Notes for Math2320
Analysis 1, 2002

John Hutchinson

email: John.Hutchinson@anu.edu.au

CHAPTER 1

Ignore this page, it is merely for typesetting purposes

1. Preliminaries

This course is both for students who do not intend at this stage doing any further theoretical mathematical courses, and also as a prerequisite for the course Analysis 2 (Lebesgue Integration and Measure Theory; and Hilbert Spaces). Interesting and significant applications to other areas of mathematics and to fields outside mathematics will be given, and the underlying theoretical ideas will be developed.

The main material for the course will be the text by Reed and these supplementary, more theoretical, notes. I will also occasionally refer to the text by Adams (fourth edition), last years MATH1115 Foundations Notes and the (supplementary) MATH1116 Calculus Notes.

There are two different printings of Reed, the second corrects some minor mistakes from the first. Most of you will have the second printing. You can tell which version you have by looking at the first expression on line 1 of page 53; the first printing has $2^{-n}(M-b)$ and the second printing has $2^{-n+1}(M-b)$. I will use the second version as the standard, but will make comments where it varies from the first.

Thanks to the many students in 2000 who pointed out various mistakes, particularly Tristan Bice, Kesava Jay, Ashley Norris and Griff Ware.

1.1. Real number system. Study Reed Chapter 1.1, it should be all review material. I will also refer ahead to parts of Chapter 2 of Reed, particularly in the $\star$ sections.

> *We begin with a review of the axioms for the real numbers. However, we will not usually make deductions directly from the axioms.*

Warning: What I call the *Cauchy completeness axiom* is called the "Completeness axiom" by Reed. What I call the *Completeness axiom* is not given a name by Reed. The terminology I use is more standard, and agrees with Adams and last years 1115 Notes.

Axioms for the real numbers. Recall that the real number system is a set[1] $\mathbb{R}$, together with two binary operations[2] "+" and "×", a binary relation[3] "$\leq$", and two particular members of $\mathbb{R}$ denoted 0 and 1 respectively. Moreover, it is required that certain axioms hold.

The first two sets of axioms are the *algebraic* and the *order axioms* respectively, see Reed (P1–P9) and (O1–O9)[4]. These axioms also hold for the rationals (but not for the irrationals, or the complex numbers).

In order to discuss the remaining two axioms, we need the set of *natural numbers* defined by

$$\mathbb{N} = \{1, 1+1, 1+1+1, \dots\} = \{1, 2, 3, \dots\}.$$

Unless otherwise clear from the context, the letters m, n, i, j, k will always denote natural numbers, or sometimes more generally will denote *integers*.

---

[1] We discuss sets in the next section

[2] A *binary* operation on $\mathbb{R}$ is a function which assigns to any *two* numbers in $\mathbb{R}$ a third number in $\mathbb{R}$. For example, the binary operation "+" assigns to the two real numbers a and b the real number denoted by $a+b$, and so in particular to the two numbers 2 and 3.4 the number 5.4.

[3] To say "$\leq$" is a *binary relation* means that for any two real numbers a and b, the expression $a \leq b$ is a *statement* and so is either true or false.

[4] See also the MATH1115 Foundations Notes, but there we had axioms for $<$ instead of $\leq$. This does not matter, since we can derive the properties of either from the axioms for the other. In fact, last year we did this for the properties of $\leq$ from axioms for $<$.

The two remaining axioms are:

- **The Archimedean axiom**[5]: *if $b > 0$ then there is a natural number n such that $n > b$.*
- **The Cauchy completeness axiom**: *for each Cauchy sequence*[6] *(a_n) of real numbers there is a real number a to which the sequence converges.*

This is the full set of axioms. All the standard properties of real numbers follow from these axioms.[7]

The (*usual*) **Completeness axiom**[8] is:

if a set of real numbers is bounded above, then it has a least upper bound.

This is equivalent to the Archimedean axiom plus the Cauchy completeness axiom. More precisely, if we assume the algebraic and order axioms then one can prove (see the following Remark) that[9]:

$$\begin{array}{c}\text{Archimedean axiom}\\ + \\ \text{Cauchy completeness axiom}\end{array} \iff \text{Completeness axiom}$$

Thus, from now on, we will assume both the Archimedean axiom and the Cauchy completeness axiom, and as a consequence also the (standard) Completeness "axiom".

Reed only assumes the algebraic, order and Cauchy completeness axioms, but not the Archimedean axiom, and then *claims* to prove first the Completeness axiom (Theorem 2.5.1) and from this the Archimedean axiom (Theorem 2.5.2). The proof in Theorem 2.5.2 is correct, but there is a mistake in the proof of Theorem 2.5.1, as we see in the following Remark. In fact, not only is the proof wrong, but in fact it is impossible to prove the Completeness axiom from the algebraic, order and Cauchy completeness axioms, as we will soon discuss[10].

REMARK 1.1.1.$\star$ The mistake in the "proof" of Theorem 2.5.1 is a hidden application of the Archimedean axiom. On page 53 line 6 Reed says "choose n so that $2^{-n}(M-b) \leq \varepsilon/2 \ldots$ ". But this is the same as choosing n so that $2^n \geq \frac{2(M-b)}{\varepsilon}$, and for this you really need the Archimedean axiom. For example, you could choose $n > \frac{2(M-b)}{\varepsilon}$ by the Archimedean axiom and then if follows that $2^n > \frac{2(M-b)}{\varepsilon}$ (since $2^n > n$ by algebraic and order properties of the reals).

What Reed really proves is that (assuming the algebraic and order axioms):

$$\begin{array}{c}\text{Archimedean axiom}\\ + \\ \text{Cauchy completeness axiom}\end{array} \implies \text{Completeness axiom}$$

---

[5]This is equivalent to the Archimedean "property" stated in Reed page 4 line 6-, which says *for any two real numbers $a, b > 0$ there is a natural number n such that $na > b$.*

Take $a = 1$ in the version in Reed to get the version here. Conversely, the version in Reed follows from the version here by first replacing b in the version here by b/a to get that $n > b/a$ for some natural number n, and then multiplying both sides of this inequality by a to get $na > b$ — this is all justified since we know the usual algebraic and order properties follow from the algebraic and order axioms.

[6]See Reed p.45 for the definition of a Cauchy sequence.

[7]Note that the rationals do not satisfy the corresponding version of the Cauchy completeness axiom. For example, take any irrational number such as $\sqrt{2}$. The decimal expansion leads to a sequence of rational numbers which is Cauchy, but there is no *rational* number to which the sequence converges. See MATH1115 Notes top of p.15. The rationals *do* satisfy the Archimedean axiom, why?

[8]See Adams page 4 and Appendix A page 23, also the MATH1115 Notes page 8

[9]"$\implies$" means "implies" and "$\iff$" means "implies and is implied by"

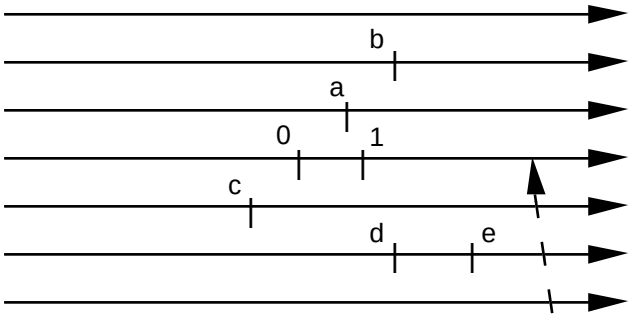
[10]Not a very auspicious beginning to the book, but fortunately that is the only really serious blunder.

The proof of "$\Longleftarrow$" is about the same level of difficulty. In Theorem 2.5.2, Reed proves in effect that the Completeness axiom implies the Archimedean axiom. The proof that the Completeness axiom implies the Cauchy completeness axiom is not too hard, and I will set it later as a $\star$ exercise.

REMARK 1.1.2.$\star\star$ We next ask: if we were smarter, would there be a way of avoiding the use of the Archimedean axiom in the proof of Theorem 2.5.1? The answer is *No*. The reason is that there are models in which the algebraic, order and Cauchy completeness axioms are true but the Archimedean axiom is false; these models are sometimes called *hyper-reals*, see below. If we had a proof that the algebraic, order and Cauchy completeness axioms implied the Completeness axiom (i.e. a correct proof of Theorem 2.5.1), then combining this with the (correct) proof in Theorem 2.5.2 we would end up with a proof that the algebraic, order and Cauchy completeness axioms imply the Archimedean axiom. But this would contradict the properties of the "hyper-reals".

These hyper-reals are quite difficult to construct rigorously, but here is a rough idea.

Part of any such model looks like a "fattened up" copy of $\mathbb{R}$, in the sense that it contains a copy of $\mathbb{R}$ together with "infinitesimals" squeezed between each real a and all reals greater than a. This part is followed and preceded by infinitely many "copies" of itself, and between any two copies there are infinitely many other copies. See the following crude diagram.



A property of absolute values. Note the properties of absolute value in Proposition 1.1.2 of Reed. Another useful property worth remembering is

$$\big|\,|x| - |y|\,\big| \leq |x - y|$$

Here is a proof from the triangle inequality, Prop. 1.1.2(c) in Reed..

PROOF.

$$\begin{aligned} |x| &= |x - y + y| \\ &\leq |x - y| + |y| \quad \text{by the triangle inequality.} \end{aligned}$$

Hence

$$|x| - |y| \leq |x - y| \tag{1}$$

and similarly

$$|y| - |x| \leq |y - x| = |x - y|. \tag{2}$$

Since $\big|\, |x| - |y| \,\big| = |x| - |y|$ or $|y| - |x|$, the result follows from (1) and (2). □

1.2. Sets and Functions. Study Reed Chapter 1.2. I provide comments and extra information below.

Sets and their properties. The notion of a set is basic to mathematics. In fact, it is possible in principle to reduce all of mathematics to set theory. But in practice, this is usually not very useful.

Sets are sometimes described by listing the members, and more often by means of some property. For example,

$$\mathbb{Q} = \{\, m/n \mid m \text{ and } n \text{ are integers, } n \neq 0 \,\}.$$

If a is a member of the set A, we write $a \in A$. Sometimes we say a is an *element* of A.

Note the definitions of $A \cap B$, $A \cup B$, $A \setminus B$, A^c and $A \subseteq B$ [11]. For given sets A and B, the first four are sets, and the fifth is a statement that is either true or false. Note that A^c is not well defined unless we know the set containing A in which we are taking the complement (this set will usually be clear from the context and it is often called the *universal set*). Often A is a set of real numbers and $\mathbb{R}$ is the universal set.

We say for sets A and B that $A = B$ if they have the same elements. This means that every member of A is a member of B and every member of B is also a member of A, that is $A \subseteq B$ and $B \subseteq A$. We usually prove $A = B$ by proving the two separate statements $A \subseteq B$ and $B \subseteq A$. For example, see the proof below of the first of DeMorgan's laws.

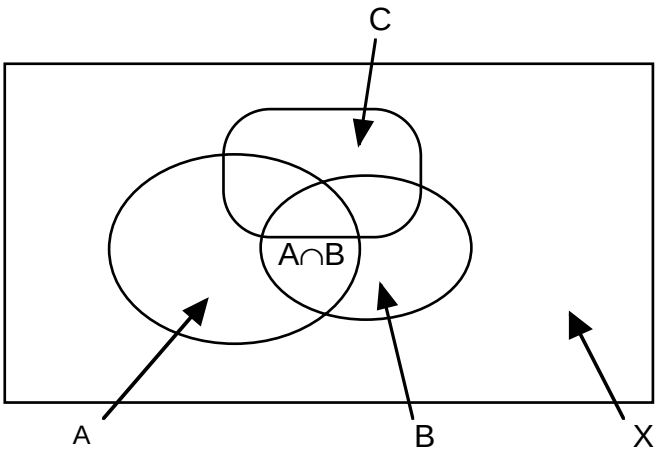
THEOREM 1.2.1 (De Morgan's laws). *For any sets A and B*

$$\begin{aligned} (A \cap B)^c &= A^c \cup B^c \\ (A \cup B)^c &= A^c \cap B^c \\ A \cap (B \cup C) &= (A \cap B) \cup (B \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (B \cup C) \end{aligned}$$

REMARK 1.2.1. The following diagram should help motivate the above. Which regions represent the eight sets on either side of one of the above four equalities?

The proofs are also motivated by the diagram. But note that all the proofs really use are the definitions of $\cap$, $\cup$, ^{c} and the logical meanings of *and*, *or*, *not*.

[11] We occasionally write $A \subset B$ instead of $A \subseteq B$. Some texts write $A \subset B$ to mean that $A \subseteq B$ but $A \neq B$; we will not do this.



PROOF. We will prove the first equality by showing $(A \cap B)^c \subseteq A^c \cup B^c$ and $A^c \cup B^c \subseteq (A \cap B)^c$.

First assume $a \in (A \cap B)^c$ (note that $a \in X$ where X is the universal set).
Then it is *not* the case that $a \in A \cap B$.
In other words: it is *not* the case that ($a \in A$ and $a \in B$).
Hence: (it is not the case that $a \in A$) or (it is not the case that $a \in B$).[12]
That is: $a \in A^c$ or $a \in B^c$ (remember that $a \in X$).
Hence: $a \in A^c \cup B^c$.

Since a was an arbitrary element in $(A \cap B)^c$, it follows that $(A \cap B)^c \subseteq A^c \cup B^c$.

The argument above can be written in essentially the reverse order (*check!*) to show that $A^c \cup B^c \subseteq (A \cap B)^c$.

This completes the proof of the first equality

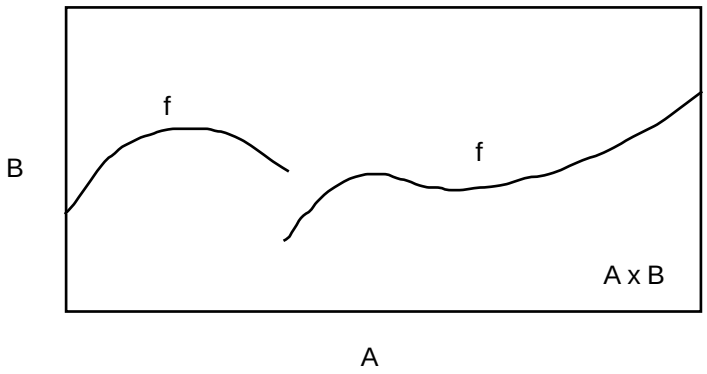
The proofs of the remaining equalities are similar, and will be left as exercises. □

Products of sets. Recall that

$$A \times B = \{\, (a, b) \mid a \in A,\ b \in B \,\}$$

For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Here (a, b) is an *ordered pair*, not an open interval!

We can represent $A \times B$ schematically as follows



REMARK 1.2.2.⋆ It is interesting to note that we can define ordered pairs in terms of sets. Important properties required for an ordered pair (a, b) are that it depends on the two objects a and b and *the order is important*, i.e. $(a, b) \neq (b, a)$

[12] If P and Q are two statements, then "not (P and Q)" has the same meaning as "(not P) or (not Q)".

unless $a = b$. This is different from the case for sets, i.e. $\{a, b\} = \{b, a\}$ for any a and b.

More generally, the only property we require of ordered pairs is that

$$(a, b) = (c, d) \quad \text{iff} \quad (a = c \text{ and } b = d). \tag{3}$$

There are a number of ways that we can define ordered pairs in terms of sets. The standard definition is

$$(a, b) := \{\{a\}, \{a, b\}\}.$$

To show this is a good definition, we need to prove (3).

PROOF. It is immediate from the definition that if $a = c$ and $b = d$ then $(a, b) = (c, d)$.

Next suppose $(a, b) = (c, d)$, i.e. $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$. We consider the two cases $a = b$ and $a \neq b$ separately.

If $a = b$ then $\{\{a\}, \{a, b\}\}$ contains exactly one member, namely $\{a\}$, and so $\{\{c\}, \{c, d\}\}$ also contains exactly the one member $\{a\}$. This means $\{a\} = \{c\} = \{c, d\}$. Hence $a = c$ and $c = d$. In conclusion, $a = b = c = d$.

If $a \neq b$ then $\{\{a\}, \{a, b\}\}$ contains exactly two (distinct) members, namely $\{a\}$ and $\{a, b\}$. Since $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ it follows $\{c\} \in \{\{a\}, \{a, b\}\}$ and so $\{c\} = \{a\}$ or $\{c\} = \{a, b\}$. The second equality cannot be true since $\{a, b\}$ contains two members whereas $\{c\}$ contains one member, and so $\{c\} = \{a\}$, and so $c = a$.

Since also $\{c, d\} \in \{\{a\}, \{a, b\}\}$ it now follows that $\{c, d\} = \{a, b\}$ (otherwise $\{c, d\} = \{a\}$, but since also $\{c\} = \{a\}$ this would imply $\{\{c\}, \{c, d\}\}$ and hence $\{\{a\}, \{a, b\}\}$ has only one member, and we have seen this is not so). Since a and b are distinct and $\{c, d\} = \{a, b\}$, it follows c and d are distinct; since $a = c$ it then follows $b = d$. In conclusion, $a = c$ and $b = d$.

This completes the proof. □

Power sets. The *power set* $\mathcal{P}(A)$ of a set A is defined to be the set of all subsets of A. That is

$$\mathcal{P}(A) = \{\, S \mid S \subseteq A \,\}.$$

1.2.1. *Functions.* We say f is a function from the set A into the set B, and write

$$f : A \to B,$$

if f "assigns" to *every* $a \in A$ exactly one member of B. This member of B is denoted by $f(a)$.

We say A is the *domain* of f, i.e. $Dom(f) = A$, and the set of all members in B of the form $f(a)$ is the *range* of f, i.e. $Ran(f) = \{\, f(a) \mid a \in A \,\}$.

Note that $Ran(f) \subseteq B$, but $Ran(f)$ need not be all of B. An example is the squaring function, see the next paragraph.

The convention on page 8 in Reed is different. He only requires that f assigns a value to *some* members of A, so that for him the domain of f need not be all of A. *This is not standard, and you should normally keep to the definition here.* In particular, in Reed Example 2 page 9 one should say f is a function from $\mathbb{R}^+$ [13] (not $\mathbb{R}$) to $\mathbb{R}$, and write $f : \mathbb{R}^+ \to \mathbb{R}$, since f only assigns a value to positive numbers.

We informally think of f as some sort of "rule". For example, the "squaring function" $f : \mathbb{R} \to \mathbb{R}$ is given by the rule $f(a) = a^2$ for each $a \in \mathbb{R}$. But functions

[13] $\mathbb{R}^+ = \{\, a \in \mathbb{R} \mid a > 0 \,\}$.

can be much more complicated than this, and in particular there may not really be any "rule" which explicitly describes the function.

The way to avoid this problem is to define functions in terms of sets. We can think of the squaring function as being given by the set of all ordered pairs $\{(a, a^2) \mid a \in \mathbb{R}\}$. This set is a subset of $\mathbb{R} \times \mathbb{R}$ and has the property that to *every* $a \in \mathbb{R}$ there corresponds *exactly one* element $b \in \mathbb{R}$ ($b = a^2$ for the squaring function). Thus we are really just identifying f with its graph.

More generally, we say *f is a function from A into B, and write $f : A \to B$, if f is a subset of $A \times B$ with the property that for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in f$.* See the preceding diagram.

Note that Reed uses a different notation for the function f thought of as a "rule" or "assignment" on the one hand, and the function f thought of as a subset of $A \times B$ on the other (he uses a capital F in the latter case). We will use he same notation in either case.

We normally *think* of f as an assignment, although the "assignment" may not be given by any rule which can readily be written down.

Note also that Reed uses the notation $a \xrightarrow{f} b$ to mean the same as $f(a) = b$. Reed's notation is not very common and we will avoid it. Do not confuse it with the notation $f : A \to B$.

2. Cardinality

Some infinite sets are bigger than others, in a sense that can be made precise.

Study Reed Chapter 1.3. In the following I make some supplementary remarks.

Note the definitions of *finite* and *infinite* sets.

A set is *countable* if it is infinite and can be put in one-one correspondence with $\mathbb{N}$, i.e. has the same cardinality as $\mathbb{N}$. (Some books include finite sets among the countable sets; if there is any likelihood of confusion, you can say *countably infinite*.) In other words, a set is countable iff it can be written as an infinite sequence (with all a_i distinct)

$$a_1, a_2, a_3, \ldots, a_n, \ldots.$$

It is *not* true that all infinite sets are countable, as we will soon see. If an infinite set is not countable then we say it is *uncountable*.

Thus every set is finite, countable or uncountable.

Important results in this section are that an infinite subset of a countable set is countable (Prop 1.3.2), the product of two countable sets is countable (Prop 1.3.4), the union of two countable sets is countable (Exercise 3), the set of rational numbers is countable (Theorem 1.3.5), the set of real numbers is uncountable (Theorem 1.3.6 and the remarks following that theorem), and the set of irrational numbers is uncountable (second last paragraph in Chapter 1.3).

A different proof of Proposition 1.3.4, which avoids the use of the Fundamental Theorem of Arithmetic and is more intuitive, is the following.

PROPOSITION 2.0.2. *If S and T are countable, then so is $S \times T$.*

PROOF. Let $S = (a_1, a_2, \ldots)$ and $T = (b_1, b_2, \ldots)$ Then $S \times T$ can be enumerated as follows:

$$\begin{array}{ccccccccc}
(a_1,b_1) & & (a_1,b_2) & \to & (a_1,b_3) & & (a_1,b_4) & \to & (a_1,b_5) & \ldots \\
\downarrow & & \uparrow & & \downarrow & & \uparrow & & \downarrow & \\
(a_2,b_1) & \to & (a_2,b_2) & & (a_2,b_3) & & (a_2,b_4) & & (a_2,b_5) & \ldots \\
 & & & & \downarrow & & \uparrow & & \downarrow & \\
(a_3,b_1) & \leftarrow & (a_3,b_2) & \leftarrow & (a_3,b_3) & & (a_3,b_4) & & (a_3,b_5) & \ldots \\
\downarrow & & & & & & \uparrow & & \downarrow & \\
(a_4,b_1) & \to & (a_4,b_2) & \to & (a_4,b_3) & \to & (a_4,b_4) & & (a_4,b_5) & \ldots \\
 & & & & & & & & \downarrow & \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & \ddots
\end{array}$$

□

It is not the case that all uncountable sets have the same cardinality. In fact $\mathcal{P}(A)$ always has a larger cardinality than A, in the sense that there is no one-one map from A *onto* $\mathcal{P}(A)$. (There is certainly a one-one map from A *into* $\mathcal{P}(A)$, define $f(A) = \{A\}$). Thus we can get larger and larger infinite sets via the sequence

$$\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\mathcal{P}(\mathbb{N})), \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N}))), \ldots.$$

(One can prove that $\mathcal{P}(\mathbb{N})$ and $\mathbb{R}$ have the same cardinality.)

⋆ In fact, this is barely the beginning of what we can do: we could take the union of all the above sets to get a set of even larger cardinality, and then take this as a new beginning set instead of using $\mathbb{N}$. And still we would hardly have begun! However, in practice we rarely go beyond sets of cardinality $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))$, i.e. of cardinality $\mathcal{P}(\mathcal{P}(\mathbb{R}))$

THEOREM 2.0.3. *For any set A, there is no map f from A* onto $\mathcal{P}(A)$. *In particular, A and $\mathcal{P}(A)$ have different cardinality.*

PROOF. * Consider any $f : A \to \mathcal{P}(A)$. We have to show that f is not onto.

Define

$$B = \{\, a \in A \mid a \notin f(a) \,\}.$$

Then $B \subseteq A$ and so $B \in \mathcal{P}(A)$. We *claim* there is no $b \in A$ such that $f(b) = B$. (This implies that f is not an onto map!)

Assume (in order to obtain a contradiction) that $f(b) = B$ for some $b \in A$. One of the two possibilities $b \in B$ or $b \notin B$ must be true.

If $b \in B$, i.e. $b \in f(b)$, then b does not satisfy the condition defining B, and so $b \notin B$.

If $b \notin B$, i.e. $b \notin f(b)$, then B does satisfy the condition defining B, and so $b \in B$.

In either case we have a contradiction, and so the *assumption* must be wrong. Hence there is no $b \in A$ such that $f(b) = B$ and in particular f is not an onto map. □

3. Sequences

Review the definitions and examples in Section 2.1 of Reed, and the propositions and theorems in Section 2.2.

3.1. Basic definitions. Recall that a sequence is an infinite list of real numbers (later, complex numbers or functions, etc.). We usually write

$$a_1, a_2, a_3, \ldots, a_n, \ldots ,$$

$(a_n)_{n\geq 1}$, $(a_n)_{n=1}^{\infty}$ or just (a_n). Sometimes we start the enumeration from 0 or another integer.

The sequence (a_n) converges to a if for any preassigned positive number (usually called ε), all members of the sequence beyond a certain member (usually called the Nth), will be within distance ε of a. In symbols:

for each $\varepsilon > 0$ there is an integer N such that

$$|a_n - a| \leq \varepsilon \quad \text{for all } n \geq N.$$

(Note that the definition implies N may (and usually will) depend on ε. We often write $N(\varepsilon)$ to emphasise this fact.)

Study Examples 1–4 in Section 2.1 of Reed. *Note* that the Archimedean axiom is needed in each of these three examples (page 30 line 9, page 31 line 8, page 32 line 1, page 33 line 3).

