of bounded variation if there exists a constant K such that whenever $0 \leq x_0 \leq x_1 \leq x_2 \cdots \leq x_n \leq 1$

we have

$$\sum_{j=1}^{n} |f(x_{j-1}) - f(x_j)| \leq K.$$

(i) Show that if $g : [0, 1] \to \mathbb{R}$ is increasing then g is of bounded variation. Show that if $h_1, h_2 : [0, 1] \to \mathbb{R}$ are increasing then $h = h_1 - h_2$ is of bounded variation.

(ii) Suppose that f is of bounded variation. Show that

$$f_1(x) = \sup\{\sum_{j=1}^{n} f(x_j) - f(y_j) : 0 \leq y_1 \leq x_1 \leq y_2 \leq x_2 \leq \cdots \leq y_n \leq x_n \leq x\}$$

is a well defined increasing function. Show that $f_1 - f$ is also an increasing function and so f is the difference of two increasing functions.

(iii) Show that if we set $f(x) = x \sin x^{-1}$ for $x \neq 0$ and $f(0) = 0$ then $f : [0, 1] \to \mathbb{R}$ is a continuous function which is not of bounded variation. [Remember diagrams may not be rigorous but they certainly help you understand what you are talking about.]

Q 20.5. If $f : [0, 1] \to \mathbb{R}$ is increasing show that if x_n is an increasing sequence then $f(x_n)$ tends to limit as $n \to \infty$. Deduce carefully that if $0 < y \leq 1$ then $f(x)$ tends to a limit $f(y-)$ as $x \to y$ through values of x with $x < y$. (We shall set $f(0-) = f(0)$.) Define $f(y+)$ similarly.

Let us call $f(y+) - f(y-)$ the jump $J(y)$ at y. Show that if $y_1, y_2, \ldots y_n$ are distinct points of $(0, 1)$ then

$$\sum_{j=1}^{n} J(y_j) \leq f(1) - f(0).$$

Show that f can have only finitely many jumps of size $1/n$ or more. Deduce that f has only countably many jumps and that f must be continuous at some point.

Q 20.6. Let $\mathcal{K}$ be the set of functions $f : [0, 1] \to \mathbb{R}$ such that f is continuous on $(0, 1]$ and f has an improper Riemann integral

$$\mathcal{I}f = \int_0^1 f(x) \, dx.$$

Show that if $f, g \in \mathcal{K}$ and $\lambda, \mu \in \mathbb{R}$ then $\lambda f + \mu g \in \mathcal{K}$ and

$$\mathcal{I}(\lambda f + \mu g) = \lambda \mathcal{I}f + \mu \mathcal{I}g.$$

By giving a proof or counter-example establish whether it is always true that if $f, g \in \mathcal{K}$ then $fg \in \mathcal{K}$.

Q 20.7. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous. By using results on the differentiation of integrals which should be quoted exactly show that

$$\frac{d}{dt} \int_a^t \left(\int_c^d f(u, v) \, dv \right) du = \frac{d}{dt} \int_c^d \left(\int_a^t f(u, v) \, du \right) dv.$$

Deduce that

$$\int_a^b \left(\int_c^d f(u, v) \, dv \right) du = \int_c^d \left(\int_a^b f(u, v) \, du \right) dv,$$

or, more informally

$$\int_a^b \int_c^d f(u, v) \, dv \, du = \int_c^d \int_a^b f(u, v) \, du \, dv.$$

Q 20.8. Prove all the results of Section 11. (A long but easy and useful exercise. In order to keep the noise level from whinging students down to tolerable levels this counts as 2 questions.)

Q 20.9. See Question 20.8.

Q 20.10. Consider $\mathcal{E}$ the space of linear maps $\alpha : \mathbb{R}^u \to \mathbb{R}^u$. We wish to show that $d(\alpha, \beta) = \|\alpha - \beta\|$ (with $\|\alpha\|$ the usual 'operator norm') defines a complete metric.