Note also the definitions of

1. *diverges*; e.g. if $a_n = n^2$ or $a_n = (-1)^n$, then (a_n) diverges.
2. *diverges to* $+\infty$; e.g. $n^2 \to \infty$, but *not* $(-1)^n n^2 \to \infty$

3.2. Convergence properties. From the definition of convergence we have the following useful theorems, see Reed Section 2.2 for proofs. While the results may appear obvious, this is partly because we may only have simple examples of sequences in mind (For a less simple example, think of a sequence which enumerates the rational numbers!).

Theorem 3.2.1.

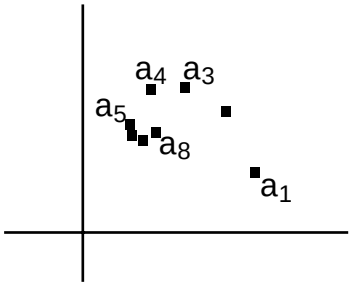
1. *If* (a_n) *converges, then* (a_n) *is bounded.*[14]
2. *If* $a_n \leq c_n \leq b_n$ *for all* n[15] *and both* $a_n \to L$, $b_n \to L$, *then* $c_n \to L$.
3. *If* $a_n \to a$ *and* $b_n \to b$ *then*
   (a) $a_n \pm b_n \to a \pm b$,
   (b) $ca_n \to ca$ *(*c *is a real number),*
   (c) $a_n b_n \to ab$,
   (d) $a_n/b_n \to a/b$, *(assuming* $b \neq 0$*)*[16]

3.3. Sequences in $\mathbb{R}^N$. We will often be interested in sequences of *vectors* (i.e. *points*) in $\mathbb{R}^N$.

[14] A sequence (a_n) is *bounded* if there is a real number M such that $|a_n| \leq M$ for all n.

[15] Or for all sufficiently large n, more precisely for all $n \geq N$ (say).

[16] Since $b \neq 0$, it follows that $b_n \neq 0$ for all $n \geq N$ (say). This implies that the sequence (a_n/b_n) is defined, at least for $n \geq N$.



Such a sequence is said to *converge* if each of the N sequences of components converges. For example,

$$\begin{bmatrix} 1 + \frac{(-1)^n}{n} \sin n \\ e^{-n} \end{bmatrix} \to \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The analogues of 1., 3(a), 3(b) hold, as is easily checked by looking at the components. More precisely

THEOREM 3.3.1.

1. *If* $(\mathbf{u}_n)$ *converges, then* $(\mathbf{u}_n)$ *is bounded.*[17]
2. *If* $\mathbf{u}_n \to \mathbf{u}$, $\mathbf{v}_n \to \mathbf{v}$ *and* $a_n \to a$ *(sequence of real numbers) then*
   (a) $\mathbf{u}_n \pm \mathbf{v}_n \to \mathbf{u} \pm \mathbf{v}$,
   (b) $a_n \mathbf{u}_n \to a\mathbf{u}$,
3. $A\mathbf{u}_n \to A\mathbf{u}$ *(where* A *is any matrix with* N *columns).*

PROOF.

1. (Think of the case $N = 2$) Since $(\mathbf{u}_n)$ converges, so does each of the N sequences of real numbers $(u_n^j)_{n\geq 1}$ (where $1 \leq j \leq N$). But this implies that each of these N sequences of real numbers lies in a fixed interval $[-M_j, M_j]$ for $j = 1, \ldots, N$. Let M be the maximum of the M_j. Since $|u_n^j| \leq M$ for all j and all n, it follows that $|\mathbf{u}_n| = \sqrt{(u_n^1)^2 + \cdots + (u_n^N)^2} \leq \sqrt{N}M$ for all n. This proves the first result.

2. Since each of the components of $\mathbf{u}_n \pm \mathbf{v}_n$ converges to the corresponding component of $\mathbf{u} \pm \mathbf{v}$ by the previous theorem, result 2(a) follows. Similarly for 2(b).

3. This follows from the fact that $A\mathbf{u}_n$ is a linear combination of the columns of A with coefficients given by the components of $\mathbf{u}_n$, while $A\mathbf{u}$ is the corresponding linear combination with coefficients given by the components of $\mathbf{u}$.

More precisely, write $A = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$, where $\mathbf{a}_1, \ldots, \mathbf{a}_N$ are the column vectors of A. Then

$$A\mathbf{u}_n = u_n^1 \mathbf{a}_1 + \cdots + u_n^N \mathbf{a}_N \quad \text{and} \quad A\mathbf{u} = u^1 \mathbf{a}_1 + \cdots + u^N \mathbf{a}_N$$

Since $u_n^1 \to u^1$, $\ldots$, $u_n^N \to u^N$ as $n \to \infty$, the result follows from 2(b) and repeated applications of 2(a). □

[17] A sequence of points $(\mathbf{u}_n)_{n\geq 1}$ in $\mathbb{R}^N$, where $\mathbf{u}_n = (u_n^1, \ldots, u_n^N)$ for each n, is *bounded* if there is a real number M such that $|\mathbf{u}_n| = \sqrt{(u_n^1)^2 + \cdots + (u_n^1)^2} \leq M$ for all n. This just means that all members of the sequence lie inside some single ball of sufficiently large radius M.

4. Markov Chains

Important examples of infinite sequences occur via Markov chains.

See Reed Section 2.3. There is a proof in Reed of Markov's Theorem (Theorem 4.0.2 in these notes) for the case of 2 states. But the proof is algebraically messy and does not easily generalise to more than 2 states. So you may omit Reed page 42 and the first paragraph on page 43. Later I will give a proof of Markov's Theorem using the Contraction Mapping Theorem.

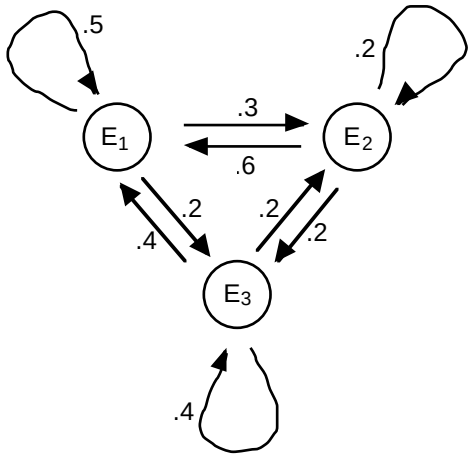
Suppose that a taxi is in one of three possible zones E_1, E_2 or E_3 (e.g. Civic, Belconnen or Woden). We are interested in its location at a certain sequence of designated times, lets say at each hour of the day.

Suppose that if the taxi is in E_1 at one hour, then the probability that it is in E_1, E_2, E_3 exactly one hour later is $.5, .3, .2$ respectively (see the first column of the following matrix, note that the sum of the probabilities must be 1). Similarly suppose that if it is in E_2 then the probability that it is in E_1, E_2, E_3 exactly one hour later is $.6, .2, .2$ (second column), and if it is in E_3 then the probability that it is in E_1, E_2, E_3 exactly one hour later is $.4, .2, .4$ (third column).

We can represent the situation by a matrix of probabilities,

$$\begin{bmatrix} .5 & .6 & .4 \\ .3 & .2 & .2 \\ .2 & .2 & .4 \end{bmatrix}.$$

The first column corresponds to starting in E_1, the second to E_2 and the third to E_3. We can also show this in a diagram as follows.



We are interested in the long-term behaviour of the taxi. For example, after a large number of hours, what is the probability that the taxi is in each of the states E_1, E_2, E_3, and does this depend on the initial starting position?

More generally, consider the following situation, called a *Markov process*. One has a system which can be in one of N states $E_1, \ldots, E_N$. $P = (p_{ij})$ is an $N \times N$ matrix where p_{ij} is the probability that if the system is in state E_j at some time then it will be in state E_i at the next time. Thus the jth column corresponds to what happens at the next time if the system is currently in state E_j. We say P is a *probability transition matrix*.

What happens after the system is in state E_2
$\downarrow$

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}$$

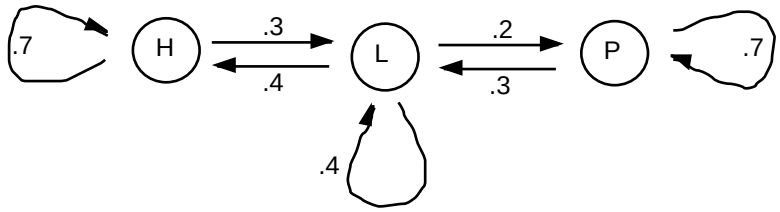
If a vector consists of numbers from the interval $[0, 1]$ and the sum of the components is one, we say the vector is a *probability vector*. We require the columns of a probability transition matrix to be probability vectors.

There are many important examples:

Cell genetics: Suppose a cell contains k particles, some of type A and some of type B. The state of the cell is the number of particles of type A, so there are $N = k + 1$ states. Suppose daughter cells are formed by subdivision. Then one can often compute from biological considerations the probability that a random daughter cell is in each of the N possible states.

Population genetics: Suppose flowers are bred and at each generation 1000 flowers are selected at random. A particular gene may occur either 0 or 1 or 2 times in each flower and so there are between 0 and 2000 occurrences of the gene in the population. This gives 2001 possible states, and one usually knows from biological considerations what is the possibility of passing from any given state for one generation to another given state for the next generation.

Random walk: Suppose an inebriated person is at one of three locations; home (H), lamppost (L) or pub (P). Suppose the probability of passing from one to the other an hour later is given by the following diagram:



The corresponding matrix of transition probabilities, with the three columns corresponding to being in the state H, L, P respectively, is

$$\begin{bmatrix} .7 & .4 & 0 \\ .3 & .4 & .3 \\ 0 & .2 & .7 \end{bmatrix}$$

Other examples occur in diffusion processes, queueing theory, telecommunications, statistical mechanics, etc. See *An Introduction to Probability Theory and Its Applications*, vol 1, by W. Feller for examples and theory.

We now return to the general theory. Suppose that at some particular time the probabilities of being in the states $E_1, \ldots, E_N$ is given by the probability vector

$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$ (so the sum of the a_i is one). Then the probability at the next time of being in state E_i is[18]

$$\begin{aligned} a_1 \times &(\text{prob. of moving from } E_1 \text{ to } E_i) + a_2 \times (\text{prob. of moving from } E_2 \text{ to } E_i) \\ &+ \cdots + a_N \times (\text{prob. of moving from } E_N \text{ to } E_i) \\ &= p_{i1}a_1 + p_{i2}a_2 + \cdots + p_{iN}a_N \\ &= (P\mathbf{a})_i. \end{aligned}$$

Thus if at some time the probabilities of being in each of the N states is given by the vector $\mathbf{a}$, then the corresponding probabilities at the next time are given by $P\mathbf{a}$, and hence at the next time by $P(P\mathbf{a}) = P^2\mathbf{a}$, and hence at the next time by $P^3\mathbf{a}$, etc.

In the case of the taxi, starting at Civic and hence starting with a probability vector $\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, the corresponding probabilities at later times are given by $P\mathbf{a} = \begin{bmatrix} .5 \\ .3 \\ .2 \end{bmatrix}$, $P^2\mathbf{a} = \begin{bmatrix} .51 \\ .25 \\ .24 \end{bmatrix}$, $P^3\mathbf{a} = \begin{bmatrix} .501 \\ .251 \\ .248 \end{bmatrix}$, $\dots$. This appears to converge to $\begin{bmatrix} .5 \\ .25 \\ .25 \end{bmatrix}$

In fact it seems plausible that after a sufficiently long period of time, the probability of being in any particular zone should be almost independent of where the taxi begins. This is indeed the case (if we pass to the limit), as we will see later.

We return again to the general situation. Suppose $\mathbf{a}$ is any probability vector with N components. We will later prove:

THEOREM 4.0.2 (Markov). *If all entries in the probability transition matrix P are greater than 0, and $\mathbf{a}$ is a probability vector, then the sequence of vectors*

$$\mathbf{a}, P\mathbf{a}, P^2\mathbf{a}, \dots,$$

converges to a probability vector $\mathbf{v}$, and this vector does not depend on $\mathbf{a}$.

Moreover, the same results are true even if we just assume P^k has all entries non-zero for some integer $k > 1$.

The vector $\mathbf{v}$ is the unique non-zero solution of $(P - I)\mathbf{v} = \mathbf{0}$.[19]

The theorem applies to the taxi, and to the inebriated person (take $k = 2$). The proof that if $\mathbf{v}$ is the limit then $(P - I)\mathbf{v} = \mathbf{0}$ can be done now.

PROOF. Assume $\mathbf{v} = \lim P^n\mathbf{a}$. Then

$$\begin{aligned} \mathbf{v} &= \lim_{n\to\infty} P^n\mathbf{a} \\ &= \lim_{n\to\infty} P^{n+1}\mathbf{a} \quad \text{(as this is the same sequence, just begun one term later)} \\ &= \lim_{n\to\infty} PP^n\mathbf{a} \\ &= P \lim_{n\to\infty} P^n\mathbf{a} \quad \text{(from Theorem 3.3.1.3)} \\ &= P\mathbf{v}. \end{aligned}$$

Hence $(P - I)\mathbf{v} = \mathbf{0}$. □

[18] From intuitive ideas of probability, if you have not done any even very basic probability theory.

[19] The matrix I is the identity matrix of the same size as P.

For the taxi, the non zero solution of

$$\begin{bmatrix} -.5 & .6 & .4 \\ .3 & -.8 & .2 \\ .2 & .2 & -.6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

with entries summing to 1 is indeed (see previous discussion) $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} .5 \\ .25 \\ .25 \end{bmatrix}$ and so in the long term the taxi spends 5025

For the inebriate, the non zero solution of

$$\begin{bmatrix} -.3 & .4 & 0 \\ .3 & -.6 & .3 \\ 0 & .2 & -.3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 4/9 \\ 1/3 \\ 2/9 \end{bmatrix}$ and so in the long term (as the evening progresses) the inebriate spends 44% of the time at home, 33% by the lamppost and 22% in the pub.

5. Cauchy sequences and the Bolzano-Weierstraß theorem

A sequence converges iff it is Cauchy. This gives a criterion for convergence which does not require knowledge of the limit. It will follow that a bounded monotone sequence is convergent.

A sequence need not of course converge, even if it is bounded. But by the Bolzano-Weierstraß theorem, every bounded sequence has a convergent subsequence.

Read Sections 2.4–6 of Reed, most of it will be review. The following notes go through the main points.

5.1. Cauchy sequences. (Study Section 2.4 of Reed.)
We first give the definition.

DEFINITION 5.1.1. A sequence (a_n) is a *Cauchy sequence* if for each number $\epsilon > 0$ there exists an integer N such that

$$|a_m - a_n| \leq \epsilon \quad \text{whenever} \quad m \geq N \text{ and } n \geq N.$$

The idea is that, given any number ("tolerance") $\varepsilon > 0$, all terms (*not* just successive ones) of the sequence from the Nth onwards are within (this tolerance) ε of one another (N will normally depend on the particular "tolerance"). The smaller we choose ε the larger we will need to choose N.

See Example 1 page 45 of Reed (note that the Archimedean axiom is used on page 45 line 8-).[20]

THEOREM 5.1.2. *A sequence is Cauchy iff it converges.*

The proof that convergent implies Cauchy is easy (Proposition 2.4.1 of Reed). The fact that Cauchy implies convergent is just the Cauchy Completeness Axiom (or what Reed calls the Completeness axiom, see Reed page 48).

Recall that a sequence (a_n) is *increasing* if $a_n \leq a_{n+1}$ for all n, and is *decreasing* if $a_n \leq a_{n+1}$ for all n. A sequence is *monotone* if it is either increasing or decreasing.

THEOREM 5.1.3. *Every bounded monotone sequence converges.*

For the proof of the theorem see Reed page 48. The proof uses a bisection argument. Begin with a closed interval which contains all members of the sequence, and keep subdividing it so that all members of the sequence after some index are in the chosen subinterval. This implies the sequence is Cauchy and so it converges by the previous theorem. (Note the use of the Archimedean Axiom on page 49 line 10- of Reed.)

Loosely speaking, the sequence of chosen subintervals "converges to a point" which is the limit of the original sequence.

5.2. Subsequences and the Bolzano-Weierstraß theorem. Study Section 2.6 of Reed (omitting the Definition and Proposition on page 56). Also read Section 3.5 of the MATH1115 Notes and Adams, Appendix III, Theorem 2.

Suppose (a_n) is a sequence of real numbers. A *subsequence* is just a sequence obtained by skipping terms. For example,

$$a_1, a_{27}, a_{31}, a_{44}, a_{101}, \ldots$$

is a subsequence. We usually write a subsequence of (a_n) as

$$a_{n_1}, a_{n_2}, a_{n_3}, a_{n_4}, \ldots, a_{n_k}, \ldots, \quad \text{or}$$
$$a_{n(1)}, a_{n(2)}, a_{n(3)}, a_{n(4)}, \ldots, a_{n(k)}, \ldots, \quad \text{or}$$

[20] line 8- means 8 lines from the bottom of the page.

$$(a_{n_k})_{k\geq 1}, \quad \text{or} \quad (a_{n(k)})_{k\geq 1}$$

Thus in the above example we have $n_1 = 1, n_2 = 27, n_3 = 31, n_4 = 44, n_5 = 101, \ldots$.

Instead of the Definition on page 56 of Reed, you may use Proposition 2.6 as the definition. That is,

DEFINITION 5.2.1. The number d is a *limit point* of the sequence $(a_n)_{n\geq 1}$ if there is a subsequence of $(a_n)_{n\geq 1}$ which converges to d.

The *limit* of a convergent sequence is also a *limit point* of the sequence. Moreover, it is the only limit point, because *any subsequence of a convergent sequence must converge to the same limit as the original sequence* (why?). On the other hand, sequences which do not converge may have many limit points, even infinitely many (see the following examples).

Now look at Examples 1 and 2 on pages 56, 57 of Reed.

Here is an example of a sequence which has every *every* number in $[0, 1]$ as a limit point. Let

$$r_1, r_2, r_3, \ldots \tag{4}$$

be an enumeration of all the rationals in the interval $[0, 1]$. We know this exists, because the rationals are countable, and indeed we can explicitly write down such a sequence.

Every number $a \in [0, 1]$ has a decimal expansion, and this actually provides a sequence (a_n) of rationals converging to a (modify the decimal expansion a little to get distinct rationals if necessary).

This sequence may not be a subsequence of (4), since the a_n may occur in a different order. But we can find a subsequence $(a_{n'})$ of (a_n) (which thus also converges to a) and which is also a subsequence of (4) as follows:

$$\begin{aligned}
a_{1'} &= a_1 \\
a_{2'} &= \text{first term in } (a_n) \text{ which occurs } \textit{after } a_{1'} \text{ in (4)} \\
a_{3'} &= \text{first term in } (a_n) \text{ which occurs } \textit{after } a_{2'} \text{ in (4)} \\
a_{4'} &= \text{first term in } (a_n) \text{ which occurs } \textit{after } a_{3'} \text{ in (4)} \\
&\vdots
\end{aligned}$$

THEOREM 5.2.2 (Bolzano-Weierstraß Theorem). *If a sequence (a_n) is bounded then it has a convergent subsequence. If all members of (a_n) belong to a closed bounded interval I, then so does the limit of any convergent subsequence of (a_n).*

Note that the theorem implies that any bounded sequence must have at least one limit point. See Reed pages 57, 58.

The proof of the theorem is via a bisection argument. Suppose I is a closed bounded interval containing all terms of the sequence. Subdivide I and choose one of the subintervals so it contains an infinite subsequence, then again subdivide and takes one of the new subintervals with an infinite subsequence of the first subsequence, etc. Now define a *new* subsequence of the original sequence by taking one term from the first subsequence, a later term from the second subsequence, a still later term from the third subsequence, etc. One can now show that this subsequence is Cauchy and so converges.

5.3. Completeness property of the reals. (See Section 2.5 of Reed, but note the errors discussed below).

The usual Completeness Axiom for the real numbers is:[21]

> *if a non empty set A of real numbers is bounded above, then there is a real number b which is the least upper bound for A.*

The real number b need not belong to A, but it is unique. For example, the least upper bound for each of the sets $(0, 1)$ and $[0, 1]$ is 1.

A common term for the "least upper bound" is the *supremum* or *sup*. Similarly, the greatest lower bound of a set is often called the *infimum* or *inf*.

See the MATH1115 2000 notes, §2.3 for a discussion. See also Adams, third edition, pages 4, A-23.

The Completeness Axiom in fact follows from the other (i.e. algebraic, order, Archimedean and Cauchy completeness) axioms. It is also possible to prove the Archimedean and Cauchy completeness axioms from the algebraic, order and Completeness axioms.

You should now go back and reread Section 1.1 of these notes.

Remarks on the text: Note that the remark at the end of Section 2.5 of Reed about the equivalence of Theorems 2.4.2, 2.4.3 and 2.5.1 is *only* correct if we assume the algebraic, order *and* Archimedean axioms. The assertion on page 53 before Theorem 2.5.2 in the corrected printing, that the Archimedean property follows from Theorem 2.5.1, is not correct, as Theorem 2.5.1 uses the Archimedean property in its proof (in line 6 of page 53).

[21] Note that a similar statement is not true of we replace "real" by "rational". For example, let

$$A = \{\, x \in \mathbb{Q} \mid x \geq 0,\ x^2 < 2 \,\}.$$

Then although A is a nonempty set of rational numbers which is bounded above, there is no *rational* least upper bound for A. The (real number) least upper bound is of course $\sqrt{2}$.

6. The Quadratic Map

This is the prototype of maps which lead to chaotic behaviour. For some excellent online software demonstrating a wide range of chaotic behaviour, go to Brian Davies' homepage, via the Department of Mathematics homepage. When you reach the page on "Exploring chaos", click on the "graphical analysis" button.

Read Section 2.7 from Reed. We will only do this section lightly, just to indicate how quite complicated behaviour can arise from seemingly simple maps.

The *quadratic map* is defined by the recursive relation

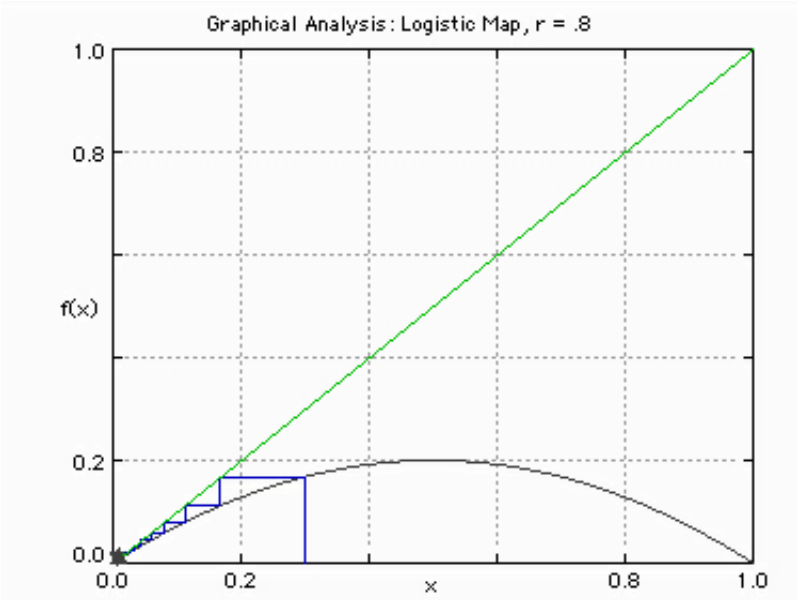
$$x_{n+1} = rx_n(1 - x_n), \tag{5}$$

where $x_0 \in [0, 1]$ and $0 < r \leq 4$. It is a simple model of certain population models, where x_n is the fraction in year n of some "maximum possible" population.

The function $F(x) = rx(1 - x)$ is positive on the interval $[0, 1]$. Its maximum occurs at $x = 1/2$ and has the value $a/4$. It follows that

$$F : [0, 1] \to [0, 1]$$

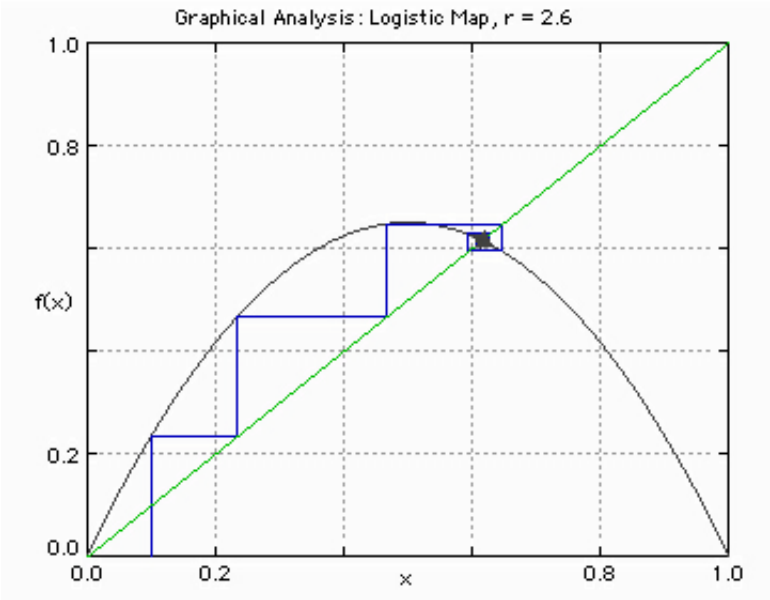
for r in the given range.