To this end consider a Cauchy sequence α_n in $\mathcal{E}$. Show that if $\mathbf{x} \in \mathbb{R}^u$ then $\alpha_n(\mathbf{x})$ tends to limit, call it $\alpha(\mathbf{x})$. Now show that $\alpha : \mathbb{R}^u \to \mathbb{R}^u$ is linear. By using the inequality

$$\|\alpha(\mathbf{x}) - \alpha_n(\mathbf{x})\| \leq \|\alpha_m(\mathbf{x}) - \alpha_n(\mathbf{x})\| + \|\alpha_m(\mathbf{x}) - \alpha(\mathbf{x})\|$$
$$\leq \sup_{p,q \geq n} \|\alpha_p(\mathbf{x}) - \alpha_q(\mathbf{x})\| + \|\alpha_m(\mathbf{x}) - \alpha(\mathbf{x})\|,$$

and allowing $m \to \infty$ show that

$$\|\alpha - \alpha_n\| \leq \sup_{p,q \geq n} \|\alpha_p - \alpha_q\|,$$

and deduce that $\|\alpha - \alpha_n\| \to \infty$ as $n \to \infty$.

Q 20.11. In Question 20.10 we showed that $\mathcal{E}$ with the appropriate metric is complete. In this question we make use of the result.

(i) Suppose $\alpha \in \mathcal{E}$ and $\|\alpha\| < 1$. Let

$$\sigma_n = \iota + \alpha + \alpha^2 + \cdots + \alpha^n.$$

Show that σ_n converges to a limit σ, say.

Compute $(\iota - \alpha)\sigma_n$ and show, carefully that $(\iota - \alpha)\sigma = \iota$.

(iii) Deduce that if $\|\iota - \gamma\| < 1$ then γ is invertible.

(iv) Show that if $\beta \in \mathcal{E}$ is invertible and $\|\beta - \tau\| < \|\beta^{-1}\|^{-1}$ then τ is invertible.

(v) Conclude that the invertible elements of $\mathcal{E}$ form an open set.

Q 20.12. Consider $C([-1, 1])$ the set of continuous functions $f : [-1, 1] \to \mathbb{R}$. Show that, if we set

$$d(f, g) = \|f - g\|_2 = \left(\int_{-1}^{1} |f(x) - g(x)|^2 \, dx \right)^{1/2}$$

then d is a metric but d is not complete.

Show that if $\|f - g\|_1$ is defined as in Example 13.8 then $2^{1/2}\|f\|_2 \geq \|f\|_1$ for all $f \in C([-1, 1])$ but that, given any constant $K > 0$ we can find a $g \in C([-1, 1])$ such that $\|g\|_2 \geq K\|g\|_1$.

The remaining question is included for general interest and not for relevance to the syllabus or to passing Tripos exams.

Q 20.13. We use the notation of Questions 20.10 and 20.11.

(i) Show that if $\alpha \in \mathcal{E}$ we can find $e^\alpha \in \mathcal{E}$ such that

$$\left\| \sum_{r=0}^{n} \frac{\alpha^r}{r!} - e^\alpha \right\| \to 0$$

as $n \to \infty$.

(ii) Show carefully that if α and β commute

$$e^\alpha e^\beta = e^{\alpha + \beta}.$$

(iii) Show that if α and β are general (not necessarily commuting) elements of $\mathcal{E}$ then

$$\left\| h^{-2} (e^{h\alpha} e^{h\beta} - e^{h\beta} e^{h\alpha}) - (\alpha\beta - \beta\alpha) \right\| \to 0$$

as the real number $h \to 0$.

Conclude that, in general, $e^\alpha e^\beta$ and $e^{\alpha + \beta}$ need not be equal.

21 Fourth Sheet of Exercises

I have tried to produce 12 rather routine questions for each sheet. The remaining questions are for interest only. Ambitious students and their supervisors should look at Tripos questions to supplement this meagre fare.

Q 21.1. Consider the two railway metrics d_1 and d_2 on $\mathbb{R}^2$ given in Examples 11.3 and 11.4. For each of the two metrics give a reasonably simple description of the open balls of radius r and centre $\mathbf{x}$ when $\mathbf{x} = \mathbf{0}$ and when $\mathbf{x} \neq \mathbf{0}$ and (a) $r \leq \|\mathbf{x}\|$, (b) $r < \|\mathbf{x}\| \leq 2r$, (c) $2r < \|\mathbf{x}\|$.

Explain the statement 'when finding open sets only balls of small radius are important so cases (b) and (c) are irrelevant'.

Give the simplest description you can find of open sets in the two metrics.