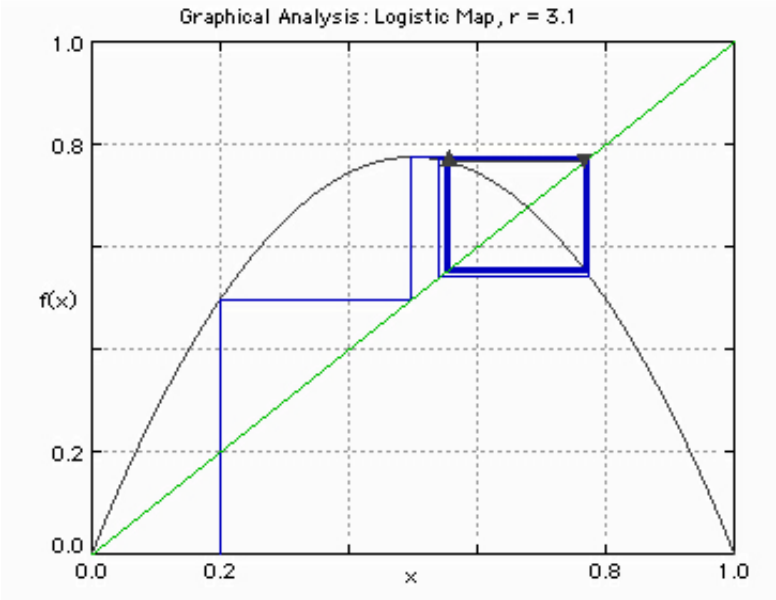
If $r \in (0, 1]$, it can be seen from the above diagram (obtained via Brian Davies' software) that $x_n \to 0$, independently of the initial value x_0. This is proved in Theorem 2.7.1 of Reed.

The proof is easy; it is first a matter of showing that the sequence (x_n) is decreasing, as is indicated by the diagram in Reed, and hence has a limit by Theorem 5.1.3. If x^* is the limit, then by passing to the limit in (5) it follows that $x^* = rx^*(1-x^*)$, and so $x^* = 0$. (The other solution of $x = rx(1-x)$ is $x = 1-1/r$, which does not lie in $[0, 1]$ — unless $a = 1$, which gives 0 again.)

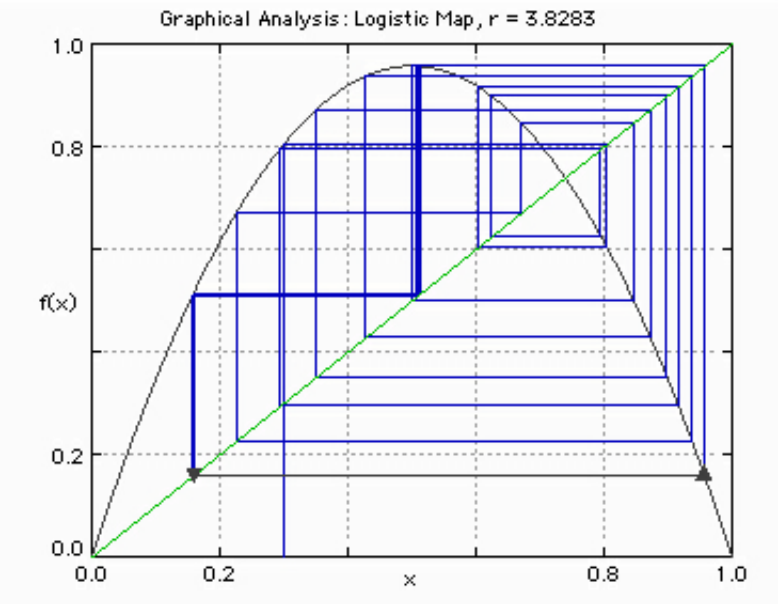
For $1 < r \leq 3$ and $0 < x_0 < 1$, the sequence converges to $x^* = 1 - 1/r$. (This is proved in Reed, Theorem 2.7.2 for $1 < r \leq 2\sqrt{2}$. The proof is not difficult, but we will not go through it.) If $x_0 = 0, 1$ then it is easy to see that the sequence converges to 0.



For $3 < r \leq 4$ and $0 < x_0 < 1$ the situation becomes more complicated. If $3 < r \leq 3.4$ (approx.) then x_n will eventually oscillate back and forth between two values which become closer and closer to two of the solutions of $x = F(F(x)) = F^2(x)$ (see the numerics on Reed page 65).



As r gets larger the situation becomes more and more complicated, and in particular one needs to consider solutions of $x = F^k(x)$ for large k.


Graphical Analysis: Logistic Map, r = 3.8283
1.0
0.8
f(x)
0.2
0.0
0.0
0.2
x
0.8
1.0

7. Continuity

Study Reed Sections 3.1, 3.2. See also the MATH1115 Notes, Chapter 4.

7.1. Definition of continuity. We will consider functions $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$ is the domain of f, also denoted by $\mathcal{D}(f)$. Usually A will be $\mathbb{R}$, an interval, or a finite union of intervals. (To fix your ideas consider the case $A = [a, b]$ or $A = (a, b)$.)

DEFINITION 7.1.1. A function f is *continuous at* $a \in \mathcal{D}(f)$ if

$$x_n \to a \quad \Longrightarrow \quad f(x_n) \to f(a)$$

whenever $(x_n) \subseteq \mathcal{D}(f)$.[22] The function f is *continuous* if it is continuous at every point in its domain.

This definition in terms of sequences is quite natural: "as x_n gets closer and closer to a, $f(x_n)$ gets closer and closer to $f(a)$". By the following theorem it is equivalent to the usual "ε–δ" definition which does not involve sequences.

THEOREM 7.1.2. *A function f is* continuous *at a point $a \in \mathcal{D}(f)$ iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that:*

$$x \in \mathcal{D}(f) \text{ and } |x - a| \leq \delta \Rightarrow |f(x) - f(a)| \leq \epsilon.$$

(For the proof see Reed Theorem 3.1.3 page 77.)

Note that if $f(x) = \sin \frac{1}{x}$ then f *is* continuous on its domain $(-\infty, 0) \cup (0, \infty)$. However, there is *no* continuous extension of f to all of $\mathbb{R}$. On the other hand, $f(x) = x \sin \frac{1}{x}$ is both continuous on its domain $(-\infty, 0) \cup (0, \infty)$ and also has a continuous extension to all of $\mathbb{R}$. (Draw diagrams!)

Another interesting example is the function g given by

$$g(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ -x & \text{if } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Then g is continuous at 0 and nowhere else. (Draw a diagram, and convince yourself via the definition.)

7.2. Limits and continuity. Note the definition of

$$\lim_{x \to a} f(x) = L,$$

on the second half of page 78 of Reed, both in terms of sequences and in terms of ε and δ.[23]

It follows (provided a is not an *isolated point*[24] of $\mathcal{D}(f)$) that f is continuous at $a \in \mathcal{D}(f)$ iff

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

(From our definitions, if a is an isolated point of $\mathcal{D}(f)$ then f is continuous at a although $\lim_{x \to a} f(x)$ is not actually defined. This is not an interesting sitation, and we would not normally consider continuity at isolated points.)

---

[22] By $(x_n) \subseteq \mathcal{D}(f)$ we mean that each term of the sequence is a member of $\mathcal{D}(f)$.

[23] In the definition of $\lim_{x \to a} f(x) = L$, we restrict to sequences $x_n \to a$ with $x_n \neq a$, or we require that $0 < |x - a| \leq c$ (thus $x \neq a$), depending on the definition used — see Reed page 78.

In the case of continuity however, because the required limit is $f(a)$ (c.f. (21) and (22)), one can see it is not necessary to make this restriction.

[24] We say $a \in A$ is an *isolated point* of A if there is no sequence $x_n \to a$ such that $(x_n) \subseteq A \setminus \{a\}$. For example, if $A = [0, 1] \cup 2$ then 2 is an isolated point of A (and is the only isolated point of A).

7.3. Properties of continuity.

THEOREM 7.3.1. *Suppose f and g are continuous. Then the following are all continuous on their domains.*

1. $f + g$
2. cf *for any* $c \in \mathbb{R}$
3. fg
4. f/g
5. $f \circ g$

The domain of each of the above functions is just the set of real numbers where it is defined. Thus the domain of $f + g$ and of fg is $\mathcal{D}(f) \cap \mathcal{D}(g)$; the domain of cf is $\mathcal{D}(f)$; the domain of f/g is $\{\, x \in \mathcal{D}(f) \cap \mathcal{D}(g) \mid g(x) \neq 0 \,\}$, and the domain of $f \circ g$ is $\{\, x \mid x \in \mathcal{D}(g) \text{ and } g(x) \in \mathcal{D}(f) \,\}$.

For the proofs of Theorem 7.3.1, see Reed Theorems 3.1.1 and 3.1.2 or the MATH1115 notes Section 4. The proofs in terms of the sequence definition of continuity are very easy, since the hard work has already been done in proving the corresponding properties for limits of sequences.

See Examples 1–4 in Section 3.1 of Reed.

7.4. Deeper properties of continuous functions. These require the completeness property of the real numbers for their proofs.

THEOREM 7.4.1. *Every continuous function defined on a closed bounded interval is bounded above and below and moreover has a maximum and a minimum value.*

PROOF. A proof using the Bolzano-Weierstraß theorem is in Reed (Theorem 3.2.1 and 3.2.2 on pp 80,81).

The proof is fairly easy, since the hard work has already been done done in proving the Bolzano-Weierstraß theorem. □

Noe that a similar result is not true on other intervals. For example, consider $f(x) = 1/x$ on $(0, 1]$.

THEOREM 7.4.2 (Intermediate Value Theorem). *Every continuous function defined on a closed bounded interval takes all values between the values taken at the endpoints.*

PROOF. See Reid p82. Again, the hard work has alredy been done in proving the Bolzano-Weierstraß theorem. □

COROLLARY 7.4.3. *Let f be a continuous function defined on a closed bounded interval with minimum value m and maximum value M. Then the range of f is the interval $[m, M]$.*

PROOF. (See Reid p83.) The function f must take the values m and M at points c and d (say) by Theorem 7.4.1. By the Intermediate Value Theorem applied with the endpoints c and d, f must take all values between m and M. This proves the theorem. □

7.5. Uniform continuity. If we consider the ε–δ definition of continuity of f at a point a as given in Theorem 7.1.2, we see that for any given ε the required δ will normally depend on a. The steeper the graph of f near a, the smaller the δ that is required. If for each $\varepsilon > 0$ there is some $\delta > 0$ which will work for *every* $a \in \mathcal{D}(f)$, then we say f is uniformly continuous.

More precisely:

DEFINITION 7.5.1. A function f is *uniformly continuous* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every $x_1, x_2 \in \mathcal{D}(f)$:

$$|x_1 - x_2| \leq \delta \quad \text{implies} \quad |f(x_1) - f(x_2)| \leq \varepsilon.$$

Any uniformly continuous function is certainly continuous. On the other hand, the function $1/x$ with domain $(0, 1]$, and the function x^2 with domain $\mathbb{R}$, are continuous but not uniformly continuous. (See Reed Example 2 page 84.)

However, we have the following important result for closed bounded intervals.

THEOREM 7.5.2. *Any continuous function defined on a closed bounded interval is uniformly continuous.*

PROOF. See Reid p 85. I will discuss this in class. □

There is an important case in which uniform continuity is easy to prove.

DEFINITION 7.5.3. A function f is *Lipschitz* if there exists an M such that

$$|f(x_1) - f(x_2)| \leq M|x_1 - x_2|$$

for all x_1, x_2 in the domain of f. We say M is a Lipshitz constant for f.

This is equivalent to claiming that

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq M$$

for all $x_1 \neq x_2$ in the domain of f. In other words, the slope of the line joining any two points on the graph of f is $\leq M$.

PROPOSITION 7.5.4. *If f is Lipschitz, then it is uniformly continuous.*

PROOF. Let $\epsilon > 0$ be any positive number.
Since $|f(x) - f(a)| \leq M|x - a|$ for any $x, a \in \mathcal{D}f$, it follows that

$$|x - a| \leq \frac{\epsilon}{M} \Rightarrow |f(x) - f(a)| \leq \epsilon.$$

This proves uniform continuity. □

It follows from the Mean Value Theorem that if the domain of f is an *interval* I (not necessarily closed or bounded), f is differentiable and $|f'(x)| \leq M$ on I, then f is Lipschitz with Lipschitz constant M.

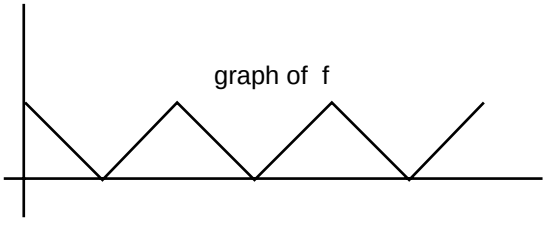
PROOF. Suppose $x_1, x_2 \in I$ and $x_1 < x_2$. Then

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} = f'(c)$$

for some $x_1 < c < x_2$ by the Mean Value Theorem. Since $|f'(c)| \leq M$, we are done.
Similarly if $x_2 < x_1$. □

♣ Where did we use the fact that the domain of f is an interval? Give a counter example in case the domain is not an interval.

There are examples of non-differentianble Lipschitz functions, such as the following "saw-tooth" function.



8. Riemann Integration

Study Reed Chapter 3.3. Also see the MATH1116 notes.

8.1. Definition. The idea is that $\int_a^b f$ (which we also write as $\int_a^b f(x)\,dx$) should be a limit of lower and upper sums corresponding to rectangles, respectively below and above the graph of f. *See Reed page 87 for a diagram.*

Suppose f is defined on the interval $[a,b]$ and f is *bounded*. (At first you may think of f as being continuous, but unless stated otherwise it is only necessary that f be bounded.)

A *partition* P of $[a,b]$ is any finite sequence of points $(x_0,\dots,x_N)$ such that

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

Let

$$m_i = \inf\{\, f(x) \mid x \in [x_{i-1},x_i]\,\}, \quad M_i = \sup\{\, f(x) \mid x \in [x_{i-1},x_i]\,\}, \tag{6}$$

for $i = 1,\dots,N$.[25] Then the *lower* and *upper sum for the partition* P are defined by

$$L_P(f) = \sum_{i=1}^{N} m_i(x_i - x_{i-1}), \quad U_P(f) = \sum_{i=1}^{N} M_i(x_i - x_{i-1}). \tag{7}$$

They correspond to the sum of the areas of the lower rectangles and the upper rectangles respectively.

Adding points to a partition increases lower sums and decreases upper sums. More precisely, if Q is a *refinement* of P, i.e. is a partition which includes all the points in P, then

$$L_P(f) \le L_Q(f), \quad U_Q(f) \le U_P(f). \tag{8}$$

(See Reed Lemma 1 page 88 for the proof. Or see Lemma 7.4 page 72 of the 1998 AA1H notes — although the proof there is for continuous functions and so uses max and min in the definition of upper and lower sums, the proof is essentially the same for general bounded functions provide one uses instead sup and inf.)

Any lower sum is $\le$ any upper sum:

$$L_P(f) \le U_Q(f). \tag{9}$$

Here, P and Q are *arbitrary* partitions. The proof is by taking a common refinement of P and Q and using (8). (See Reed Lemma 2 page 89, and the MATH1116 notes.)

For each partition P we have a number $L_P(f)$, and the set of all such numbers is bounded above (by any upper sum). The supremum of this set of numbers is denoted by

$$\sup_P\{L_P(f)\} = \sup\{\, L_P(f) \mid P\text{is a partition of [a,b]}\,\}.$$

Similarly, the infimum of the set of all upper sums,

$$\inf_P\{U_P(f)\} = \inf\{\, U_P(f) \mid P\text{is a partition of [a,b]}\,\},$$

exists. If these two numbers are equal, then we say f *is (Riemann)*[26] *integrable on* $[a,b]$ and define

$$\int_a^b f := \inf_P\{U_P(f)\} = \sup_P\{L_P(f)\}. \tag{10}$$

[25] If f is continuous, we can replace sup and inf by max and min respectively.

[26] There is a much more powerful notion called *Lebesgue integrable*. This is more difficult to develop, and will be treated in the first Analysis course in third year.

(For an example of a function which is bounded but not Riemann integrable, see Reed Problem 1 page 93.)

8.2. Basic results.

THEOREM 8.2.1. *If f is a continous function on a closed bounded interval $[a,b]$, then f is integrable on $[a,b]$.*

PROOF. (See Reed Lemma 3 and Theorem 3.3.1 on pages 89, 90, or MATH1116 notes.)

The idea is first to show, essentially from the definitions of sup and inf, that it is sufficient to prove for each $\varepsilon > 0$ that there exists a partition P for which

$$U_P(f) - L_P(f) < \varepsilon. \tag{11}$$

The existence of such a P follows from the uniform continuity of f. □

In order to numerically estimate $\int_a^b f$ we can choose a point x_i^* in each interval $[x_{i-1}, x_i]$ and compute the corresponding *Riemann sum*

$$R_P(f) := \sum_{i=1}^{N} f(x_i^*)(x_i - x_{i-1}).$$

This is not a precise notation, since $R_P(f)$ depends on the points x_i^* as well as P. It is clear that

$$L_P(f) \le R_P(f) \le U_P(f).$$

The following theorem justifies using Riemann sums. Let $\|P\|$ denote the maximum length of the intervals in the partition P. Then

THEOREM 8.2.2. *Suppose f is Riemann integrable on the interval $[a,b]$. Let P_k be a sequence of partitions of $[a,b]$ such that $\lim_{k\to\infty} \|P_k\| = 0$ and suppose $R_{P_k}(f)$ is a corresponding sequence of Riemann sums. Then*

$$R_{P_k}(f) \to \int_a^b f \quad as \quad k \to \infty.$$

See Reed page 91 for the proof.

8.3. Properties of the Riemann integral. For proofs of the following theorems see Reed pages 91–93.

THEOREM 8.3.1. *Suppose f and g are continuous on $[a,b]$ and c and d are constants. Then*

$$\int_a^b (cf + dg) = c\int_a^b f + d\int_a^b g.$$

(See Reed Theorem 3.3.3 page 92 for te proof.)

THEOREM 8.3.2. *Suppose f and g are continuous on $[a,b]$ and $f \le g$. Then*

$$\int_a^b f \le \int_a^b g.$$

THEOREM 8.3.3. *Suppose f is continuous on $[a,b]$. Then*

$$\Big|\int_a^b f\Big| \le \int_a^b |f|.$$

COROLLARY 8.3.4. *If f is continuous on $[a,b]$ and $|f| \le M$ then*

$$\Big|\int_a^b f\Big| \le M(b-a).$$

THEOREM 8.3.5. *Suppose f is continuous on $[a,b]$ and $a \le c \le b$. Then*

$$\int_a^c f + \int_c^b f = \int_a^b f.$$

REMARK 8.3.1. Results analogous to those above are still true if we assume the relevant functions are integrable, but not necessarily continuous. The proofs are then a little different, but not that much more difficult. See "Calculus" by M. Spivak, Chapter 13.

The fundamental connection between integration and differentiation is made in Reed in Section 4.2; I return to this after we discuss differentiation. In particular, *if f is continuous on $[a,b]$, then*

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

where $F'(x) = f(x)$ for all $x \in [a,b]$. This often gives a convenient way of finding integrals.

8.4. Riemann integration for discontinuous functions. See Reed Section 3.5. This material will be done "lightly".

- *Example 1 of Reed*: If f is the function defined on $[0,1]$ by

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \text{ and } x \text{ irrational} \\ 0 & 0 \le x \le 1 \text{ and } x \text{ rational} \end{cases}$$

then f is not Riemann integrable, since every upper sum is 1 and every lower sum is 0.

- However, *any bounded monotone increasing*[27] (*or decreasing*) *function on an interval $[a,b]$ is Riemann integrable.* See Reed Theorem 3.5.1.

A monotone function can have an infinite number of points where it is discontinuous (Reed Q8 p112). Also, the sum of two Riemann integrable functions is Riemann integrable (Reed Q13 p112). This shows that many quite "bad" (but still bounded) functions can be Riemann integrable.

- The next result (Theorems 3.5.2, 3) is that *if a bounded function is continuous except at a finite number of points on $[a,b]$, then it is Riemann integrable on $[a,b]$.* (Interesting examples are $f(x) = \sin 1/x$ if $-1 \le x \le 1$ & $x \ne 0$, $f(0) = 0$; and $f(x) = \sin 1/x$ if $0 < x \le 1$, $f(0) = 0$; c.f. Reed Example 2.)

Moreover, if the points of discontinuity are $a_1 < a_2 < \cdots < a_k$ and we set $a_0 = a$, $a_{k+1} = b$, then

$$\int_a^b f = \sum_{j=1}^{j=k+1} \int_{a_{j-1}}^{a_j} f,$$

and

$$\int_{a_{j-1}}^{a_j} f = \lim_{\delta \to 0+} \int_{a_{j-1}+\delta}^{a_j - \delta} f.$$

Since f is continuous on each interval $[a_{j-1}+\delta, a_j - \delta]$, we can often find $\int_{a_{j-1}+\delta}^{a_j-\delta} f$ by standard methods for computing integrals (i.e. find a function whose derivative is f) or by numerical methods, and then take the limit. (By the *one-sided limit* $\lim_{\delta \to 0+}$ we mean that δ is restricted to be positive; Reed writes $\lim_{\delta \searrow 0}$ to mean the same thing. See Reed p 108 for the definition of a one-sided limit in terms of

[27] f is monotone increasing if $f(x_1) \le f(x_2)$ whenever $x_1 < x_2$.

sequences. The "ϵ–δ" definition is in Definition 3.13 p 29 of the AA1H 1998 Notes. The two definitions are equivalent by the same argument as used in the proof of similar equivalences for continuity and also for ordinary limits, see Reid p 78.$\Big)$

• A particular case of importance is when the left and right limits of f exist at the discontinuity points a_j. The function f then is said to be *piecewise continuous*. In this case if the function f_j is defined on $[a_{j-1}, a_j]$ by

$$f_j(x) = \begin{cases} f(x) & a_{j-1} < x < a_j \\ \lim_{x \to a_{j-1}^+} f(x) & x = a_{j-1} \\ \lim_{x \to a_j^-} f(x) & x = a_j, \end{cases}$$

then f_j is continuous on $[a_{j-1}, a_j]$,

$$\int_a^b f = \sum_{j=1}^{j=k+1} \int_{a_{j-1}}^{a_j} f_j,$$

and each $\int_{a_{j-1}}^{a_j} f_j$ can often be computed by standard methods. See Reed Example 3 p 108 and Corollary 3.5.4 p 109.

8.5. Improper Riemann integrals. See Reed Section 3.8. This material will be done "lightly".

In previous situations we had a bounded function with a finite number of discontinuities. The Riemann integral existed according to our definition, and we saw that it could be obtained by computing the Riemann integrals of certain continuous functions and by perhaps taking limits.

If the function is not bounded, or the domain of integration is an infinite interval, the Riemann integral can often be *defined* by taking limits. The Riemann integral is then called an *improper integral*. For details and examples of the following, see Reed Section 3.6.

For example, if f is continuous on $[a, b)$ then we define

$$\int_a^b f = \lim_{\delta \to 0+} \int_a^{b-\delta} f,$$

provided the limit exists.

If f is continuous on $[a, \infty)$ then we define

$$\int_a^\infty f = \lim_{b \to \infty} \int_a^b f,$$

provided the limit exists.

$\Big($ I do not think that Reed actually defines what is meant by

$$\lim_{x \to \infty} g(x),$$

where here $g(x) = \int_a^x f$. But it is clear what the definition in terms of sequences should be. Namely,

$$\lim_{x\to\infty} g(x) = L \text{ if whenever a sequence } x_n \to \infty \text{ then } g(x_n) \to L.$$

The definition of $x_n \to \infty$ (i.e. (x_n) "diverges" to infinity), is the natural one and is given on p 32 of Reed.

There is also a natural, and equivalent, definition which does not involve sequences. Namely,

$$\lim_{x\to\infty} g(x) = L \text{ if for every } \epsilon > 0 \text{ there is a real number } K \text{ such that } x \geq K \text{ implies } |g(x) - L| \leq \epsilon. \Big)$$

Similar definitions apply for integrals over $(a, b]$, (a, b), $\mathbb{R}$, etc.

Note the subtlety in Reed Example 5 p 116. The integral $\int_1^\infty \frac{\sin x}{x}\, dx$ exists in the limit sense defined above, because of "cancellation" of successive positive and negative bumps, but the "area" above the axis and the "area" below the axis is infinite. This is analogous to the fact that the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \ldots$ converges, although the series of positive and negative terms each diverge.

9. Differentiation

9.1. Material from last year. Read Sections 4.1, 4.2 of Reed and MATH1116 Notes.

Suppose $f : S \to \mathbb{R}$, $x \in S$, and there is some open interval containing x which is a subset of S. (Usually S will itself be an interval, but not necessarily open or bounded).

Then $f'(x)$, the *derivative of f at x* is defined in the usual way as

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}.$$

If $S = [a, b]$, it is also convenient to define the derivatives at the endpoints a and b of S in the natural way by restricting h to be > 0 or < 0 respectively. See Reed p 126, three paragraphs from the bottom, and p 108 (or MATH1116 notes).

In Reed Section 4.1 note in particular Theorem 4.1 — *differentiability at any point implies continuity there*, Theorem 4.2 (the usual rules for differentiating sums, producs and quotients), and Theorem 4.3 (the chain rule; you may prefer the proof given in Adams).

The next main result is the justification for using derivatives to find maximum (and similarly minimum) points; *if a continuous function takes a maximum at c, and $f'(c)$ exists, then $f'(c) = 0$.*

9.2. Fundamental theorem of calculus. This connects differentiation and integration, see Reed Theorems 4.2.4 and 4.2.5. (The proof in Reed of Theorem 4.2.4 is different from that in the MATH116 notes.)

9.3. Mean value theorem and Taylor's theorem. See Reed Section 4.3 and Adams pages 285–290, 584, 585.

The Mean Value Theorem (Reed Theorem 4.2.3) says geometrically that "the slope of the chord joining two points on the graph of a differentiable function equals the derivative at some point between them".

Taylor's Theorem, particularly with the Legendre form of the remainder as in Reed page 136, Theorem 4.3.1, is a generalisation.

Note also how L'Hôpital's Rule, Theorem 4.3.2 page 138, follows from Taylor's theorem.

9.4. Newton's method. See Reed Section 4.4 and Adams pages 192–195.

Newton's method, and its generalisation to finding zeros of functions $f : \mathbb{R}^n \to \mathbb{R}^n$ for $n > 1$ (see Adams pages 745–748), is very important for both computational and theoretical reasons.

As we see in Reed page 141, if x_n is the nth approximatiion to the solution $\overline{x}$ of $f(\overline{x}) = 0$, then

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Remarks on proof of Theorem 4.4.1. We want to estimate $|x_{n+1} - \overline{x}|$ in terms of $|x_n - \overline{x}|$.

Applying the above formula, and both Taylor's Theorem and the Mean Value Theorem, gives

$$|x_{n+1} - \overline{x}| \le \frac{1}{f'(x_n)} \left(f''(\overline{\tau}_n) + \frac{f''(\tau_n)}{2!} \right) |x_n - \overline{x}|^2$$

for some $\overline{\tau}_n$ and τ_n between x_n and $\overline{x}$, see Reed page 143 formula (14).

Assume $f'(\overline{x}) \neq 0$ and f is C^2 in some interval containing $\overline{x}$.[28] Then *provided* $|x_n - \overline{x}| < \varepsilon$ (say), it follows that

$$\frac{1}{f'(x_n)}\left(f''(\overline{\tau}_n) + \frac{f''(\tau_n)}{2!}\right) \leq C \tag{12}$$

for some constant C (depending on f). Hence

$$|x_{n+1} - \overline{x}| \leq C\,|x_n - \overline{x}|^2. \tag{13}$$

We say that (x_n) converges to $\overline{x}$ *quadratically fast*. See the discusion at the end of the first paragraph on page 144 of Reed.

The only remaining point in the proof of Theorem 4.4.1 in Reed, pages 142,143, is to show that if $|x_1 - \overline{x}| < \varepsilon$ (and so (13) is true for $n = 1$) then $|x_n - \overline{x}| < \varepsilon$ for all n (and so (13) is true for all n). But from (13) with $n = 1$ we have

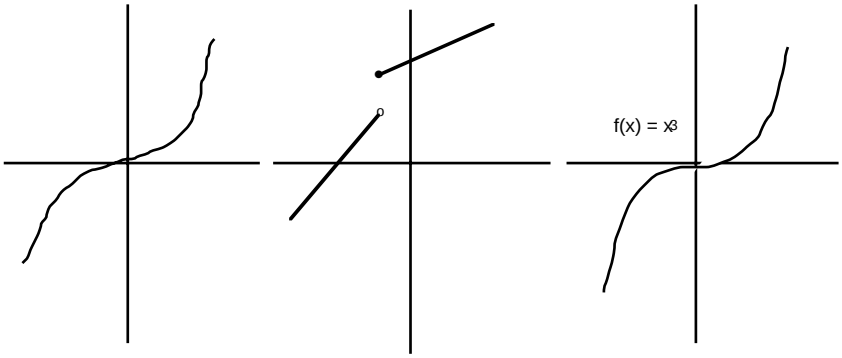
$$|x_2 - \overline{x}| \leq \frac{1}{2}|x_1 - \overline{x}|$$

provided $|x_1 - \overline{x}| \leq \frac{1}{2C}$. Thus x_2 is even closer to $\overline{x}$ than x_1 and so (13) is true with $n = 2$, etc. □

Note that even though the proof only explicitly states *linear* convergence in (16), in fact it also gives the much better *quadratic* quadratic convergence as noted on page 144 line 4.

9.5. Monotonic functions.
See Reed Section 4.5, Adams Section 4.1 and pages 258–260.

A function $f : I \to \mathbb{R}$, is *strictly increasing* if $x < y$ implies $f(x) < f(y)$. Unless stated otherwise, we alsways take the domain I to be an *interval*. Similarly one defines *strictly decreasing*. A function is *strictly monotonic* if it is either strictly increasing or strictly decreasing.



If f is differentiable on I and $f' > 0$, then it is strictly increasing (Reed page 150, problem 1. HINT: Use the Mean Value Theorem). But the derivative of a strictly increasing function *may* somewhere be zero, e.g. $f(x) = x^3$.

A strictly monotonic function f is one-one and so has an inverse f^{-1}, defined by

$$f^{-1}(y) = \text{ the unique } x \text{ such that } f(x) = y.$$

The domain of f^{-1} is the range of f and the range of f^{-1} is the domain of f.

If $f : I \to \mathbb{R}$ is strictly monotonic and continuous, then its range is an interval and its inverse exists and is strictly monotonic and continuous. See Reed page 83, Corollary 3.2.4 and page 147 Theorem 4.5.1.

[28] That is, f' and f'' exist and are continuous.

If $f : I \to \mathbb{R}$ is strictly monotonic and continuous, and is differentiable at a with $f'(a) \neq 0$, then f^{-1} is differentiable at $b = f(a)$ and[29]

$$(f^{-1})'(b) = \frac{1}{f'(a)}$$

If $f : I \to \mathbb{R}$ is strictly monotonic and continuous, and everywhere differentiable with $f'(x) \neq 0$, then by the above result we have that f^{-1} is everywhere differentiable. Moreover, the derivative is *continuous*, since

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$$

by the previous formula, and both f^{-1} and f are continuous. (This is Reed Corollary 4.5.3)

The change of variables formula, Theorem 4.5.4, does not really require ϕ to be strictly increasing, as noted in Reed page 151, problem 6 (See Adams page 326, Theorem 6, for the simpler proof of this more general fact, using the Fundamental Theorem of Calculus). The proof in Reed does have the advantage that it generalises to situations where the Fundamental Theoorem cannot be applied.

The result of Theorem 4.5.5 also follows from the Fundamental Theorem of Calculus, see Reed page 151, problem 7.

[29]Loosely speaking, $\frac{dy}{dx} = 1/\frac{dx}{dy}$. But this notation can cause problems and be ambiguous.

10. Basic metric and topological notions in $\mathbb{R}^n$

See Adams page 643 (third edition), or (better) fourth edition pages 601,602. See also Reed problem 10 page 40 and the first two paragraphs in Section 4.6. But we do considerably more.

10.1. Euclidean n-space. We define

$$\mathbb{R}^n = \{\, (x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in \mathbb{R} \,\}.$$

Thus $\mathbb{R}^n$ denotes the set of all (ordered) n-tuples $(x_1, \ldots, x_n)$ of real numbers.

We think of $\mathbb{R}^1$, $\mathbb{R}^2$ and $\mathbb{R}^3$ as the real line, the plane and 3-space respectively. But for $n \geq 4$ we cannot think of $\mathbb{R}^n$ as consisting of points in physical space. Instead, we usually think of points in $\mathbb{R}^n$ as just being given algebraically by n-tuples $(x_1, \ldots, x_n)$ of real numbers. (We sometimes say $(x_1, \ldots, x_n)$ is a *vector*, but then you should think of a vector as a point, rather than as an "arrow".)

In $\mathbb{R}^2$ and $\mathbb{R}^3$ we usually denote the coordinates of a point by (x, y) and (x, y, z), but in higer dimensions we usually use the notation $(x_1, \ldots, x_n)$, etc.

Although we usually think of points in $\mathbb{R}^n$ as n-tuples, rather than physical points, we are still motivated and guided by the geometry in the two and three dimensnional cases. So we still speak of "points" in $\mathbb{R}^n$

We define the *(Euclidean) distance*[30] between $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ by

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

This agrees with the usual distance in case $n = 1, 2, 3$.

We also call the set of points $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfying any equation of the form

$$a_1 x_1 + \cdots + a_n x_n = b,$$

a *hyperplane*. Sometimes we restrict this definition to the case $b = 0$ (in which case the hyperplane "passes through" the origin).

10.2. The Triangle and Cauchy-Schwarz inequalities. The Euclidean distance function satisfies three important properties. Any function d on a set S (not necessarily $\mathbb{R}^n$) which satisfies these properties is called a *metric*, and S is called a *metric space* with metric d. We will discuss general metrics later. But you should observe that, unless noted or otherwise clear from the context, the proofs and definitions in Sections 10–14 carry over directly to general metric spaces. We discuss this in more detail in Section 15.

THEOREM 10.2.1. *Let d denote the Euclidean distance function in $\mathbb{R}^n$. Then*

1. $d(\mathbf{x}, \mathbf{y}) \geq 0$*;* $d(\mathbf{x}, \mathbf{y}) = 0$ *iff* $\mathbf{x} = \mathbf{y}$. *(positivity)*
2. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$. *(symmetry)*
3. $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$. *(triangle inequality)*

PROOF. The first two properties are immediate.

The third follows from the Cauchy-Schwarz inequality, which we prove in the next theorem.

To see this let

$$u_i = x_i - z_i, \quad v_i = z_i - y_i,$$

Then

$$d(\mathbf{x}, \mathbf{y})^2 = \sum_i (x_i - y_i)^2 = \sum_i (x_i - z_i + z_i - y_i)^2$$

[30] Later we will define other distance functions (i.e. metrics, see the next section) on $\mathbb{R}^n$, but the Euclidean distance function is the most important.

$$= \sum_i (u_i + v_i)^2 = \sum_i u_i^2 + 2u_i v_i + v_i^2$$
$$= \sum_i u_i^2 + 2\sum_i u_i v_i + \sum_i v_i^2$$
$$\leq \sum_i u_i^2 + 2\Big(\sum_i u_i^2\Big)^{1/2}\Big(\sum_i u_i^2\Big)^{1/2} + \sum_i v_i^2$$

(by the Cauchy-Schwarz inequality)

$$= \left(\left(\sum_i u_i^2\right)^{\frac{1}{2}} + \left(\sum_i v_i^2\right)^{\frac{1}{2}}\right)^2 = \Big(d(\mathbf{x},\mathbf{z}) + d(\mathbf{z},\mathbf{y})\Big)^2.$$

The triangle inequality now follows. □

The following is a fundamental inequality.[31] Here is one of a number of possible proofs.

THEOREM 10.2.2 (Cauchy-Schwarz inequality). *Suppose* $(u_1, \ldots, u_n) \in \mathbb{R}^n$ *and* $(v_1, \ldots, v_n) \in \mathbb{R}^n$. *Then*

$$\left|\sum_i u_i v_i\right| \leq \Big(\sum_i u_i^2\Big)^{1/2}\Big(\sum_i v_i^2\Big)^{1/2}. \tag{14}$$

Moreover, equality holds iff at least one of $(u_1, \ldots, u_n)$, $(v_1, \ldots, v_n)$ *is a multiple of the other (in particular if one is* $(0, \ldots, 0)$*).*

PROOF. Let

$$f(t) = \sum_i (u_i + t v_i)^2 = \sum_i u_i^2 + 2t\sum_i u_i v_i + t^2 \sum_i v_i^2.$$

Then f is a quadratic in t (provided $(v_1, \ldots, v_n) \neq (0, \ldots, 0)$), and from the first expression for $f(t)$, $f(t) \geq 0$ for all t.

It follows that $f(t) = 0$ has no real roots (in case $f(t) > 0$ for all t) or two equal real roots (in case $f(t) = 0$ for some t). Hence

$$4\Big(\sum_i u_i v_i\Big)^2 - 4\sum_i u_i^2 \sum_i v_i^2 \leq 0. \tag{15}$$

This immediately gives the Cauchy-Schwarz inequality.

If one of $(u_1, \ldots, u_n)$, $(v_1, \ldots, v_n)$ is $(0, \ldots, 0)$ or is a multiple of the other then equality holds in (14). (*why?*)

Conversely, if equality holds in (14), i.e. in (15), and $(v_1, \ldots, v_n) \neq (0, \ldots, 0)$, then $f(t)$ is a quadratic (since the coefficient of t^2 is non-zero) with equal roots and in particular $f(t) = 0$ for some t. This t gives $\mathbf{u} = -t\mathbf{v}$, from the first expression for $f(t)$.

This completes the proof of the theorem. □

REMARK 10.2.1. In the preceding we discussed the (standard) metric on $\mathbb{R}^n$. It can be defined from the (standard) norm and conversely. That is

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \quad \text{and} \quad \|\mathbf{x}\| = d(\mathbf{0},\mathbf{x})$$

The triangle inequlity for the metric follows easily from the triangle inequality for the norm (i.e. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$), and conversely (*Exercise*).

[31] The Cauchy-Schwarz inequality can be interpreted in terms of inner products as saying that $|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$. But we will not need to refer to inner products here.

10.3. Preliminary Remarks on Metric Spaces. Any function d on a set $\mathcal{M}$ (not necessarily $\mathbb{R}^n$) which satisfies the three properties in Theorem 10.2.1 is called a *metric*, and $\mathcal{M}$ together with d is called a *metric space*.

You should now study Section 15.1, except possibly for Example 15.1.3.

We will discuss general metrics later. But you should observe that unless noted or otherwise clear from context, the proofs and definitions in Sections 10–14 carry over directly to general metric spaces. We discuss this in more detail in Section 15.

10.4. More on intersections and unions. The intersection and union of two sets was defined in Reed page 7. The intersection and union of more than two sets is defined similarly. The intersection and union of a (possibly infinite) collection of sets indexed by the members of some set J are defined by

$$\bigcap_{\lambda\in J} A_\lambda = \{\, x \mid x \in A_\lambda \text{ for } \textit{all } \lambda \in J \,\},$$
$$\bigcup_{\lambda\in J} A_\lambda = \{\, x \mid x \in A_\lambda \text{ for } \textit{some } \lambda \in J \,\}.$$

For example:

$$\bigcap_{n\in\mathbb{N}} \Big[0, \frac{1}{n}\Big] = \{0\},$$
$$\bigcap_{n\in\mathbb{N}} \Big(0, \frac{1}{n}\Big) = \emptyset,$$
$$\bigcap_{n\in\mathbb{N}} \Big(-\frac{1}{n}, \frac{1}{n}\Big) = \{0\},$$
$$\bigcap_{n\in\mathbb{N}} [n, \infty) = \emptyset,$$
$$\bigcup_{r\in\mathbb{Q}} (r-\epsilon, r+\epsilon) = \mathbb{R} \quad \text{(for any fixed } \epsilon > 0),$$
$$\bigcup_{r\in\mathbb{R}} \{r\} = \mathbb{R}.$$

It is an easy generalisation of De Morgan's law that

$$(A_1 \cup \cdots \cup A_k)^c = A_1^c \cap \cdots \cap A_k^c.$$

More generally,

$$\Big(\bigcup_{\lambda\in J} A_\lambda\Big)^c = \bigcap_{\lambda\in J} A_\lambda^c.$$

Also,

$$(A_1 \cap \cdots \cap A_k)^c = A_1^c \cup \cdots \cup A_k^c$$

and

$$\Big(\bigcap_{\lambda\in J} A_\lambda\Big)^c = \bigcup_{\lambda\in J} A_\lambda^c.$$

10.5. Open sets. The definitions in this and the following sections generalise the notions of open and closed intervals and endpoints on the real line to $\mathbb{R}^n$. You should think of the cases $n = 2, 3$.

A *neighbourhood* of a point $\mathbf{a} \in \mathbb{R}^n$ is a set of the form

$$B_r(\mathbf{a}) = \{\, \mathbf{x} \in \mathbb{R}^n \mid d(\mathbf{x}, \mathbf{a}) < r \,\}$$

for some $r > 0$. We also say that $B_r(\mathbf{a})$ is the *open ball* of radius r centred at $\mathbf{a}$. In case $n = 1$ we say $B_r(\mathbf{a})$ is an *open interval* centred at $\mathbf{a}$, and in case $n = 2$ also call it the *open disc* of radius r centred at $\mathbf{a}$.

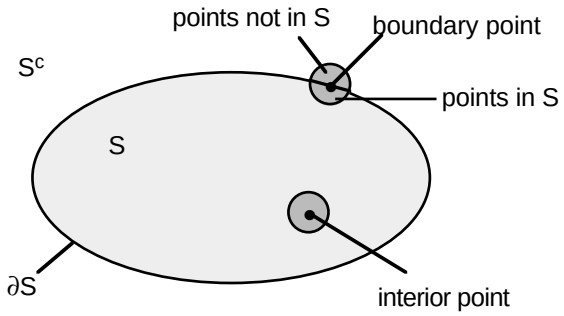
DEFINITION 10.5.1. A set $S \subseteq \mathbb{R}^n$ is *open* (in $\mathbb{R}^n$)[32] if for every $\mathbf{a} \in S$ there is a neighbourhood of $\mathbf{a}$ which is a subset of S.

Every neighbourhood is itself open (*why?*). Other examples in $\mathbb{R}^2$ are

1. $\{\, \mathbf{x} \in \mathbb{R}^2 \mid d(\mathbf{x}, \mathbf{a}) > r \,\}$,
2. $\{\, (x, y) \mid x > 0 \,\}$,
3. $\{\, (x, y) \mid y > x^2 \,\}$,
4. $\{\, (x, y) \mid y \neq x^2 \,\}$,
5. $\{\, (x, y) \mid x > 3 \text{ or } y < x^2 \,\}$, etc.

Typically, sets described by strict inequalities "$<$", "$>$", or "$\neq$", are open[33]. An example in $\mathbb{R}^3$ is

$$\{\, (x, y, z) \mid x^2 + y^2 < z^2 \text{ and } \sin x < \cos y \,\}.$$



By the following theorem, the intersection of a *finite* number of open sets is open, but the union of *any* number (possibly infinite) of open sets is open. The third example in the previous section shows that the intersection of an infinite collection of open sets need not be open.

THEOREM 10.5.2.

1. *The sets $\emptyset$ and $\mathbb{R}^n$ are both open.*
2. *If $A_1, \ldots, A_k$ is a* finite *collection of open subsets of $\mathbb{R}^n$, then their intersection $A_1 \cap \cdots \cap A_k$ is open.*
3. *If A_λ is an open subset of $\mathbb{R}^n$ for each $\lambda \in J$, then the union $\bigcup_\lambda A_\lambda$ of all the sets A_λ is also open.*

PROOF.

1. The first part is trivial (the empty set is open because every point in it certainly has the required property!)

2. Consider any point $\mathbf{a} \in A_1 \cap \cdots \cap A_k$. Then $\mathbf{a} \in A_k$ for *every* k, and since A_k is open there is a real number $r_k > 0$ such that $B_{r_k}(\mathbf{a}) \subseteq A_k$. Choosing $r =$

[32] We sometimes say S is open *in* $\mathbb{R}^n$ because there is a more general notion of S being open *in* E for any set E such that $S \subseteq E$. When we say S is *open* we will always mean open *in* $\mathbb{R}^n$, unless otherwise stated.

[33] Let $S = \{\, (x_1, \ldots, x_n) \mid f(x_1, \ldots, x_n) > g(x_1, \ldots, x_n) \,\}$. If f and g are continuous (we rigorously define continuity for functions of more than one variable later) and $\mathbf{a} = (a_1, \ldots, a_n)$ is a point in S, then all points in a sufficiently small neighbourhood of $\mathbf{a}$ will also satisfy the corresponding inequality and so be in S — see later. It follows that S is open.

If a set is described by a finite number of strict inequalities in terms of "and" and "or", then it will be obtained by taking finite unions and intersections of open sets as above, and hence be open by Theorem 10.5.2.

$\min\{r_1, \ldots, r_k\}$ we see that $r > 0$, and $B_r(\mathbf{a}) \subseteq A_k$ for each k. It follows[34] that

$$B_r(\mathbf{a}) \subseteq A_1 \cap \cdots \cap A_k.$$

Hence $A_1 \cap \cdots \cap A_k$ is open.

3. Consider any point $\mathbf{a} \in \bigcup_{\lambda \in J} A_\lambda$. Then $\mathbf{a} \in A_\lambda$ for *some* $\lambda \in J$. For this λ there is an $r > 0$ such that $B_r(\mathbf{a}) \subseteq A_\lambda$. It follows (from the definition of $\bigcup$) that

$$B_r(\mathbf{a}) \subseteq \bigcup_{\lambda} A_\lambda.$$

Hence $\bigcup_{\lambda \in J} A_\lambda$ is open. □

The reason the proof of **2.** does not show the intersection of an *infinite* number of open sets is always open is that we would need to take r to be the infimum of an *infinite* set of positive numbers. Such an infimum may be 0. Consider, for example, $\bigcap_{n \in \mathbb{N}} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$ from the previous section. If $\mathbf{a} = 0$ then $r_n = 1/n$, and the infimum of all the r_n is 0.

10.6. Closed sets.

DEFINITION 10.6.1. A set C is *closed* (in $\mathbb{R}^n$)[35] if its complement (in $\mathbb{R}^n$) is open.

If, in the examples after Definition 10.5.1, "$>$", "$<$" and $\neq$ are replaced by "$\geq$", "$\leq$" and $=$, then we have examples of closed sets. Generally, sets defined in terms of nonstrict inequalities and $=$ are closed.[36] Closed intervals in $\mathbb{R}$ are closed sets.

THEOREM 10.6.2.

1. *The sets $\emptyset$ and $\mathbb{R}^n$ are both closed.*
2. *If $A_1, \ldots, A_k$ is a* finite *collection of closed subsets of $\mathbb{R}^n$, then their union $A_1 \cup \cdots \cup A_k$ is closed.*
3. *If A_λ is a closed subset of $\mathbb{R}^n$ for each $\lambda \in J$, then the intersection $\bigcap_\lambda A_\lambda$ of all the sets $A_{\lambda \in J}$ is also closed.*

PROOF.

1. Since $\emptyset$ and $\mathbb{R}^n$ are complements of each other, the result follows from the corresponding part of the previous theorem.

2. By DeMorgan's Law,

$$(A_1 \cup \cdots \cup A_k)^c = A_1^c \cap \cdots \cap A_k^c.$$

Since $A_1^c \cap \cdots \cap A_k^c$ is open from part **2** of the previous theorem, it follows that $(A_1 \cup \cdots \cup A_k)$ is closed.

3. The proof is similar to **2**. Namely,

$$\left(\bigcap_{\lambda \in J} A_\lambda\right)^c = \bigcup_{\lambda \in J} A_\lambda^c,$$

and so the result follows from **3** in the previous theorem. □

[34] If $B \subseteq A_1, \ldots, B \subseteq A_k$, then it follows from the definition of $A_1 \cap \cdots \cap A_k$ that $B \subseteq A_1 \cap \cdots \cap A_k$.

[35] By *closed* we will always mean closed *in* $\mathbb{R}^n$. But as for open sets, there is a more general concept of being closed in an arbitrary E such that $S \subseteq E$.

[36] The complement of such a set is a set defined in terms of *strict* inequalities, and so is open. See the previous section.

10.7. Topological Spaces*. Motivated by the idea of open sets in $\mathbb{R}^n$ we make the following definition.

DEFINITION 10.7.1. A set $\mathcal{S}$, together with a collection of subsets of $\mathcal{S}$ (which are called the *open sets* in $\mathcal{S}$) which satisfy the following three properties, is called a *topological space*.

1. The sets $\emptyset$ and $\mathcal{S}$ are both open.
2. If $A_1, \ldots, A_k$ is a *finite* collection of open subsets of $\mathcal{S}$, then their intersection $A_1 \cap \cdots \cap A_k$ is open.
3. If A_λ is an open subset of $\mathcal{S}$ for each $\lambda \in J$, then the union $\bigcup_\lambda A_\lambda$ of all the sets A_λ is also open.

Essentially the same arguments as used in the case of $\mathbb{R}^n$ with the Euclidean metric d allow us to define the notion of a neighbourhood of a point in *any* metric space $(\mathcal{M}, d)$, and then to define open sets in any metric space, and then to show as in Theorem 10.5.2 that this gives a topological space.

In any topological space (which as we have just seen includes all metric spaces) one can define the notion of a closed set in an analogous way to that in Section 10.6 and prove the analogue of Theorem 10.6.2. The notion of boundary, interior and exterior of an open set in any topologoical space is defined as in Section 10.8.

The theory of topological spaces is fundamental in contemporary mathematics.

10.8. Boundary, Interior and Exterior.

DEFINITION 10.8.1. The point $\mathbf{a}$ is a *boundary point* of S if every neighbourhood of $\mathbf{a}$ contains both points in S and points not in S. The *boundary* of S ($\operatorname{bdry} S$ or ∂S) is the set of boundary points of S.