Q 21.2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable and let $g(x) = f(x, c - x)$ where c constant. Show that $g : \mathbb{R} \to \mathbb{R}$ is differentiable and find its derivative

(i) directly from the definition of differentiability

and also

(ii) by using the chain rule.

Deduce that if $f_{,1} = f_{,2}$ throughout $\mathbb{R}$ then $f(x, y) = h(x + y)$ for some differentiable function h.

Q 21.3 (Traditional). Consider the functions $f_n : [0, 1] \to \mathbb{R}$ defined by $f_n(x) = n^p x \exp(-n^q x)$ where $p, q > 0$.

(i) Show that f_n converges pointwise on $[0, 1]$.

(ii) Show that if $p < q$ then f_n converges uniformly on $[0, 1]$.

(iii) Show that if $p \geq q$ then f_n does not converge uniformly on $[0, 1]$. Does f_n converge uniformly on $[0, 1 - \epsilon]$? Does f_n converge uniformly on $[\epsilon, 1]$? (Here $0 < \epsilon < 1$, you should justify your answers.)

Q 21.4. Let (X, d) be a metric space. We say that E is dense in X if whenever $x \in X$ we can find $e_n \in E$ with $e_n \to x$.

(i) Show that the rationals are dense in the reals (i.e. $\mathbb{Q}$ is dense in $\mathbb{R}$ if we give $\mathbb{R}$ the usual metric). Show that the irrationals are dense in the reals. [You should make explicit any use you make of the axiom of Archimedes.]

(ii) Consider the space $C([0, 1])$ of continuous functions with the uniform norm. Show that the piecewise linear functions are dense in $C([0, 1])$.

Q 21.5. Recall from your courses in mathematical methods that, if $f : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π then we write

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \exp(-int) \, dt.$$

Show, quoting carefully any results that you use, that, if $a_n \in \mathbb{C}$ and $\sum_{n=-N}^{N} |a_n|$ converges, then $\sum_{n=-N}^{N} a_n \exp(int)$ converges uniformly to a continuous periodic function $f(t)$, say, with period 2π. Show that $\hat{f}(n) = a_n$. for each n.

[This chain of thought is continued in Question 21.14.]

Q 21.6. Recall that $\| \; \|$ is said to be a norm on a real vector space V if the following conditions hold.

(1) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in V$.

(2) $\|\mathbf{x}\| = 0$ implies $\mathbf{x} = \mathbf{0}$.

(3) $\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$ for all $\lambda \in \mathbb{R}$ and all $\mathbf{x} \in V$.

(4) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}$.

(i) Let $\| \; \|_2$ be the usual Euclidean norm on $\mathbb{R}^m$ and let $\| \; \|$ be another norm on $\mathbb{R}^m$. Let $I : (\mathbb{R}^m, \| \; \|_2) \to (\mathbb{R}^m, \| \; \|)$ be the identity map, given by $I\mathbf{x} = \mathbf{x}$. By considering basis vectors or otherwise that there exists a constant K such that $\|\mathbf{x}\| \leq K\|\mathbf{x}\|_2$ for all $\|\mathbf{x}\| \in \mathbb{R}^m$. Deduce that I is continuous.

By applying a theorem on continuous functions on closed bounded sets show that there exists a $k > 0$ such that $\|\mathbf{x}\| \geq k$ for all $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$. Conclude that

$$k\|\mathbf{x}\|_2 \leq \|\mathbf{x}\| \leq K\|\mathbf{x}\|_2$$

for all $\|\mathbf{x}\| \in \mathbb{R}^m$.

[Thus all norms on a finite dimensional space are essentially the same.]

(ii) Consider the real vector space $\mathbb{R}^{\mathbb{N}}$ of all real sequences $\mathbf{x} = (x_1, x_2, \dots)$ with the usual vector addition and multiplication by scalars. Let V be the set of all sequences $\mathbf{x} = (x_1, x_2, \dots)$ such that $x_j \neq 0$ for only finitely many j. Show that V is a subspace of $\mathbb{R}^{\mathbb{N}}$ and so a vector space in its own right. By considering norms of the form $\|\mathbf{x}\| = \max_j \kappa_j |x_j|$, or otherwise, find two norms $\| \; \|_A$ and $\| \; \|_B$ such that

$$\sup_{\|\mathbf{x}\|_A = 1} \|\mathbf{x}\|_B = \sup_{\|\mathbf{x}\|_B = 1} \|\mathbf{x}\|_A = \infty.$$

Q 21.7. (i) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a twice differentiable function such that

$$\left| \frac{f(x)f''(x)}{f'(x)^2} \right| \leq \lambda$$

for all x and some $|\lambda| < 1$ Show that the mapping

$$Tx = x - \frac{f(x)}{f'(x)}$$

is a contraction mapping and deduce that f has a unique root y.