Thus $\partial B_r(\mathbf{a})$ is the set of points whose distance from $\mathbf{a}$ equals r. See the diagram in Section 10.5.

DEFINITION 10.8.2. The point $\mathbf{a}$ is an *interior point* of S if $\mathbf{a} \in S$ but $\mathbf{a}$ is not a boundary point of S. The point $\mathbf{a}$ is an *exterior point* of S if $\mathbf{a} \in S^c$ but $\mathbf{a}$ is not a boundary point of S.

The *interior* of S ($\operatorname{int} S$) consists of all interior points of S. The *exterior* of S ($\operatorname{ext} S$) consists of all exterior points of S.

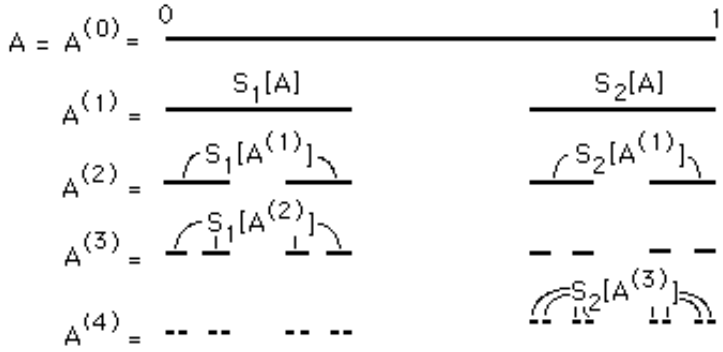
It follows that:

- $\mathbf{a}$ *is an interior point of* S *iff* $B_r(\mathbf{a}) \subseteq S$ *for some* $r > 0$,
- $\mathbf{a}$ *is an exterior point of* S *iff* $B_r(\mathbf{a}) \subseteq S^c$ *for some* $r > 0$
- $S = \operatorname{int} S \cup \partial S \cup \operatorname{ext} S$, *and every point in* $\mathbb{R}^n$ *is in exactly one of these three sets.*

Why?

Do exercises 1–8, page 647 of Adams, to reinforce the basic definitions.

10.9. Cantor Set. We next sketch a sequence of approximations $A = A^{(0)}$, $A^{(1)}$, $A^{(2)}$, ... to the *Cantor Set* C, the most basic fractal.



We can think of C as obtained by first removing the *open middle third* $(1/3, 2/3)$ from $[0, 1]$; then removing the open middle third from each of the two closed intervals which remain; then removing the open middle third from each of the four closed interval which remain; etc.

More precisely, let

$$\begin{aligned} A = A^{(0)} &= [0,1] \\ A^{(1)} &= [0,1/3] \cup [2/3,1] \\ A^{(2)} &= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] \\ &\vdots \end{aligned}$$

Let $C = \bigcap_{n\geq 0} A^{(n)}$. Since C is the intersection of a family of closed sets, C is closed.

Note that $A^{(n+1)} \subseteq A^{(n)}$ for all n and so the $A^{(n)}$ form a *decreasing* family of sets.

Consider the ternary expansion of numbers $x \in [0, 1]$, i.e. write each $x \in [0, 1]$ in the form

$$x = .a_1 a_2 \dots a_n \dots = \frac{a_1}{3} + \frac{a_2}{3^2} + \cdots + \frac{a_n}{3^n} + \cdots \tag{16}$$

where $a_n = 0$, 1 or 2. Each number has either one or two such representations, and the only way x can have two representations is if

$$x = .a_1 a_2 \dots a_n 222 \cdots = .a_1 a_2 \dots a_{n-1}(a_n+1)000 \dots$$

for some $a_n = 0$ or 1. For example, $.210222 \cdots = .211000 \dots$.

Note the following:

- $x \in A^{(n)}$ iff x has an expansion of the form (16) with each of $a_1, \dots, a_n$ taking the values 0 or 2.
- It follows that $x \in C$ iff x has an expansion of the form (16) with *every* a_n taking the values 0 or 2.
- Each endpoint of any of the 2^n intervals associated with $A^{(n)}$ has an expansion of the form (16) with each of $a_1, \dots, a_n$ taking the values 0 or 2 and the remaining a_i either *all* taking the value 0 or *all* taking the value 2.

Exercise: Show that C is uncountable.

Next let

$$S_1(x) = \frac{1}{3}x, \quad S_2(x) = 1 + \frac{1}{3}(x-1).$$

Notice that S_1 is a dilation with dilation ratio 1/3 and fixed point 0. Similarly, S_2 is a dilation with dilation ratio 1/3 and fixed point 1.

Then

$$A^{(n+1)} = S_1[A^{(n)}] \cup S_2[A^{(n)}].$$

Moreover,

$$C = S_1[C] \cup S_2[C].$$

The Cantor set is a *fractal*. It is composed of two parts each obtained from the original set by contracting by scaling (by 1/3).

If you think of a one, two or three dimensional set (e.g. line, square or cube), contracting by 1/3 should give a set which is $1/3^d$ of the original set in "size".

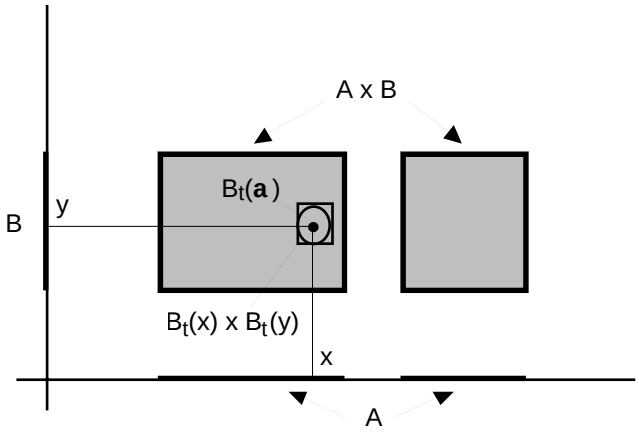
Thus if d is the "dimension" of the Cantor set, then we expect

$$\frac{1}{2} = \frac{1}{3^d}.$$

This gives $\log 2 = d \log 3$ or $d = \log 2/\log 3 \approx .6309$.

10.10. Product sets.

THEOREM 10.10.1. *If $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$ are both open (closed) then so is the product $A \times B \subseteq \mathbb{R}^{m+n}$.*



PROOF. For simplicity of notation, take $m = n = 1$.

First suppose A and B are open.

If $\mathbf{a} = (x, y) \in A \times B \subseteq \mathbb{R}^2$ then $x \in A$ and $y \in B$. Choose r and s so $B_r(x) \subseteq A$ and $B_s(y) \subseteq B$.

Let $t = \min\{r, s\}$. Then

$$B_t(\mathbf{a}) \subseteq B_r(x) \times B_s(y) \subseteq A \times B.$$

The first "$\subseteq$" is because

$$\begin{aligned}
(x_0, y_0) \in B_t(\mathbf{a}) \quad &\Rightarrow \quad d\big((x_0, y_0), (x, y)\big) < t \\
&\Rightarrow \quad d(x_0, x) < t \text{ and } d(y_0, y) < t \\
&\Rightarrow \quad x_0 \in B_t(x) \text{ and } y_0 \in B_t(y) \\
&\Rightarrow \quad (x_0, y_0) \in B_r(x) \times B_s(y).
\end{aligned}$$

Since $B_t(\mathbf{a}) \subseteq A \times B$ it follows that $A \times B$ is open.

Next suppose A and B are closed. Note that

$$(A \times B)^c = A^c \times \mathbb{R} \ \cup \ \mathbb{R} \times B^c,$$

since $(x, y) \notin A \times B$ iff (either $x \notin A$ or $y \notin B$ (or both)). Hence the right side is the union of two open sets by the above, and hence is open. Hence $A \times B$ is closed. $\square$

11. Sequences in $\mathbb{R}^p$

See Reed Problem 8 page 51, Problems 10–12 on page 59 and the first Definition on page 152. But we do much more!

11.1. Convergence.

DEFINITION 11.1.1. A sequence of points $(\mathbf{x}_n)$ from $\mathbb{R}^p$ converges to $\mathbf{x}$ if $d(\mathbf{x}_n, \mathbf{x}) \to 0$.

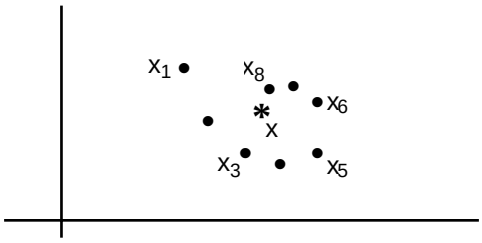
Thus the notion of convergence of a sequence from $\mathbb{R}^p$ is reduced to the notion of convergence of a sequence of *real* numbers to 0.

EXAMPLE 11.1.2. Let $\theta \in \mathbb{R}$ and $\mathbf{a} = (x^*, y^*) \in \mathbb{R}^2$ be fixed, and let

$$\mathbf{a}_n = \Big(x^* + \frac{1}{n}\cos n\theta,\ y^* + \frac{1}{n}\sin n\theta\Big).$$

Then $\mathbf{a}_n \to \mathbf{a}$ as $n \to \infty$. The sequence $(\mathbf{a}_n)$ spirals around $\mathbf{a}$ with $d(\mathbf{a}_n, \mathbf{a}) = \frac{1}{n}$ and with rotation by the angle θ in passing from $\mathbf{a}_n$ to $\mathbf{a}_{n+1}$.

The next theorem shows that a sequence from $\mathbb{R}^p$ converges iff each sequence of components converges. Think of the case $p = 2$.



The shown sequence converges, as does its projection on to the x and y-axes

PROPOSITION 11.1.3. *The sequence* $\mathbf{x}_n \to \mathbf{x}$ *iff* $x_n^{(k)} \to x^{(k)}$ *for* $k = 1, \ldots, p$.

PROOF. Assume $\mathbf{x}_n \to \mathbf{x}$. Then $d(\mathbf{x}_n, \mathbf{x}) \to 0$. Since $|x_n^{(k)} - x^{(k)}| \le d(\mathbf{x}_n, \mathbf{x})$, it follows that $|x_n^{(k)} - x^{(k)}| \to 0$ and so $x_n^{(k)} \to x^{(k)}$, as $n \to \infty$.

Next assume $x_n^{(k)} \to x^{(k)}$ for each k, i.e. $|x_n^{(k)} - x^{(k)}| \to 0$, as $n \to \infty$. Since

$$d(\mathbf{x}_n, \mathbf{x}) = \sqrt{(x_n^{(1)} - x^{(1)})^2 + \cdots + (x_n^{(p)} - x^{(p)})^2},$$

it follows that $d(\mathbf{x}_n, \mathbf{x}) \to 0$ and so $\mathbf{x}_n \to \mathbf{x}$. □

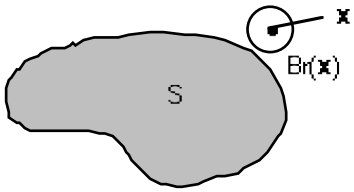
11.2. Closed sets. The reason we call a set "closed" is that it is closed under the operation of taking limits of any sequence of points from the set. This is the content of the next theorem.

THEOREM 11.2.1. *A set* $S \subseteq \mathbb{R}^p$ *is closed iff every convergent sequence* $(\mathbf{x}_n) \subseteq S$ *has its limit in* S.

PROOF. First suppose S is closed. Let $(\mathbf{x}_n) \subseteq S$ be a convergent sequence with limit $\mathbf{x}$.

Assume $\mathbf{x} \notin S$, i.e. $\mathbf{x} \in S^c$. Since S^c is open, there is an $r > 0$ such that $B_r(\mathbf{x}) \subseteq S^c$. But since $\mathbf{x}_n \to \mathbf{x}$, ultimately[37] $\mathbf{x}_n \in B_r(\mathbf{x}) \subseteq S^c$.

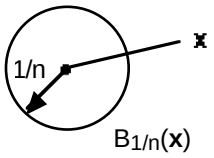
[37] *Ultimately* means there is an N such that this is true for all $n \ge N$.



If $\mathbf{x}_n \to \mathbf{x}$ then ultimately $\mathbf{x}_n \in B_r(\mathbf{x})$

This contradicts the fact $\mathbf{x}_n \in S$ for all n. Hence the assumption is wrong and so $\mathbf{x} \in S$

Next suppose that every convergent sequence $(\mathbf{x}_n) \subseteq S$ has its limit in S. *Assume* S is not closed, i.e. S^c is not open. Then there is a point $\mathbf{x} \in S^c$ such that for no $r > 0$ is it true that $B_r(\mathbf{x}) \subseteq S^c$. In particular, for each natural number n there is a point in $B_{1/n}(\mathbf{x})$ which belongs to S. Choose such a point and call it $\mathbf{x}_n$. Then $\mathbf{x}_n \to \mathbf{x}$ since $d(\mathbf{x}_n, \mathbf{x}) \leq 1/n \to 0$.



If for each n there is a member of S in $B_{1/n}(\mathbf{x})$,
then there is a sequence from S converging to **x** .

This means that the sequence $(\mathbf{x}_n) \subseteq S$ but its limit is not in S. This is a contradiction and so the assumption is wrong. Hence S is closed. □

11.3. Bounded sets.

DEFINITION 11.3.1. A set $S \subseteq \mathbb{R}^p$ is *bounded* if there is a point $\mathbf{a} \in \mathbb{R}^p$ and a real number R such that $S \subseteq B_R(\mathbf{a})$.

A sequence $(\mathbf{x}_n)$ is *bounded* if there is a point $\mathbf{a} \in \mathbb{R}^p$ and a real number R such that $\mathbf{x}_n \in B_R(\mathbf{a})$ for every n.

Thus a sequence is bounded iff the corresponding set of members is bounded.

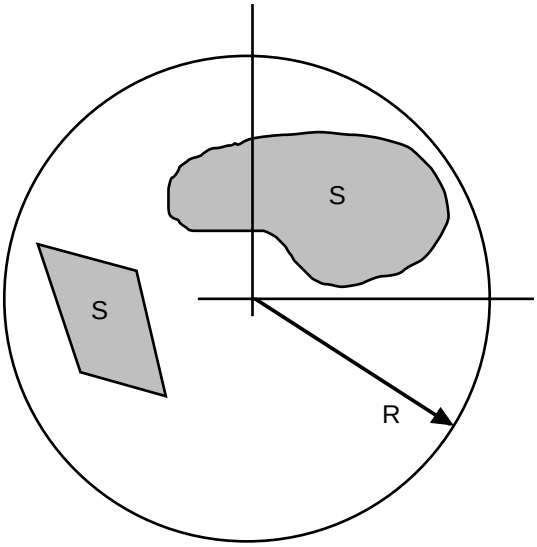
PROPOSITION 11.3.2. *A set is bounded iff there is a ball (of finite radius) centred at the origin which includes the set.*

PROOF. This is clear in $\mathbb{R}^2$ from a diagram. It is proved from the triangle inequality as follows.

Suppose $S \subseteq B_R(\mathbf{a})$. Let $r = d(\mathbf{a}, \mathbf{0})$. Then for any $\mathbf{x} \in S$,

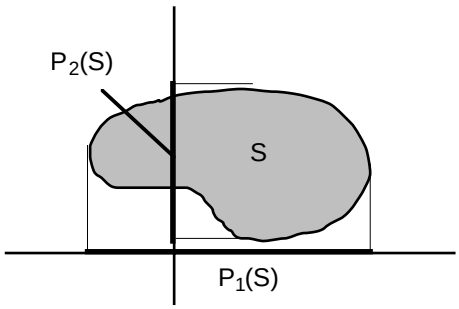
$$d(\mathbf{x}, \mathbf{0}) \leq d(\mathbf{x}, \mathbf{a}) + d(\mathbf{a}, \mathbf{0}) < R + r.$$

Hence $S \subseteq B_{R+r}(\mathbf{0})$. □



The following theorem says that a set S is bounded iff its projection on to each of the axes is bounded. The projection onto the kth axis is defined by

$$P_k(S) = \{\, a \mid \mathbf{x} \in S \text{ for some } \mathbf{x} \text{ whose } k\text{th component is } a \,\}.$$



THEOREM 11.3.3. *A set $S \subseteq \mathbb{R}^p$ is bounded iff $P_k(S)$ is bounded for each $k = 1, \ldots, p$.*

PROOF. If $S \subseteq B_R(\mathbf{0})$ then each $P_k(S) \subseteq [-R, R]$ and so is bounded.

Conversely, if each $P_k(S)$ is bounded then by choosing the largest interval, we may assume $P_k(S) \subseteq [-M, M]$ for every k. It follows that if $\mathbf{x} \in S$ then every component of $\mathbf{x}$ is at most M in absolute value. It follows that

$$|\mathbf{x}| \leq \sqrt{kM^2},$$

i.e. $S \subseteq B_R(\mathbf{0})$ where $R = \sqrt{k}M$. □

In particular, if a sequence of points from $\mathbb{R}^p$ is bounded, then so are the sequences of real numbers corresponding to the first component, to the second component, etc.

PROPOSITION 11.3.4. *A convergent sequence is bounded.*

PROOF. Suppose $\mathbf{x}_n \to \mathbf{x}$. Then ultimately $\mathbf{x}_n \in B_1(\mathbf{x})$, say for $n \geq N$, and so the set of terms $\{\, \mathbf{x}_n \mid n \geq N \,\}$ is bounded. The set $\{\mathbf{x}_1, \ldots, \mathbf{x}_{N-1}\}$ is also bounded (being finite) and so the set of *all* terms is bounded, being the union of two bounded sets. In other words, the sequence is bounded. □

11.4. Bolzano-Weierstraß theorem. As for sequences of real numbers, a *subsequence* of a sequence is obtained by skipping terms.

PROPOSITION 11.4.1. *If* $(\mathbf{x}_n) \subseteq \mathbb{R}^p$ *and* $\mathbf{x}_n \to \mathbf{x}$*, then any subsequence also converges to* $\mathbf{x}$*.*

PROOF. Let $(\mathbf{x}_{n(k)})_{k\geq 1}$ be a subsequence of $(\mathbf{x}_n)$. Suppose $\epsilon > 0$ is given.

Since $\mathbf{x}_n \to \mathbf{x}$ there is an integer N such that

$$k \geq N \quad \Rightarrow \quad d(\mathbf{x}_k, \mathbf{x}) \leq \epsilon.$$

But if $k \geq N$ then also $n(k) \geq N$ and so

$$k \geq N \quad \Rightarrow \quad d(\mathbf{x}_{n(k)}, \mathbf{x}) \leq \epsilon.$$

That is, $\mathbf{x}_{n(k)} \to \mathbf{x}$. □

The following theorem is analogous to the Bolzano-Weierstraß Theorem 5.2.2 for real numbers, and in fact follows from it.

THEOREM 11.4.2 (Bolzano-Weierstraß Theorem). *If a sequence* $(\mathbf{a}_n) \subseteq \mathbb{R}^p$ *is bounded then it has a convergent subsequence. If all members of* $(\mathbf{a}_n)$ *belong to a closed bounded set* S*, then so does the limit of any convergent subsequence of* $(\mathbf{a}_n)$*.*

PROOF. For notational simplicity we prove the case $p = 2$, but clearly the proof easily generalises to $p > 2$.

Suppose then that $(\mathbf{a}_n) \subseteq \mathbb{R}^p$ is bounded.

Write $\mathbf{a}_n = (x_n, y_n)$. Then the sequences (x_n) and (y_n) are also bounded (see the paragraph after Theorem 11.3.3).

Let $(x_{n(k)})$ be a convergent subsequence of (x_n) (this uses the Bolzano-Weierstraß Theorem 5.2.2 for real sequences).

Now consider the sequence $(x_{n(k)}, y_{n(k)})$. Let $(y_{n'(k)})$ be a convergent subsequence of $(y_{n(k)})$ (again by the Bolzano-Weierstraß Theorem for real sequences).

Note that $(x_{n'(k)})$ is a subsequence of the convergent sequence $(x_{n(k)})$, and so also converges.

Since $(x_{n'(k)})$ and $(y_{n'(k)})$ converge, so does $(\mathbf{a}_{n'(k)}) = ((x_{n'(k)}, y_{n'(k)}))$.[38]

Finally, since S is closed, any convergent subsequence from S must have its limit in S. □

.

11.5. Cauchy sequences.

DEFINITION 11.5.1. A sequence $(\mathbf{x}_n)$ is a *Cauchy sequence* if for any $\epsilon > 0$ there is a corresponding N such that

$$m, n \geq N \quad \Rightarrow \quad d(\mathbf{x}_m, \mathbf{x}_n) \leq \epsilon.$$

That is, ultimately *any* two members of the sequence (not necessarily consecutive members) are within ϵ of each another.

[38] After a bit of practice, we would probably write the previous argument as follow:

PROOF. Since $(\mathbf{a}_n) \subseteq \mathbb{R}^p$, where $\mathbf{a}_n = (x_n, y_n)$, is bounded, so is each of the component sequences.

By the Bolzano-Weierstraß Theorem for real sequences, on passing to a subsequence (but without changing notation) we may assume (x_n) converges. Again by the Bolzano-Weierstraß Theorem for real sequences, on passing to a further subsequence (without changing notation) we may assume that (y_n) also converges. (This gives a new subsequence (x_n), but it converges since any subsequence of a convergent sequence is also convergent).

Since the subsequences (x_n) and (y_n) converge, so does the subsequence $(\mathbf{a}_n) = ((x_n, y_n))$. □

PROPOSITION 11.5.2. *The sequence $(\mathbf{x}_n)_{n\geq 1}$ in $\mathbb{R}^p$ is Cauchy iff the component sequences $(x_n^{(k)})_{n\geq 1}$ (of real numbers) are Cauchy for $k = 1, \ldots, p$.*

PROOF. (The proof is similar to that for Proposition 11.1.3.)

Assume $(\mathbf{x}_n)$ is Cauchy. Since $|x_m^{(k)} - x_n^{(k)}| \leq d(\mathbf{x}_m, \mathbf{x}_n)$, it follows that $(x_n^{(k)})_{n\geq 1}$ is a Cauchy sequence, for $k = 1, \ldots, p$

Conversely, assume $(x_n^{(k)})_{n\geq 1}$ is a Cauchy sequence, for $k = 1, \ldots, p$. Since

$$d(\mathbf{x}_m, \mathbf{x}_n) = \sqrt{(x_m^{(1)} - x_n^{(1)})^2 + \cdots + (x_m^{(p)} - x_n^{(p)})^2},$$

it follows that $(\mathbf{x}_n)$ is Cauchy. □

THEOREM 11.5.3. *A sequence in $\mathbb{R}^p$ is Cauchy iff it converges.*

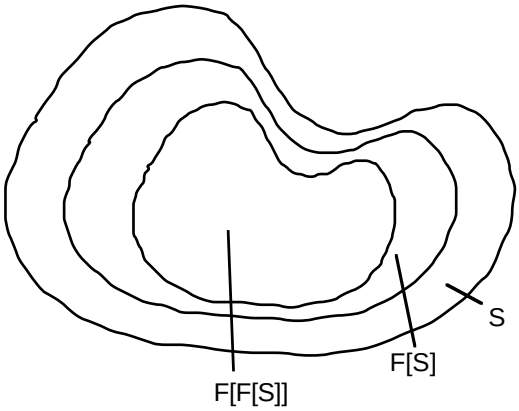
PROOF. A sequence is Cauchy iff each of the component sequences is Cauchy (Proposition 11.5.2) iff each of the component sequences converges (Theorem 5.1.2) iff the sequence converges (Proposition 11.1.3). □

11.6. Contraction Mapping Principle in $\mathbb{R}^p$. The reference is Reed Section 5.7, but with $\mathbb{R}^p$ and d instead of $\mathcal{M}$ and ρ. In Section 15.5 of these notes we will generalise this to any *complete metric space* with the same proof. But until then, omit Reed Examples 1 and 3 and all except the first four lines on page 208.

Suppose $S \subseteq \mathbb{R}^p$. We say $F : S \to S$ is a *contraction* if there exists $\lambda \in [0, 1)$ such that

$$d(F(\mathbf{x}), F(\mathbf{y})) \leq \lambda d(\mathbf{x}, \mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in S$.



* A contraction map is continuous, and in fact uniformly continuous. More generally, any Lipschitz map F is continuous, where Lipschitz means for some $\lambda \in [0, \infty)$ and all $x, y \in \mathcal{D}(f)$,

$$d(F(x), F(y)) \leq \lambda d(x, y).$$

The definitions of continuity and uniform continuity are essentially the same as in the one dimensional case, see Section 13.1.3. The proof that Lipschitz implies uniform continuity is essentially the same as in the one dimensional case, see Proposition 7.5.4.

A simple example of a contraction map on $\mathbb{R}^p$ is the map

$$F(\mathbf{x}) = \mathbf{a} + r(\mathbf{x} - \mathbf{b}), \tag{17}$$

for $0 \leq r < 1$. In this case $\lambda = r$, as is easily checked. Since

$$\mathbf{a} + r(\mathbf{x} - \mathbf{b}) = \big(\mathbf{b} + r(\mathbf{x} - \mathbf{b})\big) + \mathbf{a} - \mathbf{b},$$

we see (17) is dilation about $\mathbf{b}$ by the factor r, followed by translation by the vector $\mathbf{a}-\mathbf{b}$.

We say $\mathbf{z}$ is a *fixed point* of the map $F : S \to S$ if $F(\mathbf{z}) = \mathbf{z}$.