(ii) Suppose that $F : \mathbb{R} \to \mathbb{R}$ is a twice differentiable function such that

$$\left| \frac{F(x)F''(x)}{F'(x)^2} \right| \leq \lambda$$

for all $|x| \leq a$ and some $|\lambda| < 1$ and that $F(0) = 0$. Consider the mapping

$$Tx = x - \frac{F(x)}{F'(x)}.$$

Show that $T^n x \to 0$.

Suppose that

$$\frac{\sup_{|t| \leq a} |f'(t)| \sup_{|t| \leq a} |f''(t)|}{\inf_{|t| \leq a} |f'(t)|^2} = M.$$

By using the mean value theorem twice, show that if $|x| \leq a$ then

$$|Tx| \leq Mx^2.$$

(iii) If you know what the Newton–Raphson method is, comment on the relevance of the results of (i) and (ii) to that method.

Q 21.8. We work in $\mathbb{R}^m$ with the usual 'Euclidean norm'. The following line of thought is extremely important in later work. Suppose we want the solution $\mathbf{x}_0$ of

$$\mathbf{x} + \boldsymbol{\epsilon}(\mathbf{x}) = \mathbf{y}$$

where $\boldsymbol{\epsilon}(\mathbf{x})$ is 'a small error term' or 'of first order compared to $\mathbf{x}$. The following method of solution is traditional (for the excellent reason that it usually works like a Spanish charm). Start by ignoring the small $\boldsymbol{\epsilon}$ term and guess $\mathbf{x}_0 \approx \mathbf{x}_1 = \mathbf{y}$. Since this initial guess is good we estimate the small $\boldsymbol{\epsilon}$ term as $\boldsymbol{\epsilon}(\mathbf{x}_1)$. Feeding this estimate back into the equation we now guess $\mathbf{x}_0 \approx \mathbf{x}_2 = \mathbf{y} - \boldsymbol{\epsilon}(\mathbf{x}_1)$. Repeating this process gives us the rule for successive approximations

$$\mathbf{x}_{n+1} = \mathbf{y} - \boldsymbol{\epsilon}(\mathbf{x}_n).$$

We now try to justify this in certain cases.

(i) Suppose that $\boldsymbol{\epsilon}(\mathbf{0}) = \mathbf{0}$ and

$$\|\boldsymbol{\epsilon}(\mathbf{a}) - \boldsymbol{\epsilon}(\mathbf{b})\| < \|\mathbf{a} - \mathbf{b}\|/2$$

for all $\|\mathbf{a}\|, \|\mathbf{b}\| \leq \delta$ where $\delta > 0$. (Note that these are rather stronger conditions than might have been expected. However, this is not to say that the general idea might not work under weaker conditions.) We shall show that the method works for $\|\mathbf{y}\| \leq \delta/2$.

Consider the closed ball $B = \{\mathbf{x} : \|\mathbf{x}\| \leq \delta\}$. Show that the equation

$$T(\mathbf{x}) = \mathbf{y} - \boldsymbol{\epsilon}(\mathbf{x})$$

defines a map $T : B \to B$. Now use the contraction mapping theorem to show that

$$\mathbf{x} + \boldsymbol{\epsilon}(\mathbf{x}) = \mathbf{y}$$

has a unique solution $\mathbf{x}(\mathbf{y})$ with $\|\mathbf{x}(\mathbf{y})\| \leq \delta$ and that $T^n(\mathbf{y}) \to \mathbf{x}(\mathbf{y})$ as $n \to \infty$.

(ii) Suppose that $f : \mathbb{R}^m \to \mathbb{R}^m$ is a differentiable function with derivative continuous at $\mathbf{0}$, $f(\mathbf{0}) = \mathbf{0}$ and $Df(\mathbf{0}) = I$. Show that we can find a $\delta > 0$ such that whenever $\|\mathbf{y}\| \leq \delta/2$ the equation

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

has a unique solution $\mathbf{x}(\mathbf{y})$ with $\|\mathbf{x}(\mathbf{y})\| \leq \delta$.

(iii) Suppose that $f : \mathbb{R}^m \to \mathbb{R}^m$ is a differentiable function with derivative continuous at $\mathbf{0}$, $f(\mathbf{0}) = \mathbf{0}$ and $Df(\mathbf{0})$ invertible. Show that we can find a $\delta_1, \delta_2 > 0$ such that whenever $\|\mathbf{y}\| \leq \delta_1$ the equation

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

has a unique solution $\mathbf{x}(\mathbf{y})$ with $\|\mathbf{x}(\mathbf{y})\| \leq \delta_2$.

Q 21.9. (A method of Abel) (i) Suppose that a_j and b_j are sequences of complex numbers and that $S_n = \sum_{j=1}^{n} a_j$ for $n \geq 1$ and $S_0 = 0$. Show that, if $1 \leq u \leq v$ then

$$\sum_{j=u}^{v} a_j b_j = \sum_{j=u}^{v} S_j(b_j - b_{j+1}) - S_{u-1}b_u + S_v b_{v+1}.$$

(This is known as partial summation, for obvious reasons.)

(ii) Suppose now that, in addition, the b_j form a decreasing sequence of positive terms and that $|S_n| \leq K$ for all n. Show that

$$\left| \sum_{j=u}^{v} a_j b_j \right| \leq 2K b_u.$$

Deduce that if $b_j \to 0$ as $j \to \infty$ then $\sum_{j=1}^{\infty} a_j b_j$ converges.

Deduce the alternating series test.

(iii) If b_j is a decreasing sequence of positive terms with $b_j \to 0$ as $j \to \infty$ show that $\sum_{j=1}^{\infty} b_j z^j$ converges uniformly in the region given by $|z| \leq 1$ and $|z - 1| \geq \epsilon$ for all $\epsilon > 0$.

Q 21.10. We work in $\mathbb{R}^m$ with the usual distance.

(i) Show that if A_1, A_2, are non-empty, closed and bounded with $A_1 \supseteq A_2 \supseteq \ldots$ then $\bigcap_{j=1}^{\infty} A_j$ is non empty. Is this result true if we merely assume A_j closed and non-empty? Give reasons.

(ii) If A is non-empty, closed and bounded show that we can find $\mathbf{a}', \mathbf{b}' \in A$ such that

$$\|\mathbf{a}' - \mathbf{b}'\| \geq \|\mathbf{a} - \mathbf{b}\|$$

for all $\mathbf{a}, \mathbf{b} \in A$. Is this result true if we merely assume A bounded and non-empty? Give reasons.

Q 21.11. We work in $\mathbb{R}^m$ with the usual distance. Let E be a closed non-empty subset of $\mathbb{R}^m$ and let T be a map $T : E \to E$.

(i) Suppose $\|T(\mathbf{a}) - T(\mathbf{b})\| < \|\mathbf{a} - \mathbf{b}\|$ for all $\mathbf{a}, \mathbf{b} \in E$ with $\mathbf{a} \neq \mathbf{b}$. We saw in Example 15.2 that T need not have a fixed point. Show that, if T has a fixed point, it is unique.

(ii) Suppose $\|T(\mathbf{a}) - T(\mathbf{b})\| > \|\mathbf{a} - \mathbf{b}\|$ for all $\mathbf{a}, \mathbf{b} \in E$ with $\mathbf{a} \neq \mathbf{b}$. In Question 21.17 it is shown that T need not have a fixed point. Show that, if T has a fixed point, it is unique.

(iii) Suppose $\|T(\mathbf{a}) - T(\mathbf{b})\| = \|\mathbf{a} - \mathbf{b}\|$ for all $\mathbf{a}, \mathbf{b} \in E$. Show that T need not have a fixed point. and that, if T has a fixed point, it need not be unique.

(iv) Suppose now that E is non-empty, closed and bounded and

$$\|T(\mathbf{a}) - T(\mathbf{b})\| < \|\mathbf{a} - \mathbf{b}\|$$

for all $\mathbf{a}, \mathbf{b} \in E$ with $\mathbf{a} \neq \mathbf{b}$. By considering $\inf_{mathbf{x} \in E} \|\mathbf{x} - T(mathbf{x})\|$, or otherwise show that T has a fixed point.