In the preceding example, the unique fixed point for *any* $r \neq 1$ (even $r > 1$) is $(\mathbf{a} - r\mathbf{b})/(1-r)$, as is easily checked.

The following theorem (and its generalisations to closed subsets of a complete metric spaces) is known as the *Contraction Mapping Theorem*, the *Contraction Mapping Principle* or the *Banach Fixed Point Theorem*. It has many important applications.

THEOREM 11.6.1. *Let $F : S \to S$ be a contraction map, where $S \subseteq \mathbb{R}^p$ is closed. Then F has a unique fixed point $\mathbf{x}$.*[39]

Moreover, iterates of F applied to any initial point $\mathbf{x}_0$ converge to $\mathbf{x}$, and

$$d(\mathbf{x}_n, \mathbf{x}) \leq \frac{\lambda^n}{1-\lambda}\, d(\mathbf{x}_0, F(\mathbf{x}_0)). \tag{18}$$

PROOF. We will find the fixed point as the limit of a Cauchy sequence.

Let $\mathbf{x}_0$ be any point in S and define a sequence $(\mathbf{x}_n)_{n\geq 0}$ by

$$\mathbf{x}_1 = F(\mathbf{x_0}),\ \mathbf{x}_2 = F(\mathbf{x}_1),\ \mathbf{x}_3 = F(\mathbf{x}_2), \dots,\ \mathbf{x}_n = F(\mathbf{x}_{n-1}), \dots .$$

Let λ be the contraction ratio.

1. Claim: $(\mathbf{x}_n)$ *is Cauchy.*

We have

$$d(\mathbf{x}_n, \mathbf{x}_{n+1}) = d(F(\mathbf{x}_{n-1}), F(\mathbf{x}_n) \leq \lambda d(\mathbf{x}_{n-1}, \mathbf{x}_n).$$

By iterating this we get

$$d(\mathbf{x}_n, \mathbf{x}_{n+1}) \leq \lambda^n d(\mathbf{x}_0, \mathbf{x}_1).$$

Thus if $m > n$ then

$$\begin{aligned} d(\mathbf{x}_n, \mathbf{x}_m) &\leq d(\mathbf{x}_n, \mathbf{x}_{n+1}) + \cdots + d(\mathbf{x}_{m-1}, \mathbf{x}_m) \\ &\leq (\lambda^n + \cdots + \lambda^{m-1})\, d(\mathbf{x}_0, \mathbf{x}_1). \end{aligned} \tag{19}$$

But

$$\begin{aligned} \lambda^n + \cdots + \lambda^{m-1} &\leq \lambda^n(1 + \lambda + \lambda^2 + \cdots) \\ &= \frac{\lambda^n}{1-\lambda} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

It follows (why?) that $(\mathbf{x}_n)$ is Cauchy.

Hence $(\mathbf{x}_n)$ converges to some limit $\mathbf{x} \in \mathbb{R}^p$. Since S is closed it follows that $\mathbf{x} \in S$.

2. Claim: $\mathbf{x}$ *is a fixed point of F.*

We will show that $d(\mathbf{x}, F(\mathbf{x})) = 0$ and so $\mathbf{x} = F(\mathbf{x})$. In fact

$$\begin{aligned} d(\mathbf{x}, F(\mathbf{x})) &\leq d(\mathbf{x}, \mathbf{x}_n) + d(\mathbf{x}_n, F(\mathbf{x})) \\ &= d(\mathbf{x}, \mathbf{x}_n) + d(F(\mathbf{x}_{n-1}), F(\mathbf{x})) \\ &\leq d(\mathbf{x}, \mathbf{x}_n) + \lambda d(\mathbf{x}_{n-1}, \mathbf{x}) \\ &\to 0 \end{aligned}$$

as $n \to \infty$. This establishes the claim.

3. Claim: The fixed point is unique.

[39] In other words, F has *exactly one* fixed point.

If $\mathbf{x}$ and $\mathbf{y}$ are fixed points, then $F(\mathbf{x}) = \mathbf{x}$ and $F(\mathbf{y}) = \mathbf{y}$ and so

$$d(\mathbf{x}, \mathbf{y}) = d(F(\mathbf{x}), F(\mathbf{y})) \leq \lambda d(\mathbf{x}, \mathbf{y}).$$

Since $0 \leq \lambda < 1$ this implies $d(\mathbf{x}, \mathbf{y}) = 0$, i.e. $\mathbf{x} = \mathbf{y}$.

4. From (19) and the lines which follow it, we have

$$d(\mathbf{x}_n, \mathbf{x}_m) \leq \frac{\lambda^n}{1-\lambda} d(\mathbf{x}_0, F(\mathbf{x}_0)).$$

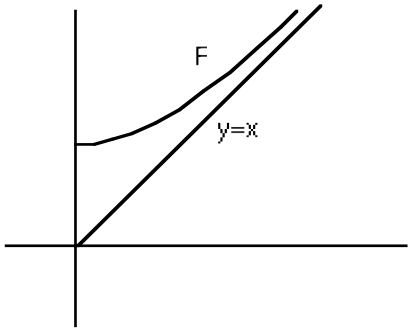
Since $\mathbf{x}_m \to \mathbf{x}$, (18) follows. □

REMARK 11.6.1. It is essential that there is a *fixed* $\lambda < 1$. For example, the function $F(x) = x^2$ is a contraction on $[0, a]$ for each $0 \leq a < \frac{1}{2}$ but is not a contraction on $[0, \frac{1}{2}]$. However, in this case there is still a fixed point in $[0, \frac{1}{2}]$, namely 0.

To obtain an example where $d(F(x), F(y)) < d(x, y)$ for all $x, y \in S$ but there is no fixed point, consider a function F with the properties

$$x < F(x), \quad 0 \leq F'(x) < 1,$$

for all $x \in [0, \infty)$, see the following diagram. (Can you write down an analytic expression?)



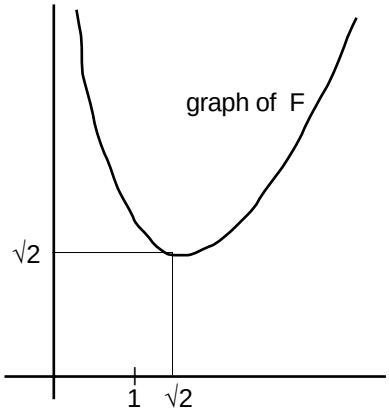
By the Mean Value Theorem, $d(F(x), F(y)) < d(x, y)$ for all $x, y \in \mathbb{R}$. But since $x < F(x)$ for all x, there is no fixed point.

EXAMPLE 11.6.2. (*Approximating* $\sqrt{2}$) The map

$$F(x) = \frac{1}{2}\left(x + \frac{2}{x}\right)$$

takes $[1, \infty)$ into itself and is a contraction with $\lambda = \frac{1}{2}$.

The first claim is clear from a graph, and can be then checked in various standard ways.



Since $F'(x) = \frac{1}{2} - \frac{1}{x^2}$, it follows for $x \in [1, \infty)$ that

$$-\frac{1}{2} \leq F'(x) < \frac{1}{2}.$$

Hence

$$|F(x_1) - F(x_2)| \leq \frac{1}{2}|x_1 - x_2|$$

by the Mean Value Theorem. This proves F is a contraction with $\lambda = \frac{1}{2}$.

The unique fixed point is $\sqrt{2}$. It follows that we can approximate $\sqrt{2}$ by beginning from any $x_0 \in [1, \infty)$ and iterating F. In particular, the following provide approximations to $\sqrt{2}$:

$$1,\ \frac{3}{2},\ \frac{17}{12} \approx 1.417,\ \frac{577}{408} \approx 1.4142156.$$

(Whereas $\sqrt{2} = 1.41421356\ldots$.) This was known to the Babylonians, nearly 4000 years ago.

EXAMPLE 11.6.3 (Stable fixed points). See Reed, Theorem 5.7.2 on page 206 and the preceding paragraph.

Suppose $F : S \to S$ is continuously differentiable on some open interval $S \subseteq \mathbb{R}$ and $F(x^*) = x^*$ (i.e. x^* is a fixed point of F). We say x^* is a *stable* fixed point if there is an $\varepsilon > 0$ such that $F^n(x_0) \to x^*$ for every $x_0 \in [x^* - \varepsilon, x^* + \varepsilon]$.

If $|F'(x^*)| < 1$ *it follows* from the Mean Value Theorem and the Contraction Mapping Principle *that* x^* *is a stable fixed point of* F (Reed, Theorem 5.7.2 page 206).

EXAMPLE 11.6.4 (The Quadratic Map). See Reed Example 4 on page 208.

In Section 6 we considered the quadratic map

$$F : [0, 1] \to [0, 1], \quad F(x) = rx(1 - x).$$

The only fixed points of F are $x = 0$ and $x = x^* := 1 - 1/r$. As noted in Section 6, if $1 < r \leq 3$ it can be shown that, beginning from *any* $x_0 \in (0, 1)$, iterates of F converge to x^*. The proof is quite long.

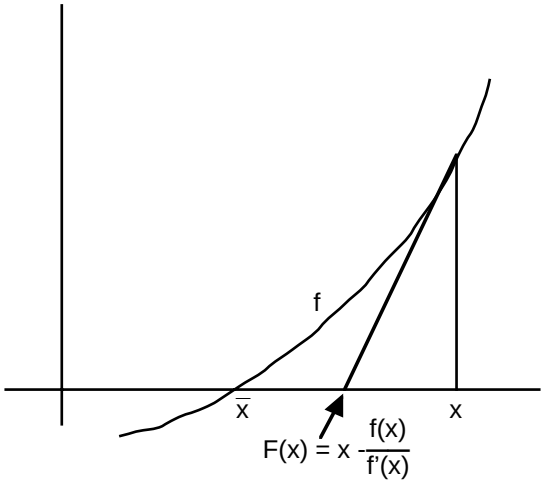
However, we can prove less, but much more easily.

Note that $F'(x) = r(1 - 2x)$ and so $F'(x^*) = 2 - r$. If $1 < r < 3$ then $|F'(x^*)| < 1$. Hence x^* is a stable fixed point of F, and so iterates $F^n(x_0)$ converge to x^* provided x_0 is *sufficiently close* to x^*.

EXAMPLE 11.6.5. (*Newton's Method*) (Recall Reed Theorem 4.4.1) Suppose that $f(\overline{x}) = 0$, $f'(\overline{x}) \neq 0$ and f'' exists and is bounded in some interval containing $\overline{x}$.

Then Newton's Method is to iterate the function

$$F(x) = x - \frac{f(x)}{f'(x)}.$$



Here is a very short proof that the method works: Since

$$F'(x) = \frac{f(x)}{(f'(x))^2} f''(x),$$

we see $F'(\overline{x}) = 0$. Hence $\overline{x}$ is a stable fixed point of F and so iterates $F^n(x_0)$ converge to $\overline{x}$ for all x_0 sufficiently close to $\overline{x}$.

* We can even show quadratic convergence with a little extra work. The main point is that the contraction ratio converges to 0 as $x \to \overline{x}$.

12. Compact subsets of $\mathbb{R}^p$

In this section we see that a subset of $\mathbb{R}^p$ is closed and bounded iff it is sequentially compact iff it is compact. We also see that a subset of an *arbitrary* metric space is sequentially compact iff it is compact, that either implies the subset is closed and bounded, but that (closed and bounded) need not imply compact (or sequentially compact).

In the next section we will see that compact sets have nice properties with respect to continuous functions (Theorems 13.2.1 and 13.5.1).

12.1. Sequentially compact sets.

DEFINITION 12.1.1. A set $A \subseteq \mathbb{R}^p$ is *sequentially compact* if every $(\mathbf{x}_n) \subseteq A$ has a convergent subsequence with limit in A.

We saw in the Bolzano-Weierstraß Theorem 11.4.2 that a closed bounded subset of $\mathbb{R}^p$ is sequentially compact. The converse is also true.

THEOREM 12.1.2. *A set $A \subseteq \mathbb{R}^p$ is closed and bounded iff it is sequentially compact.*

PROOF. We only have one direction left to prove.

So suppose A is sequentially compact.

To show that A is *closed* in $\mathbb{R}^p$, suppose $(\mathbf{x}_n) \subseteq A$ and $\mathbf{x}_n \to \mathbf{x}$. We need to show that $\mathbf{x} \in A$.

By sequential compactness, some subsequence $(\mathbf{x}_{n'})$ converges to some $\mathbf{x}' \in A$. But from Proposition 11.4.1 any subsequence of $(\mathbf{x}_n)$ must converge to the *same* limit as $(\mathbf{x}_n)$. Hence $\mathbf{x}' = \mathbf{x}$ and so $\mathbf{x} \in A$. Thus A is closed.

To show that A is *bounded*, assume otherwise. Then for each n we can choose some $\mathbf{a}_n \in A$ with $\mathbf{a}_n \notin B_n(\mathbf{0})$, i.e. with $d(\mathbf{a}_n, \mathbf{0}) > n$. It follows that any subsequence of $(\mathbf{a}_n)$ is unbounded. But this means no subsequence is convergent by Proposition 11.3.4. This contradicts the fact A is compact, and so the assumption is false, i.e. A is bounded. □

REMARK 12.1.1. The above result is true in an arbitrary metric space in only one direction. Namely, sequentially compact implies closed and bounded and the proof is essentially the same. The other direction requires the Heine Borel theorem, which need not hold in an arbitrary metric space.

12.2. Compact sets. An *open cover* of a set A is a collection of open sets A_λ (indexed by $\lambda \in J$, say) such that

$$A \subseteq \bigcup_{\lambda \in J} A_\lambda.$$

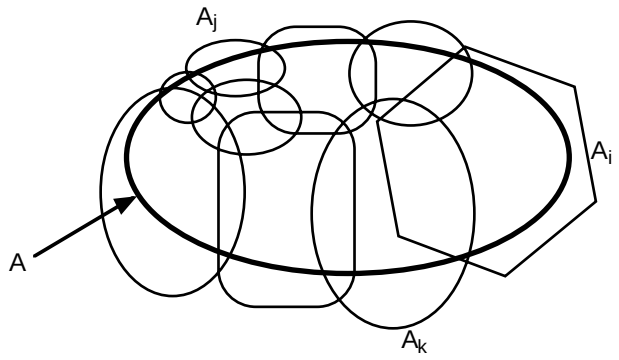
We denote the collection of sets by $\{ A_\lambda \mid \lambda \in J \}$.

A *finite subcover* of the given cover is a finite subcollection which still covers A. That is, a finite subcover is a collection of sets $\{ A_\lambda \mid \lambda \in J_0 \}$ for some finite $J_0 \subseteq J$, such that

$$A \subseteq \bigcup_{\lambda \in J_0} A_\lambda.$$

(Thus a finite subcover exists if one can keep just a finite number of sets from the original collection and still cover A.)

The following is an example of a cover of A. This particular cover is already finite.



EXAMPLE 12.2.1. An open cover of a set which has *no* finite subcover is

$$(0,1) \subseteq \bigcup_{n\geq 2} \left(\frac{1}{n}, 1\right).$$

This is easy to see, *why*?

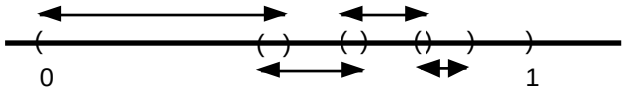
Another example is

$$(0,1) \subseteq \bigcup_{n\geq 2} \left(1-\frac{2}{n}, 1-\frac{1}{n}\right).$$

We can also write the cover as

$$\left(0,\frac{1}{2}\right) \cup \left(\frac{1}{3},\frac{2}{3}\right) \cup \left(\frac{2}{4},\frac{3}{4}\right) \cup \left(\frac{3}{5},\frac{4}{5}\right) \cup \left(\frac{4}{6},\frac{5}{6}\right) \cup \ldots$$

The following is a rough sketch of the first four intervals in the cover, indicated by lines with an arrow at each end.



The intervals overlap since $1-\frac{1}{n} > 1-\frac{2}{n+1}$ if $n \geq 2$ (because this inequality is equivalent to the inequality $\frac{1}{n} < \frac{2}{n+1}$, which in turn is equivalent to the inequality $\frac{n+1}{n} < 2$, which in turn is equivalent to the inequality $1+\frac{1}{n} < 2$, which is certainly true.) The intervals cover $(0,1)$ because they overlap and because $1-\frac{1}{n} \to 1$.

But it is clear that there is no finite subcover. (If n is the largest integer such that $\left(1-\frac{2}{n}, 1-\frac{1}{n}\right)$ is in some finite collection of such intervals, then no x between $1-\frac{1}{n}$ and 1 is covered.)

DEFINITION 12.2.2. A set $A \subseteq \mathbb{R}^p$ is *compact* if every open cover of A has a finite subcover.

Note that we require a finite subcover of *every* open cover of A.

EXAMPLE 12.2.3. Every finite set A is compact. (This is a simple case of the general result that every closed bounded set is compact — see Remark **??**.)

To see this, suppose $A \subseteq \bigcup_{\lambda\in J} A_\lambda$. For each $a \in A$ choose one A_λ which contains a. The collection of A_λ chosen in this way is a finite subcover of A.

We next prove that if a subset of $\mathbb{R}^p$ is compact, then it is closed and bounded. The converse direction is true for subsets of $\mathbb{R}^p$, but not for subsets of an arbitrary metric space — we show all this later.

THEOREM 12.2.4. *If $A \subseteq \mathbb{R}^p$ is compact, then it is closed and bounded.*

PROOF. Suppose $A \subseteq \mathbb{R}^p$ is compact.

For *any* set $A \subseteq \mathbb{R}^p$,

$$A \subseteq \bigcup_{n \geq 1} B_n(\mathbf{0}),$$

since the union of all such balls is $\mathbb{R}^p$. Since A is compact, there is a finite subcover. If N is the radius of the largest ball in the subcover, then

$$A \subseteq B_N(\mathbf{0}).$$

Hence A is bounded.

If A is not closed then there exists a sequence $(\mathbf{a}_n) \subseteq A$ with $\mathbf{a}_n \to \mathbf{b} \notin A$.

For each $\mathbf{a} \in A$, let $r_{\mathbf{a}} = \frac{1}{2} d(\mathbf{a}, \mathbf{b})$. Clearly,

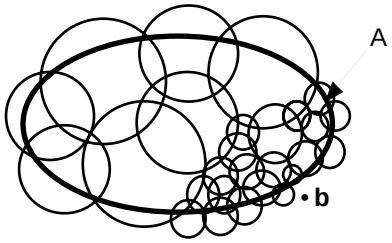
$$A \subseteq \bigcup_{\mathbf{a} \in A} B_{r_{\mathbf{a}}}(\mathbf{a}).$$

By compactness there is a finite subcover:

$$A \subseteq B_{r_1}(\mathbf{a}_1) \cup \cdots \cup B_{r_n}(\mathbf{a}_n),$$

say.

Let r be the minimum radius of any ball in the subcover ($r > 0$ as we are taking the minimum of a *finite* set of positive numbers). Every point $\mathbf{x} \in A$ is in some $B_{r_i}(\mathbf{a}_i)$ $(i = 1, \ldots, n)$ and so within distance r_i of $\mathbf{a}_i$. Since the distance from $\mathbf{a}_i$ to $\mathbf{b}$ is $2r_i$, the distance from $\mathbf{x}$ to $\mathbf{b}$ is at least r_i and hence at least r. But this contradicts the fact that there is a sequence from A which converges to $\mathbf{b}$.



Hence A *is* closed. □

12.3. Compact sets. We next prove that sequentially compact sets and compact sets in $\mathbb{R}^p$ are the same. The proofs in both directions are starred. They both generalise to subsets of arbitrary metric spaces.

It follows that the three notions of closed and bounded, of sequential compactness, and of compactness, all agree in $\mathbb{R}^p$. To prove this, without being concerned about which arguments generalise to arbitrary metric spaces, it is sufficient to consider Theorem 12.1.2 (closed and bounded implies sequentially compact), Theorem 12.2.4 (compact implies closed and bounded), and the following Theorem 12.3.1 just in one direction (sequentially compact implies compact).

THEOREM 12.3.1. *A set $A \subseteq \mathbb{R}^p$ is sequentially compact iff it is compact.*

PROOF*. First suppose A is sequentially compact. Suppose

$$A \subseteq \bigcup_{\lambda \in J} A_\lambda,$$

where $\{ A_\lambda \mid \lambda \in J \}$ is an open cover.

We first *claim* there is a subcover which is either countable or finite. (This is true for *any* set A, not necessarily closed or bounded.)

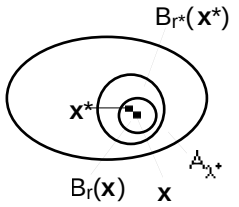
To see this, consider each point $\mathbf{x} = (x_1, \ldots, x_p) \in A$ all of whose coordinates are rational, and consider each rational $r > 0$. For such an $\mathbf{x}$ and r, if the ball

$B_r(\mathbf{x}) \subseteq A_\lambda$ for some $\lambda \in J$, choose one such A_λ. Let the collection of A_λ chosen in this manner be indexed by $\lambda \in J'$, say.

Thus the set J' corresponds to a subset of the set $\overset{1}{\mathbb{Q}} \times \cdots \times \overset{p}{\mathbb{Q}} \times \mathbb{Q}$, which is countable by Reed Proposition 1.3.4 applied repeatedly (products of countable sets are countable). Since a subset of a countable set is either finite or countable (Reed page 16 Proposition 1.3.2), it follows that J' is either finite or countable.

We next show that the collection of sets $\{ A_\lambda \mid \lambda \in J' \}$ is a cover of A.

To see this consider any $\mathbf{x}^* \in A$. Then $\mathbf{x}^* \in A_{\lambda^*}$ for some $\lambda^* \in J$, and so $B_{r^*}(\mathbf{x}^*) \subseteq A^*_\lambda$ for some real number $r^* > 0$ since A_{λ^*} is open.



It is clear from the diagram that we can choose $\mathbf{x}$ with *rational* components close to $\mathbf{x}^*$, and then choose a suitable *rational* $r > 0$, such that

$$\mathbf{x}^* \in B_r(\mathbf{x}) \quad \text{and} \quad B_r(\mathbf{x}) \subseteq B_{r^*}(\mathbf{x}^*).$$

(The precise argument uses the triangle inequality.[40]) In particular, $B_r(\mathbf{x}) \subseteq A_{\lambda^*}$ and so $B_r(\mathbf{x})$ must be one of the balls used in constructing J'. In other words, $B_r(\mathbf{x}) \subseteq A_\lambda$ for some $\lambda \in J'$.

The collection of all $B_r(\mathbf{x})$ obtained in this manner must be a cover of A (because every $\mathbf{x}^* \in A$ is in at least one such $B_r(\mathbf{x})$). It follows that the collection $\{ A_\lambda \mid \lambda \in J' \}$ is also a cover of A.

But we have seen that the set $\{ A_\lambda \mid \lambda \in J' \}$ is finite or countable, and so we have proved the claim.

We next *claim* that there is a *finite* subcover of any *countable* cover.

To see this, write the countable cover as

$$A \subseteq A_1 \cup A_2 \cup A_3 \cup \dots . \tag{20}$$

If there is *no* finite subcover, then we can choose

$$\begin{aligned} &\mathbf{a}_1 \in A \setminus A_1 \\ &\mathbf{a}_2 \in A \setminus (A_1 \cup A_2) \\ &\mathbf{a}_3 \in A \setminus (A_1 \cup A_2 \cup A_3) \\ &\quad \vdots \end{aligned}$$

By sequential compactness of A there is a subsequence $(\mathbf{a}_{n'})$ which converges to some $\mathbf{a} \in A$. From (20), $\mathbf{a} \in A_k$ for some k. Because A_k is open, $\mathbf{a}_{n'} \in A_k$ for all sufficiently large n'. But from the construction of the sequence $(\mathbf{a}_n)$ $\mathbf{a}_n \notin A_k$ if $n \geq k$. This is a contradiction.

Hence there is a finite subcover in (20), and so we have proved the claim.

---

[40] For example, choose $\mathbf{x}$ with rational components so $d(\mathbf{x}, \mathbf{x}^*) < r^*/4$ and then choose rational r so $r^*/4 < r < r^*/2$.

Then $\mathbf{x}^* \in B_r(\mathbf{x})$ since $d(\mathbf{x}, \mathbf{x}^*) < r^*/4 < r$.

Moreover, if $\mathbf{y} \in B_r(\mathbf{x})$ then

$$d(\mathbf{x}^*, \mathbf{y}) < d(\mathbf{x}^*, \mathbf{x}) + d(\mathbf{x}, \mathbf{y}) < r^*/4 + r < r^*/4 + r^*/2 < r^*.$$

Hence $\mathbf{y} \in B_{r^*}(\mathbf{x}^*)$ and so $B_r(\mathbf{x}) \subseteq B_{r^*}(\mathbf{x}^*)$.

Putting the two claims together, we see that there is a finite subcover of any cover of A, and so A is compact.

This proves one direction in the theorem.

To prove the other direction, suppose A is compact.

Assume A is not sequentially compact.

This means we can choose a sequence $(\mathbf{a}_n) \subseteq A$ that has no subsequence converging to any point in A. We saw in the previous theorem that A is closed. This then implies that in fact $(\mathbf{a}_n)$ has no subsequence converging to any point in $\mathbb{R}$. Let the set of values taken by $(\mathbf{a}_n)$ be denoted

$$S = \{\, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots \}.$$

(Some values may be listed more than once.)