Q 21.12. From time to time numerical analysts mention the *spectral radius*. It forms no part of any of non-optional 1B courses but the reader may be interested to see what it is.

(i) Give an example of a linear map $\beta : \mathbb{R}^m \to \mathbb{R}^m$ such that $\beta^{m-1} \neq \mathbf{0}$ but $\beta^m = \mathbf{0}$.

(ii) Let $\alpha : \mathbb{R}^m \to \mathbb{R}^m$ be linear. If $n = jk + r$ explain why

$$\|\alpha^{jk+r}\| \leq \|\alpha^j\|^k \|\alpha\|^r.$$

(iii) Continuing with the hypotheses of (ii), show that $\Delta = \inf_n \|\alpha^n\|^{1/n}$ is well defined and, by using the result of (ii), or otherwise, that

$$\|\alpha^n\|^{1/n} \to \Delta.$$

We call Δ the spectral radius of α and write $\rho(\alpha) = \Delta$.

(iv) If $\alpha : \mathbb{R}^m \to \mathbb{R}^m$ is diagonalisable show that

$$\rho(\alpha) = \max\{|\lambda| : \lambda \text{ an eigenvalue of } \alpha\}.$$

(v) Give an example of linear maps $\alpha, \beta : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\rho(\alpha) = \rho(\beta) = 0$ but $\rho(\alpha + \beta) = 1$.

(vi) Recalling Question 20.11 of the third sheet, show that if $\rho(\iota - \gamma) < 1$ then γ is invertible.

The remaining questions are included for general interest and not for relevance to the syllabus or to passing Tripos exams.

Q 21.13. We work in $\mathbb{R}^m$, $\|\mathbf{x} - \mathbf{y}\|$ will represent the usual Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$.

(i) If K is a closed non-empty bounded set and $\mathbf{x}$ is any point, show that there exists a point $\mathbf{k}' \in K$ such that

$$\|\mathbf{x} - \mathbf{k}\| \geq \|\mathbf{x} - \mathbf{k}'\|$$

for all $\mathbf{k} \in K$. Is $\mathbf{k}'$ necessarily unique?

(ii) If E is a non-empty closed set and $\mathbf{x}$ is any point, show that there exists a point $\mathbf{e}' \in E$ such that

$$\|\mathbf{x} - \mathbf{e}\| \geq \|\mathbf{x} - \mathbf{e}'\|$$

for all $\mathbf{e} \in E$. We write $d(\mathbf{x}, E) = \|\mathbf{x} - \mathbf{e}'\|$.

(iii) With the notation of (ii) show that

$$d(\mathbf{x}, E) + \|\mathbf{x} - \mathbf{y}\| \geq d(\mathbf{y}, E).$$

Show that $d(\ , E) : \mathbb{R}^m \to \mathbb{R}$ is continuous.

(iv) If E is closed and non-empty and K closed, bounded and non-empty show that there exist $\mathbf{e}' \in E$ and $\mathbf{k}' \in K$ such that

$$\|\mathbf{e} - \mathbf{k}\| \geq \|\mathbf{e}' - \mathbf{k}'\|.$$

Would this result be true if we only assumed E and K closed and non-empty.

Q 21.14. (i) Show that if $f : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π such that $\hat{f}(n) = 0$ for all n then

$$\int_{-\pi}^{\pi} (1 - \epsilon_1 + \epsilon_2 \cos t)^N f(t) \, dt = 0$$

for all N.

(ii) Show that, given $\delta > 0$ we can find $\epsilon_1, \ \epsilon_2 > 0$ and an $\eta > 0$ such that

$$(1 - \epsilon_1 + \epsilon_2 \cos t)^N \to \infty \text{ uniformly for } |t| \le \eta$$
$$(1 - \epsilon_1 + \epsilon_2 \cos t)^N \to 0 \text{ uniformly for } \delta \le |t| \le \pi$$

as $N \to \infty$.

(iii) Show that if $f : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π such that $f(0)$ is real and $f(0) > 0$ we can find $\epsilon_1, \ \epsilon_2 > 0$ such that

$$\Re \int_{-\pi}^{\pi} (1 - \epsilon_1 + \epsilon_2 \cos t)^N f(t) \, dt \to \infty$$

as $N \to \infty$.

(iv) Show that if $f : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π such that $\hat{f}(n) = 0$ for all n then $f(t) = 0$ for all t.

(v) By using Question 21.5, or otherwise, show that if if $F : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π such that $\sum |\hat{F}(n)|$ converges then

$$F(t) = \sum_{n=-\infty}^{\infty} \hat{F}(n) \exp(int).$$

Q 21.15. (i) Suppose that $g : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π. Show that

$$\frac{1}{2\pi} \int_0^{2\pi} \left| g(t) - \sum_{r=-N}^{N} \hat{g}(r) \exp(irt) \right|^2 dt = \frac{1}{2\pi} \int_0^{2\pi} |g(t)|^2 \, dt - \sum_{r=-N}^{N} |\hat{g}(r)|^2.$$

Deduce that

$$\sum_{r=-N}^{N} |\hat{g}(r)|^2 \le \frac{1}{2\pi} \int_0^{2\pi} |g(t)|^2 \, dt.$$

(This is a version of Bessel's inequality.)

(ii) Now suppose that $f : \mathbb{R} \to \mathbb{C}$ is a continuously differentiable periodic function with period 2π. If f has derivative g, obtain a simple relation between $\hat{f}(n)$ and $\hat{g}(n)$. By applying the Cauchy-Schwarz inequality to $\sum_{n \neq 0, \ |n| \leq N} |n| |\hat{g}(n)|$ show that

$$\sum_{n \neq 0, \ |n| \leq N} |\hat{f}(n)| \leq A \frac{1}{2\pi} \int_0^{2\pi} |g(t)|^2 \, dt,$$

for some constant A. Conclude that $\sum |\hat{f}(n)|$ converges.

(iii) Deduce, using Question 21.14, that if $f : \mathbb{R} \to \mathbb{C}$ is a continuously differentiable periodic function with period 2π then

$$f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n) \exp(int).$$

Q 21.16. We continue with the ideas of Question 21.13. From now on all sets will belong to the collection $\mathcal{K}$ of closed bounded non-empty subsets of $\mathbb{R}^m$. Define

$$d(E, F) = \sup_{e \in E} d(e, F) + \sup_{f \in F} d(f, E).$$

Show that d is a metric on $\mathcal{K}$.

Show also that d is a complete metric.

Q 21.17. If (X, d) is a complete metric space and $T : X \to X$ is a surjective map such that

$$d(Tx, Ty) \geq K d(x, y)$$

for all $x, y \in X$ and some $K > 1$ show that T has a unique fixed point.

By considering the map $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = 1 + 4n + 2x$ for $0 \leq x < 1$ and n an integer, or otherwise show that the condition T surjective can not be dropped.

Q 21.18 (A continuous nowhere differentiable function). The following construction is due to Van der Waerden.

(i) Sketch the graph of the function g given by the condition

$$g(x + k) = |x| \quad \text{if } k \text{ is any integer and } -1/2 < x \leq 1/2.$$

(ii) Let $g_n(t) = 2^{-n} g(8^n t)$. Sketch the graphs of g_1, g_2 and g_3.

(iii) Let $f_n(t) = g_1(t) + g_2(t) + \ldots + g_n(t)$. Sketch the graphs of f_1, f_2 and f_3.

(iv) Show that f_n converges uniformly to a continuous function f.

(v) If n is a positive integer and r an odd integer give an expression for $f(r2^{-n})$. Show that when n is large then $2^n|f((r+2)2^{-n}) - f(r2^{-n})|$ is large. Conclude that if $r2^{-n} < t < (r+2)2^{-n}$ then

$$\max\left(\left|\frac{f(t) - f(r2^{-n})}{t - r2^{-n}}\right|, \left|\frac{f((r+2)2^{-n}) - f(t)}{(r+2)2^{-n} - t}\right|\right).$$

(vi) By developing the ideas of part (v) show that f is nowhere differentiable.

Final Note To Supervisors

Let me reiterate my request for corrections and improvements *particularly to the exercises*. The easiest path for me is e-mail. My e-mail address is `twk@dpmms`. I have to express my gratitude to Drs Barden and Pinch for finding a multitude of errors but I am sure that once the first veil of error has been lifted a further veil of deeper errors will appear.