Claim: S is infinite. To see this, suppose S is finite. Then at least one value $\mathbf{a}$ would be taken an infinite number of times. There is then a constant subsequence all of whose terms equal $\mathbf{a}$ and which in particular converges to $\mathbf{a}$. This contradicts the above and so establishes the claim.

Claim: S is closed in $\mathbb{R}^p$. If S is *not* closed then from Theorem 11.2.1 there is a sequence $(\mathbf{b}_n)$ from S which converges to $\mathbf{b} \notin S$. Choose a subsequence $(\mathbf{b}_{n'})$ of $(\mathbf{b}_n)$ so that $\mathbf{b}_{1'} = \mathbf{b}_1$, so that $\mathbf{b}_{2'}$ occurs further along in the sequence $(\mathbf{a}_n)$ than does $\mathbf{b}_{1'}$, so that $\mathbf{b}_{3'}$ occurs further along in the sequence $(\mathbf{a}_n)$ than does $\mathbf{b}_{2'}$, etc. Then the sequence $(\mathbf{b}_{n'})$ is a subsequence of $(\mathbf{a}_n)$, and it converges to $\mathbf{b}$. This contradicts what we said before about $(\mathbf{a}_n)$. Hence S *is* closed in $\mathbb{R}^p$ as claimed.

A similar argument shows that for each $\mathbf{a} \in S$ there is a neighbourhood $U(\mathbf{a}) = B_r(\mathbf{a})$ (for some r depending on $\mathbf{a}$) such that the only point in $S \cap U(\mathbf{a})$ is $\mathbf{a}$. (Otherwise we could construct in a similar manner to the previous paragraph a subsequence of $(\mathbf{a}_n)$ which converges to $\mathbf{a}$.)

Next consider the following open cover of A (in fact of all of $\mathbb{R}^p$):

$$S^c,\ U(\mathbf{a}_1),\ U(\mathbf{a}_2),\ U(\mathbf{a}_3),\ \dots\ .$$

By compactness there is a finite subcover of A, and by adding more sets if necessary we can take a finite subcover of the form

$$S^c,\ U(\mathbf{a}_1),\ \dots,\ U(\mathbf{a}_N)$$

for some N.

But this is impossible, as we see by choosing $\mathbf{a} \in S\ (\subseteq A)$ with $\mathbf{a} \neq \mathbf{a}_1, \dots, \mathbf{a}_N$ (remember that S is infinite) and recalling that $S \cap U(\mathbf{a}_i) = \{\mathbf{a}_i\}$. In particular, $\mathbf{a} \notin S^c$ and $\mathbf{a} \notin U(\mathbf{a}_i)$ for $i = 1, \dots, N$.

Thus the assumption is false and so we have proved the theorem, □

12.4. More remarks. We have seen that $A \subseteq \mathbb{R}^p$ is closed and bounded iff it is sequentially compact iff it is compact.

Remark 12.4.1.***** In an arbitrary metric space the second "iff" is still true, with a "similar" argument to that of Theorem 12.3.1. The notion of points with rational coordinates is replaced by the notion of a "countable dense subset" of A.

In an arbitrary metric space sequentially compact implies closed and bounded (with essentially the same proof as in Theorem 12.1.2), but closed and bounded does not imply sequentially compact (the proof of Theorem 11.4.2 uses components and does not generalise to arbitrary metric spaces). The general result is that complete and *totally* bounded (these are the same as closed and bounded in $\mathbb{R}^p$) is equivalent to sequentially compact is equivalent to compact.

13. Continuous functions on $\mathbb{R}^p$

Review Section 7 and Reed page 152 line 3- the end of paragraph 2 on page 154.

We consider functions $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^p$.

(Much of what we say will generalise to $f : A \to \mathbb{R}$ where $A \subseteq X$ and (X, ρ) is a general metric space. It also mostly further generalises to replacing $\mathbb{R}$ by Y where (Y, ρ^*) is another metric space.)

13.1. Basic results. Exactly as for $p = 1$, we have:

DEFINITION 13.1.1. A function $f : A \to \mathbb{R}$ is *continuous at* $\mathbf{a} \in A$ if

$$\mathbf{x}_n \to \mathbf{a} \quad \Longrightarrow \quad f(\mathbf{x}_n) \to f(\mathbf{a}) \tag{21}$$

whenever $(\mathbf{x}_n) \subseteq A$.

THEOREM 13.1.2. *A function $f : A \to \mathbb{R}$ is* continuous *at $\mathbf{a} \in A$ iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that:*

$$\mathbf{x} \in A \text{ and } d(\mathbf{x}, \mathbf{a}) \leq \delta \quad \Longrightarrow \quad d(f(\mathbf{x}), f(\mathbf{a})) \leq \epsilon. \tag{22}$$

(The proof is the same as the one dimensional case, see Reed Theorem 3.1.3 page 77.)

The usual properties of sums, products, quotients (when defined) and compositions of continuous functions being continuous, still hold and have the same proofs.

DEFINITION 13.1.3. A function $f : A \to \mathbb{R}$ is *uniformly continuous* iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that:

$$\mathbf{x} \in A \text{ and } d(\mathbf{x}_1, \mathbf{x}_2) \leq \delta \quad \Longrightarrow \quad d(f(\mathbf{x}), f(\mathbf{a})) \leq \epsilon. \tag{23}$$

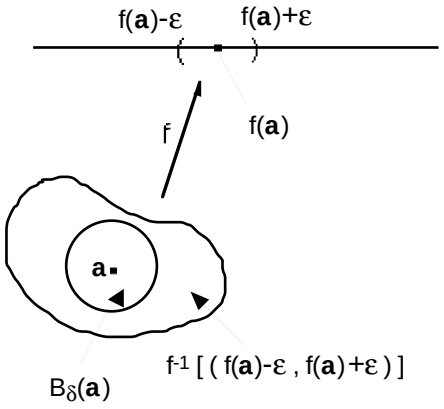
This is also completely analogous to the case $p = 1$.

REMARK 13.1.1. In the previous Theorem and Definition we can replace either or both "$\leq$" by "$<$". This follows essentially from the fact that we require the statements to be true for *every* $\varepsilon > 0$. This is frequently very convenient and we will usually do so without further comment. See also Adams §1.5, Definition 1.9.

REMARK 13.1.2. We can think of (22) geometrically as saying either (using $<$ instead of $\leq$)

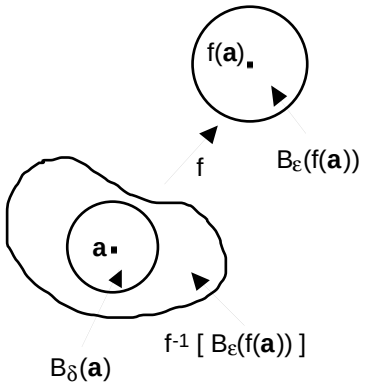
1. $f[A \cap B_\delta(\mathbf{a})] \subseteq (f(\mathbf{a}) - \varepsilon, f(\mathbf{a}) + \varepsilon)$, *or*
2. $A \cap B_\delta(\mathbf{a}) \subseteq f^{-1}[\,(f(\mathbf{a}) - \varepsilon, f(\mathbf{a}) + \varepsilon)\,]$.

See Section 13.4 for the notation, although the following diagram in case $A = \mathbb{R}^p$ should make the idea clear.

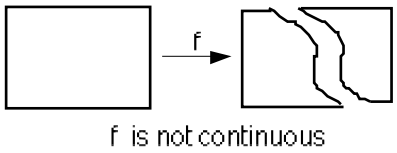


REMARK 13.1.3. Definitions 13.1.1, 13.1.3 and Theorem 13.1.2 generalise immediately, without even changing notation, to functions $f : A \to \mathbb{R}^q$. We can again replace either or both "$\leq$" by "$<$".

The previous diagram then becomes:



A function f which has "rips" or "tears" are not continuous.



If $f : A \to \mathbb{R}^q$ we can write

$$f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_q(\mathbf{x})).$$

It follows from Definition 13.1.1 (or Theorem 13.1.2) applied to functions with values in $\mathbb{R}^q$, that f is continuous iff each component function $f_1, \ldots, f_q$ is continuous.

For example, if

$$f : \mathbb{R}^2 \to \mathbb{R}^3, \quad f(x, y) = (x^2 + y^2, \sin xy, e^x),$$

then f is continuous since the component functions

$$f_1(x, y) = x^2 + y^2, \quad f_2(x, y) = \sin xy, \quad f_3(x, y) = e^x,$$

are all continuous.

In this sense we can always consider real valued functions instead of functions into $\mathbb{R}^q$. But this is often not very useful — in linear algebra, for example, we usually want to think of linear transformations as maps into $\mathbb{R}^q$ rather than as q component maps into $\mathbb{R}$. And it does not help if we want to consider functions which map into a general metric space.

13.2. Deeper properties of continuous functions.

THEOREM 13.2.1. *Suppose $f : A \to \mathbb{R}$ is continuous where $A \subseteq \mathbb{R}^p$ is closed and bounded. Then f is bounded above and below and has a maximum and a minimum value. Moreover, f is uniformly continuous of A.*

This generalises the result for functions defined on a closed bounded interval. The proof is essentially the same as in the case $p = 1$. See Reed Theorem 4.6.1 page 153.

The theorem is also true for $A \subseteq X$ where X is an arbitrary metric space, provided A is sequentially compact (which is the same as compact, as we have proved). The proof is essentially the same as in the case of $X = \mathbb{R}$ in Section 3.2 of Reed. The Bolzano-Wierstrass theorem is not needed since we know immediately

from the definition of sequential compactness that every sequence from A has a convergent subsequence.

***** There is also a generalisation of the Intermediate Value Theorem, but for this we need the set A to be "connected". One definition of *connected* is that we cannot write

$$A = (E \cap A) \cup (F \cap A)$$

where E, F are open and $E \cap A, F \cap A$ are nonempty. If f is continuous and A is connected, then one can prove that the image of f is an interval.

13.3. Limits. As for the case $p = 1$, limits in general can be defined either in terms of sequences or in terms of ε and δ. See Definition 13.3.1 and Theorem 13.3.2.

We say $\mathbf{a}$ is a *limit point* of A if there is a sequence from $A \setminus \{\mathbf{a}\}$ which converges to $\mathbf{a}$. We do not require $\mathbf{a} \in A$. From Theorem 11.2.1, a set contains all its limit points iff it is closed.

We say $\mathbf{a} \in A$ is an *isolated point* of A if there is no sequence $\mathbf{x}_n \to \mathbf{a}$ such that $(\mathbf{x}_n) \subseteq A \setminus \{\mathbf{a}\}$.

For example, if $A = (0, 1] \cup 2 \subseteq \mathbb{R}$ then 2 is an isolated point of A (and is the only isolated point of A). The limit points of A are given by the set $[0, 1]$.

The following Definition, Theorem and its proof, are completely analogous to the case $p = 1$.

Definition 13.3.1. Suppose $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^p$ and $\mathbf{a}$ is a limit point of A. If for any sequence $(\mathbf{x}_n)$:

$$(\mathbf{x}_n) \subseteq A \setminus \{\mathbf{a}\} \text{ and } \mathbf{x}_n \to \mathbf{a} \quad \Rightarrow \quad f(\mathbf{x}_n) \to L,$$

then we say *the limit of* $f(\mathbf{x})$ *as* $\mathbf{x}$ *approaches* $\mathbf{a}$ *is* L, or *the limit of* f *at* $\mathbf{a}$ *is* L, and write

$$\lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in A}} f(\mathbf{x}) = L, \quad \text{or} \quad \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L.$$

The reason for restricting to sequences, none of whose terms equal $\mathbf{a}$, is the usual one. For example, if

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} x^2 & x \neq 0 \\ 1 & x = 0 \end{cases}$$

or

$$g : \mathbb{R} \setminus \{0\} \to \mathbb{R}, \quad g(x) = x^2,$$

then we want

$$\lim_{x\to 0} f(x) = 0, \quad \lim_{x\to 0} g(x) = 0.$$

Theorem 13.3.2. *Suppose* $f : A \to \mathbb{R}$, *where* $A \subseteq \mathbb{R}^p$ *and* $\mathbf{a}$ *is a limit point of* A. *Then* $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$ *iff* [41]*:*

for every $\varepsilon > 0$ *there is a corresponding* $\delta > 0$ *such that*

$$d(\mathbf{x}, \mathbf{a}) \leq \delta \text{ and } \mathbf{x} \neq \mathbf{a} \quad \Rightarrow \quad |f(\mathbf{x}) - L| \leq \epsilon.$$

It follows, exactly as in the case $p \neq 1$, that if $\mathbf{a} \in \mathcal{D}(f)$ is a limit point of $\mathcal{D}(f)$) then f is continuous at $\mathbf{a}$ iff

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

(If $\mathbf{a}$ is an isolated point of $\mathcal{D}(f)$ then f is always continuous at $\mathbf{a}$ according to the definition — although $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ is not actually defined. This is not an

[41] As usual, we could replace either or both "$\leq$ " by "$<$ "

interesting sitation, and we would not normally consider continuity at isolated points.)

The usual properties of sums, products and quotients of limits hold, with the same proofs as in the case $p = 1$.

EXAMPLE 13.3.3. Functions of two or more variables exhibit more complicated behaviour than functions of one variable. For example, let

$$f(x, y) = \frac{xy}{x^2 + y^2}$$

for $(x, y) \neq (0, 0)$. (Setting $x = r\cos\theta$, $y = r\sin\theta$, we see that in polar coordinates this is just the function $f(r, \theta) = \frac{1}{2}\sin 2\theta$.)

If $y = ax$ then $f(x, y) = a(1 + a^2)^{-1}$ for $x \neq 0$. Hence

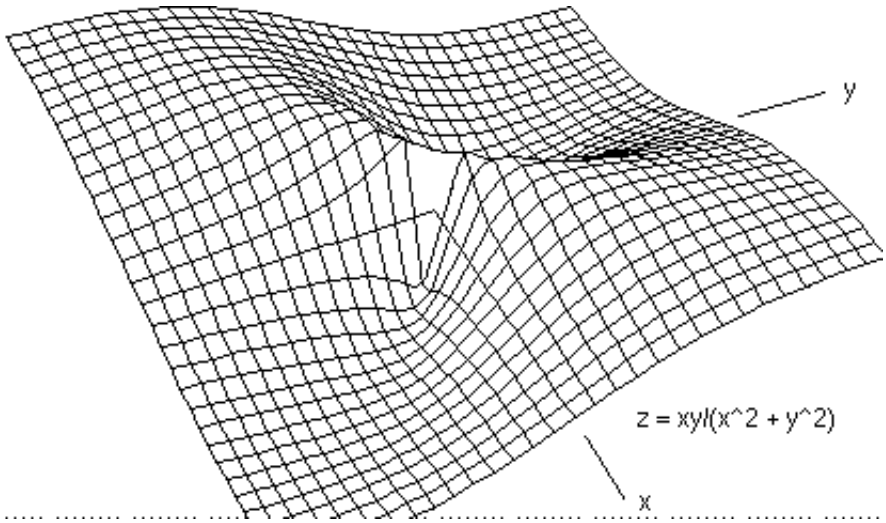
$$\lim_{\substack{(x,y)\to a \\ y=ax}} f(x, y) = \frac{a}{1 + a^2}.$$

Thus we obtain a different limit of f as $(x, y) \to (0, 0)$ along different lines. It follows that

$$\lim_{(x,y)\to(0,0)} f(x, y)$$

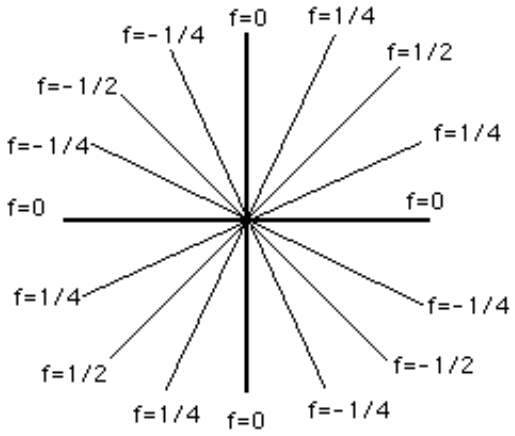
does not exist.

A partial diagram of the graph of f is:



Probably a better way to visualize f is by sketching *level sets*[42] of f as shown in the next diagram. Then you can visualise the graph of f as being swept out by a straight line rotating around the origin at a height as indicated by the level sets. This may also help in understanding the previous diagram.

[42] A level set of f is a set on which f is constant.



EXAMPLE 13.3.4. Let

$$f(x,y) = \frac{x^2 y}{x^4 + y^2}$$

for $(x,y) \neq (0,0)$.

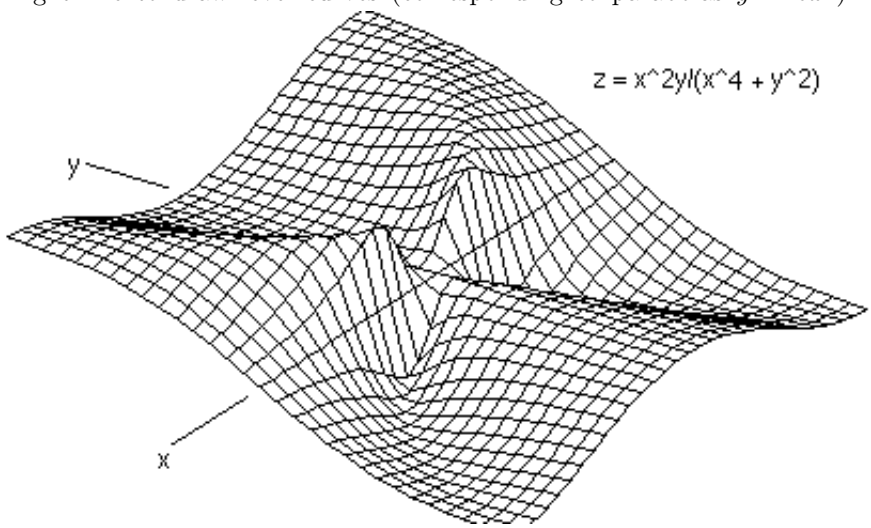
Then

$$\begin{aligned} \lim_{\substack{(x,y)\to(0,0)\\ y=ax}} f(x,y) &= \lim_{x\to 0} \frac{ax^3}{x^4 + a^2x^2} \\ &= \lim_{x\to 0} \frac{ax}{x^2 + a^2} \\ &= 0. \end{aligned}$$

Thus the limit of f as $(x,y) \to (0,0)$ along *any* line $y = ax$ is 0. The limit along the y-axis $x = 0$ is also easily seen to be 0.

But it is still not true that $\lim_{(x,y)\to(0,0)} f(x,y)$ exists. For if we consider the limit of f as $(x,y) \to (0,0)$ along the parabola $y = bx^2$ we see that $f = b(1+b^2)^{-1}$ on this curve and so the limit is $b(1+b^2)^{-1}$.

You might like to draw level curves (corresponding to parabolas $y = bx^2$).



13.4. Another characterisation of continuity. One can define continuity in terms of open sets (or closed sets) without using either sequences or the ε–δ definition. See Theorem 13.4.3.

But for this we first need to introduce some general notation for functions.

DEFINITION 13.4.1. Suppose $f : X \to Y$. The *image* of $E \subseteq X$ under f is the set

$$f[E] = \{\, y \in Y \mid y = f(x) \text{ for some } x \in E \,\}.$$

The *inverse image* of $E \subseteq Y$ under f is the set

$$f^{-1}[E] = \{\, x \in X \mid f(x) \in E \,\}.$$

It is important to realise that $f^{-1}[E]$ is always defined for $E \subseteq Y$, even if f does not have an inverse. In fact the inverse function will exist iff $f^{-1}\{y\}$[43] contains at most one element for each $y \in Y$.

The following are straightforward to check. Note that inverse images are better behaved than images. The results generalise immediately to intersections and unions of more than two, including infinitely many, sets.

THEOREM 13.4.2. *Suppose $f : X \to Y$ and $E, E_1, E_2 \subseteq Y$. Then*

$$\begin{aligned} f^{-1}[E_1 \cap E_2] &= f^{-1}[E_1] \cap f^{-1}[E_2], \\ f^{-1}[E_1 \cup E_2] &= f^{-1}[E_1] \cup f^{-1}[E_2], \\ f^{-1}[E_1 \setminus E_2] &= f^{-1}[E_1] \setminus f^{-1}[E_2], \\ E_1 \subseteq E_2 \Rightarrow f^{-1}[E_1] &\subseteq f^{-1}[E_2] \\ f[f^{-1}[E]\,] &\subseteq E. \end{aligned}$$

If $E, E_1, E_2 \subseteq X$, then

$$\begin{aligned} f[E_1 \cap E_2] &\subseteq f[E_1] \cap f[E_2], \\ f[E_1 \cup E_2] &= f[E_1] \cup f[E_2], \\ f[E_1 \setminus E_2] &\supseteq f[E_1] \setminus f[E_2] \\ E_1 \subseteq E_2 \Rightarrow f[E_1] &\subseteq f[E_2] \\ E &\subseteq f^{-1}[f[E]\,]. \end{aligned}$$

PROOF. For the first, $x \in f^{-1}[E_1 \cap E_2]$ iff $f(x) \in E_1 \cap E_2$ iff ($f(x) \in E_1$ and $f(x) \in E_2$) iff ($x \in f^{-1}[E_1]$ and $x \in f^{-1}[E_2]$) iff $x \in f^{-1}[E_1] \cap f^{-1}[E_2]$. The others are similar. □

To see that equality need not hold in the sixth assertion let $f(x) = x^2$, $X = Y = \mathbb{R}$, $E_2 = (-\infty, 0]$ and $E_1 = [0, \infty)$. Then $f[E_1 \cap E_2] = \{0\}$ and $f[E_1] \cap f[E_2] = [0, \infty)$. Similarly, $f[E_1 \setminus E_2] = (0, \infty)$ while $f[E_1] \setminus f[E_2] = \emptyset$. Also, $f^{-1}[f[E_1]] = \mathbb{R} \supsetneq E_1$ and $f[f^{-1}[\mathbb{R}]] = [0, \infty) \neq \mathbb{R}$.

You should now go back and look again at Remarks 13.1.2 and 13.1.3 and the two diagrams there.

We first state and prove the following theorem for functions defined on all of $\mathbb{R}^p$. We then discuss the generalisation to $f : A\ (\subseteq \mathbb{R}^p) \to B\ (\subseteq \mathbb{R})$.

THEOREM 13.4.3. *Let $f : \mathbb{R}^p \to \mathbb{R}$. Then the following are equivalent:*

1. *f is continuous;*
2. *$f^{-1}[E]$ is open in $\mathbb{R}^p$ whenever E is open in $\mathbb{R}$;*
3. *$f^{-1}[C]$ is closed in $\mathbb{R}^p$ whenever C is closed in $\mathbb{R}$.*

[43] We sometimes write $f^{-1}E$ for $f^{-1}[E]$.

PROOF.

(1) $\Rightarrow$ (2): Assume (1). Let E be open in $\mathbb{R}$. We wish to show that $f^{-1}[E]$ is open (in $\mathbb{R}^p$).

Let $\mathbf{x} \in f^{-1}[E]$. Then $f(\mathbf{x}) \in E$, and since E is open there exists $r > 0$ such that $(f(\mathbf{x}) - r, f(\mathbf{x}) + r) \subseteq E$.

From Remark 13.1.2 there exists $\delta > 0$ such that

$$B_\delta(\mathbf{x}) \subseteq f^{-1}(f(\mathbf{x}) - r, f(\mathbf{x}) + r).$$

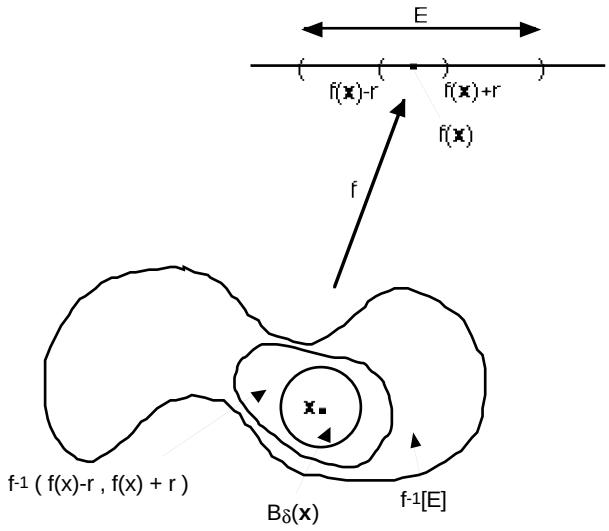
But

$$f^{-1}(f(\mathbf{x}) - r, f(\mathbf{x}) + r) \subseteq f^{-1}[E],$$

and so

$$B_\delta(\mathbf{x}) \subseteq f^{-1}[E].$$

Thus for every point $\mathbf{x} \in f^{-1}[E]$ there is a $\delta = \delta_{\mathbf{x}} > 0$ such that the above line is true, and so $f^{-1}[E]$ is open.



(2) $\Leftrightarrow$ (3): Assume (2), i.e. $f^{-1}[E]$ is open in $\mathbb{R}^p$ whenever E is open in $\mathbb{R}$. If C is closed in $\mathbb{R}$ then C^c is open and so $f^{-1}[C^c]$ is open. But $(f^{-1}[C])^c = f^{-1}[C^c]$. Hence $f^{-1}[C]$ is closed.

We can similarly show (3) $\Rightarrow$ (2).

(2) $\Rightarrow$ (1): Assume (2).

Let $\mathbf{x} \in \mathbb{R}^p$. In order to prove f is continuous at $\mathbf{x}$ take any $r > 0$.

Since $(f(\mathbf{x}) - r, f(\mathbf{x}) + r)$ is open it follows that $f^{-1}(f(\mathbf{x}) - r, f(\mathbf{x}) + r)$ is open.

Since $\mathbf{x} \in f^{-1}(f(\mathbf{x}) - r, f(\mathbf{x}) + r)$ and this set is open, it follows there exists $\delta > 0$ such that $B_\delta(\mathbf{x}) \subseteq f^{-1}(f(\mathbf{x}) - r, f(\mathbf{x}) + r)$.

It now follows from Remark 13.1.2 that f is continuous at $\mathbf{x}$.

Since $\mathbf{x}$ was an arbitrary point in $\mathbb{R}^p$, this finishes the proof. □

EXAMPLE 13.4.4. We can now give a simple proof that the examples at the beginning of Section 10.5 are indeed open. For example, if

$$A = \{\, (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 < z^2 \text{ and } \sin x < \cos y \,\}$$

then $A = A_1 \cap A_2$ where

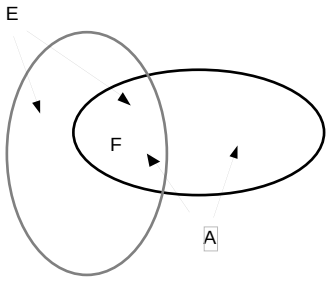
$$A_1 = f_1^{-1}(-\infty, 0), \quad f_1(x, y, z) = x^2 + y^2 - z^2,$$
$$A_2 = f_2^{-1}(-\infty, 0), \quad f_2(x, y, z) = \sin x - \cos y.$$

Thus A_1 and A_2 are open, being the inverse image of open sets, and hence their intersection is open.

REMARK 13.4.1.* In order to extend the previous characterisation of continuity to functions defined just on a subset of $\mathbb{R}^p$, we need the following definitions (which are also important in other settings).

Suppose $A \subseteq \mathbb{R}^p$. We say $F \subseteq A$ is *open (closed) in A* or *relatively open (closed) in A* if there is an open (closed) set E in $\mathbb{R}^p$ such that

$$F = A \cap E.$$



E does not include its boundary, A does.
F is relatively open in A, but is not open in $\mathbf{R}^2$.

It follows (*exercise*) that $F \subseteq A$ is open in A iff for each $\mathbf{x} \in F$ there is an $r > 0$ such that

$$A \cap B_r(\mathbf{x}) \subseteq F.$$

EXAMPLE 13.4.5. Consider the interval $A = (0, 1]$. Then the set $F = (\frac{1}{2}, 1]$ is open in A, as is A itself. The set $F = (0, \frac{1}{2}]$ is closed in A, as is A itself. (*What could we take as the set E in each case?*)

For another example, consider the square

$$F = \{\, (x, y) \mid 0 \leq x \leq 1,\ 0 < y < 1 \,\},$$

which does not contain its vertical sides but does contain the remainder of its horizontal sides. Let

$$A = \{\, (x, y) \mid 0 \leq x \leq 1 \,\}, \quad B = \{\, (x, y) \mid 0 < y < 1 \,\},$$

Then A is closed and B is open. Since $F = A \cap B$ it follows that F is open in A and closed in B.

The interval $(0, 1)$ is open in $\mathbb{R}$. But it is not open when considered as a subset of $\mathbb{R}^2$ via the usual inclusion of $\mathbb{R}$ into $\mathbb{R}^2$.

Theorem 13.4.3 is true for functions $f : A\ (\subseteq \mathbb{R}^p) \to \mathbb{R}$ if we replace "in $\mathbb{R}^p$ by "in A".

For (1) $\Rightarrow$ (2) we suppose E is open. We need to show that $f^{-1}[E]$ is open in A. Just replace $B_\delta(\mathbf{x})$ by $B_\delta(\mathbf{x}) \cap A$ and $B_{\delta_\mathbf{x}}(\mathbf{x})$ by $B_{\delta_\mathbf{x}}(\mathbf{x}) \cap A$ throughout the proof.

The proof of (2) $\Leftrightarrow$ (3) is similar, by taking complements in A and noting that a set is open in A iff its complement in A is closed in A. (*Why?*)

For (2) $\Rightarrow$ (1) we suppose the inverse image of an open set in $\mathbb{R}$ is open in A. In the proof, replace $B_\delta(\mathbf{x})$ by $A \cap B_\delta(\mathbf{x})$ (and use the *exercise* at the beginning of this Remark).

Finally, we observe that the theorem also generalises easily to functions $f : A\ (\subseteq \mathbb{R}^p)\mathbb{R}^q$. Just replace the interval $(f(\mathbf{x}) - r, f(\mathbf{x}) + r)$ by the ball $B_r(f(\mathbf{x}))$ throughout this Remark.

REMARK 13.4.2. It is not true that the continuous image of an open (closed) set need be open (closed).

For example, let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Then $f(-1, 1) = [0, 1)$ which is neither open nor closed.

If $f(x) = e^{-x^2}$ then $f[\mathbb{R}] = (0, 1]$ which again is neither open nor closed.

13.5. Continuous image of compact sets*. The inverse image of a compact set under a continuous function need not be compact.

For example, if $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = e^{-x^2}$ then $f^{-1}[0, 1] = \mathbb{R}$.

However, continuous images of compact sets are compact.

THEOREM 13.5.1. *Suppose $f : E\ (\subseteq \mathbb{R}^p) \to \mathbb{R}^q$ is continuous and E is compact. Then $f[E]$ is compact.*

PROOF. Suppose

$$f[E] \subseteq \bigcup_{\lambda \in J} A_\lambda$$

is an open cover. Then

$$E \subseteq f^{-1}[f[E]\,] \subseteq f^{-1}\Big[\bigcup_{\lambda \in J} A_\lambda\Big] = \bigcup_{\lambda \in J} f^{-1}[A_\lambda].$$

This gives a cover of E by relatively open sets, and by compactness[44] there is a finite subcover

$$E \subseteq \bigcup_{\lambda \in J_0} f^{-1}[A_\lambda].$$

It follows from Theorem 13.4.2 that

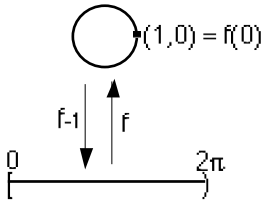
$$f[E] \subseteq f\Big[\bigcup_{\lambda \in J_0} f^{-1}[A_\lambda]\Big] = \bigcup_{\lambda \in J_0} f[f^{-1}[A_\lambda]\,] \subseteq \bigcup_{\lambda \in J_0} A_\lambda.$$

Thus the original cover of $f[E]$ has a finite subcover.

Hence $f[E]$ is compact. □

It is not always true that if f is continuous and one-to-one then its inverse is continuous. For example, let

$$f : [0, 2\pi) \to \{\,(x, y) \mid x^2 + y^2 = 1\,\}\ (\subseteq \mathbb{R}^2), \quad f(t) = (\cos t, \sin t).$$



Then the inverse function is given by

$$f^{-1} : \{\,(x, y) \mid x^2 + y^2 = 1\,\} \to [0, 2\pi), \quad f(\cos t, \sin t) = t.$$

But this is not continuous, as we see by considering what happens near $(1, 0)$.

For example,

$$\left(\cos 2\pi - \frac{1}{n}, \sin 2\pi - \frac{1}{n}\right) \to (1, 0),$$

[44]Each A_λ can be extended to an open set in $\mathbb{R}^p$. By compactness there is a finite subcover from these extended sets. Taking their restriction to E gives a finite subcover of the original cover.

but

$$f^{-1}\left(\cos 2\pi - \frac{1}{n}, \sin 2\pi - \frac{1}{n}\right) = 2\pi - \frac{1}{n} \not\to f^{-1}(1,0) = 0.$$

However,

THEOREM 13.5.2. *Suppose $f : E \to \mathbb{R}^q$ is one-to-one and continuous, and E is compact. Then the inverse function f^{-1} is also continuous (on its domain $f[E]$).*

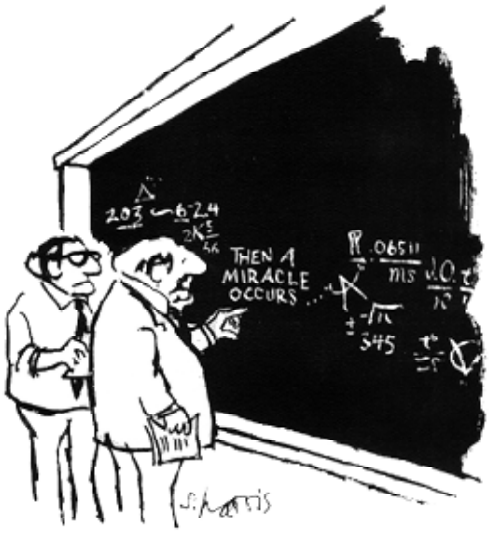
PROOF. Since f is one-to-one and onto its image, then its inverse function $f^{-1} : f[E] \to E$ is certainly well-defined (namely, $f^{-1}(\mathbf{y}) = \mathbf{x}$ iff $f(\mathbf{x}) = \mathbf{y}$).

To show that f^{-1} is continuous we need to show that the inverse image under f^{-1}, of any set $C \subseteq E$ which is closed in E, is closed in $f[E]$. Since f is one-to-one and onto, this inverse image $(f^{-1})^{-1}[C]$ under f^{-1} is the same as the image $f[C]$ under f.

Since C is a relatively closed subset of a compact set, it is itself compact. (C is the intersection of the closed bounded set E with a closed set, and so is itself closed and bounded, hence compact.)[45] Hence $f[C]$ is compact by the previous theorem, and in particular is closed in $\mathbb{R}^q$. It follows that $f[C]$ is closed in $f[E]$. □

[45]On can also prove this directly from the definition of sequential compactness or from the definition of compactness via open covers. Thus this result generalises to any metric space or even any topological space.

14. Sequences and series of functions.



"I think you should be more explicit here in step two."

The reference for this Chapter is Reed Chapter 5, §1–3, and Example 2 page 197; and Chapter 6, §6.3

14.1. Pointwise and uniform convergence. Study Reed §5.1 carefully; the definitions and all the examples.

14.2. Properties of uniform convergence. Study Reed §5.2 carefully.

The point is that continuity and integration behave well under uniform convergence, but not under pointwise convergence.

Differentiation does not behave well, unless we also require the derivatives to converge uniformly.

Theorem 5.2.4, concerning differentiation under the integral sign, is also important.

14.3. The supremum norm and supremum metric. Study Reed §5.3 and Example 2 page 197 carefully.

One way of measuring the "size" of a bounded function f is by means of the *supremum norm* $\|f\|_\infty$ (Reed page 175). We can then measure the *distance* between two functions f and g (with the same domain E) by using the *sup distance* (or *sup metric*)

$$\rho_\infty(f,g) = \|f-g\|_\infty = \sup_{x\in E} |f(x)-g(x)|.$$

In Reed, $E \subseteq \mathbb{R}$, but exactly the same definition applies for $E \subseteq \mathbb{R}^n$.

In Proposition 5.3.1 it is shown that the sup norm satisfies positivity, the triangle inequality, and is well behaved under scalar multiples.

In Example 2 on page 197 it is shown that as a consequence the sup metric is indeed a metric (i.e. satisfies positivity, symmetry and transitivity — see the first paragraph in Section 10.2).

(See Reed page 177) A sequence of functions (f_n) defined on the same domain E *converges to* f *in the sup metric*[46] if $\rho_\infty(f_n, f) \to 0$ as $n \to \infty$. (Note that this is completely analogous to Definition 11.1.1 for convergence of a sequence of points in $\mathbb{R}^n$.)

One similarly defines the notion of *Cauchy in the sup metric* (or equivalently *Cauchy in the sup norm*).

One proves (page 178) that if a sequence (f_n) of continuous functions defined on a closed bounded interval $[a, b]$ is Cauchy in the sup metric, then it converges in the sup metric and the limit is also continuous on $[a, b]$. (This is analogous to the fact that if a sequence from $\mathbb{R}^n$ is Cauchy then it converges to a point in $\mathbb{R}^n$.)

The same theorem and proof works if $[a, b]$ is replaced by any set $E \subseteq \mathbb{R}^n$, but then we need to restrict to functions that are continuous and *bounded* (otherwise the definition of $\rho_\infty(f, g)$ may give ∞). (If E is compact we need not assume boundedness, as this follows from continuity by Theorem 13.2.1.)

We say that the set of bounded and continuous functions defined on E is *complete in the sup metric (or norm).*

14.4. Integral norms and metrics. Study Reed pp 179–181, particularly Example 1. We cannot do much on this topic, other than introduce it.

The important "integral" norms for functions f with domain E are defined by

$$\|f\|_p = \left(\int_E |f|^p \right)^{1/p}.$$

for any $1 \leq p < \infty$.

The most important are $p = 2$ and $p = 1$ (in that order). If E is not an interval in $\mathbb{R}$ then we need a more general notion of integration; to do it properly requires the Lebesgue integral. But $\int_E g$ is still interpreted as the area between the graph of f and the set $E \subseteq \mathbb{R}$ — being negative where it lies below the axis. In higher dimensional cases we replace "area" by "volume" of the appropriate dimension.)

It is possible to prove for $\|f\|_p$ the three properties of a "norm" in Proposition 5.3.1 of Reed.

Actually, $\|f\|_p$ does not quite give a norm. The problem is that if f is not continuous then $\|f\|_p$ may equal 0 even though f is not the zero function, i.e. not everywhere zero. However f must then be zero "almost everywhere" — in a sense that can be made precise. This problem does not arise if we restrict to continuous functions, since if f is continuous and $\int_a^b |f| = 0$ then $f = 0$ everywhere. But it is usually not desirable to restrict to continuous functions for other reasons.

The metrics corresponding to these norms are

$$\rho_p(f, g) = \|f - p\|_p,$$

again for $1 \leq p < \infty$.

REMARK 14.4.1. ***** The set S of continuous functions defined on a closed bounded set is a metric space in any of the metrics ρ_p, but it is not a *complete* metric space (unless $p = \infty$). If we take the "completion" of the set S we are lead to the so-called Lebesgue measurable functions and the theory of Lebesgue integration.

[46] Reed says "converges in the sup norm".

14.5. Series of functions. Study Reed §4.3 carefully.

Suppose the functions f_j all have a common domain (typically $[a, b]$).

The infinite *series* of functions $\sum_{j=1}^{\infty} f_j$ is said to *converge pointwise* to the function f iff the corresponding series of real numbers $\sum_{j=1}^{\infty} f_j(x)$ converges to $f(x)$ for each x in the (common) domain of the f_j.

From the definition of convergence of a series of real numbers, this means that the infinite *sequence* (S_n) of partial sums

$$S_n = f_1 + \cdots + f_n$$

converges in the pointwise sense to f.

We say that the series $\sum_j f$ *converges uniformly* if the sequence of partial sums converges uniformly.

Note the important *Weierstrass M-test*, Theorem 6.3.1 of Reed. This gives a very useful criterion for uniform convergence of a series of functions.

The results about uniform convergence of a sequence of functions in Section14.2 (Reed §5.2) immediately give corresponding results for series (Reed Theorem 6.3.1 — last sentence; Theorem 6.3.2, Theorem 6.3.3). In particular:

- A uniformly convergent series of continuous functions has a continuous limit.
- Integration behaves well in the sense that for a uniformly convergent series of continuous functions on a closed bounded interval, the integral of the sum is the sum of the integrals (i.e. summation and integration can be interchanged).
- If a series of C^1 functions converges uniformly on an interval to a function f, ***and*** if the derivatives also converge uniformly to g say, then f is C^1 and $f' = g$ (i.e. summation and differentiation can be interchanged).

Study example 1 on page 240 of Reed. Example 2 is * material (it gives an example of a continuous and nowhere differentiable function).

15. Metric spaces

Study Reed §5.6.

The best way to study convergence and continuity is to do it abstractly by means of metric spaces. This has the advantage of simplicity and generality.

15.1. Definition and examples. The reference is Reed §5.6 up to the end of the first paragraph on page 201.

The basic idea we need is the notion of a "distance function" or a "metric".

DEFINITION 15.1.1. A metric space is a set $\mathcal{M}$ together with a function $\rho : \mathcal{M} \times \mathcal{M} \to [0, \infty)$ (called a *metric*) which satisfies for all $x, y, z \in \mathcal{M}$

1. $\rho(x, y) \geq 0$ and $\rho(x, y) = 0$ iff $x = y$, (positivity)
2. $\rho(x, y) = \rho(y, x)$, (symmetry)
3. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$, (triangle inequality)

The idea is that $\rho(x, y)$ measures the "distance (or length of a shortest route) from x to y". The three requirements are natural. In particular, the third might be thought of as saying that one route from x to y is to take a shortest route from x to z and then a shortest route from z to y — but there may be an even shorter route from x to y which does not go via z.

EXAMPLE 15.1.2 (Metrics on $\mathbb{R}^n$). We discussed the Euclidean metric on $\mathbb{R}^n$ in Sections 10.1 and 10.2

Example 3 on page 197 of Reed is important. It can be generalised to give the following metrics on $\mathbb{R}^n$:

$$\rho_p(\mathbf{x}, \mathbf{y}) = ((x_1 - y_1)^p + \dots (x_n - y_n)^p)^{1/p} = \left(\sum_{i=1}^{n} (x_i - y_i)^p \right)^{1/p},$$

$$\rho_{\max}(\mathbf{x}, \mathbf{y}) = \max\{\, |x_i - y_i| : i - 1, \dots, n \,\},$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $1 \leq p < \infty$.

We sometimes write ρ_∞ for $\rho_{\max}$.

Note the diagrams in Reed showing $\{\, \mathbf{x} \in \mathbb{R}^2 | \rho_p(\mathbf{x}, \mathbf{0}) \leq 1 \,\}$ for $p = 1, 2, \infty$.

The proof that ρ_p is a metric has already been given for $p = 2$ (the Euclidean metric). The cases $p = 1, \infty$ are Problem 1 page 202 of Reed in case $n = 2$, but the proofs trivially generalise to $n > 2$. The case of arbitrary p is trickier, and I will set it later as an assignment problem (with hints!).

EXAMPLE 15.1.3 (Metrics on function spaces). These are extremely important. See also Section 14.3.

The basic example is the sup metric ρ_∞ on $\mathcal{C}[a, b]$, the set of continuous functions $f : [a, b] \to \mathbb{R}$. See Reed Example 2 page 197.

More generally, one has the sup metric on

1. $\mathcal{B}(E)$, the set of bounded functions $f : E\ (\subseteq \mathbb{R}^n) \to \mathbb{R}$,
2. $\mathcal{BC}(E)$, the set of bounded continuous functions $f : E\ (\subseteq \mathbb{R}^n) \to \mathbb{R}$.

In the second case, if E is compact, we need only require continuity as this already implies boundedness.

The metrics ρ_p $(1 \leq p < \infty)$ on $\mathcal{C}[a, b]$ (and generalisations to other sets than $[a, b]$) have also been discussed, see Section 14.4.

***** If we regard $(a_1, \dots, a_n)$ as a function

$$a : \{1, \dots, n\} \to \mathbb{R},$$

and interpret

$$\int |a|^p \quad \text{as} \quad \sum_{i=1}^{n} |a(i)|^p,$$

(which is indeed the case for an appropriate "measure"), then the metrics ρ_p ($1 \leq p < \infty$) on function spaces give the metrics ρ_p on $\mathbb{R}^n$ as a special case.

EXAMPLE 15.1.4 (Subsets of a metric space). Any subset of a metric space is itself a metric space with the same metric, see the first paragraph of Reed page 199.

See also the first example in Example 4 of Reed, page 199.

EXAMPLE 15.1.5 (The discrete metric). A simple metric on any set $\mathcal{M}$ is given by

$$d(x,y) = \begin{cases} 1 & x \neq y, \\ 0 & x = y. \end{cases}$$

Check that this is a metric. It is usually called the *discrete metric*. (It is a particular case of what Reed calls a "discrete metric" in Problem 10 page 202.) It is not very useful, other than as a source of counterexamples to possible conjectures.

EXAMPLE 15.1.6 (Metrics on strings of symbols and DNA sequences). Example 5 in Reed is a discussion about metrics as used to estimate the difference between two DNA molecules. This is important in studying evolutionary trees.

EXAMPLE 15.1.7 (The natural metric on a normed vector space). Any normed vector space gives rise to a metric space defined by

$$\rho(x,y) = \|x - y\|.$$

The three properties of a metric follow from the properties of a norm. Examples are the metrics ρ_p for $1 \leq p \leq \infty$ on both $\mathbb{R}^n$ and the metrics ρ_p for $1 \leq p \leq \infty$ on spaces of functions.

But metric spaces are much more general. In particular, the discrete metric, and Examples 4 and 5 on pages 199, 200 of Reed are not metrics which arise from norms.

15.2. Convergence in a metric space. Reed page 201 paragraphs 2–4.

Once one has a metric, the following is the natural notion of convergence.

DEFINITION 15.2.1. A sequence (x_n) in a metric space $(\mathcal{M}, \rho)$ converges to $x \in \mathcal{M}$ if $\rho(x_n, x) \to 0$ as $n \to \infty$.

Notice that convergence in a metric space is reduced to the notion of convergence of a sequence of real numbers.

In the same way as for sequences in $\mathbb{R}$, there can be at most one limit.

We discussed convergence in $\mathbb{R}^n$ in Section 11.1. This corresponds to convergence with respect to the Euclidean metric. We will see in the next section that convergence with respect to any of the metrics ρ_p on $\mathbb{R}^n$ is the same.

Convergence of a sequence of functions in the sup norm is the same as convergence in the sup metric. But convergence in the other metrics ρ_p is *not* the same.

For example, let

$$f_n(x) = \begin{cases} 0 & -1 \leq x \leq 0 \\ nx & 0 \leq x \leq \frac{1}{n} \\ 1 & \frac{1}{n} \leq x \leq 1 \end{cases}, \qquad f(x) = \begin{cases} 0 & -1 \leq x \leq 0 \\ 1 & 0 < x \leq 1 \end{cases}.$$

Then $\rho_\infty(f_n, f) = 1$ and $\rho_1(f_n, f) = \frac{1}{2n}$ (*why?*). Hence $f_n \to f$ in $(\mathcal{R}[-1,1], \rho_1)$ (the space of Riemann integrable functions on $[-1,1]$ with the ρ_1 metric). But $f_n \not\to f$ in $(\mathcal{R}[-1,1], \rho_\infty)$[47].

15.3. Uniformly equivalent metrics. The reference here is the Definition and Theorem on Reed page 201, *and* Problem 11 page 203. However, we will only consider *uniformly* equivalent metrics, rather than equivalent metrics.

DEFINITION 15.3.1. Two metrics ρ and σ on the same set $\mathcal{M}$ are *uniformly equivalent*[48] if there are positive numbers c_1 and c_2 such that

$$c_1\rho(x,y) \le \sigma(x,y) \le c_2\rho(x,y)$$

for all $x, y \in \mathcal{M}$.

It follows that

$$c_2^{-1}\sigma(x,y) \le \rho(x,y) \le c_1^{-1}\sigma(x,y).$$

Reed also defines a weaker notion of "equivalent" ("uniformly equivalent" implies "equivalent" but not conversely). We will just need the notion of equivalent. The following theorem has a slightly simpler proof than the analogous one in the text for equivalent metrics.

THEOREM 15.3.2. *Suppose ρ and σ are uniformly equivalent metrics on $\mathcal{M}$. Then $x_n \to x$ with respect to ρ iff $x_n \to x$ with respect to σ.*

PROOF. Since $\sigma(x_n, x) \le c_2\rho(x_n, x)$, it follows that $\rho(x_n, x) \to 0$ implies $\sigma(x_n, x) \to 0$. Similarly for the other implication. □

The following theorem is a generalisation of Problem 12 page 203.

EXAMPLE 15.3.3. *The metrics ρ_p on $\mathbb{R}^n$ are uniformly equivalent to one another.*

PROOF. It is sufficient to show that ρ_p for $1 \le p < \infty$ is equivalent to ρ_∞ (since if two metrics are each uniformly equivalent to a third metric, then from Definition15.3.1 it follows fairly easily that they are uniformly equivalent to one another — *Exercise*).

But

$$\begin{aligned}\rho_p(\mathbf{x},\mathbf{y}) &= \left(|x_1 - y_1|^p + \cdots + |x_n - y_n|^p\right)^{1/p} \\ &\le \left(n \max_{1\le i\le n} |x_i - y_i|^p\right)^{1/p} = n^{1/p}\rho_\infty(\mathbf{x},\mathbf{y}),\end{aligned}$$

and

$$\rho_\infty(\mathbf{x},\mathbf{y}) = \max_{1\le i\le n} |x_i - y_i| \le \rho_p(\mathbf{x},\mathbf{y}).$$

This completes the proof. □

REMARK 15.3.1. *The metrics ρ_p on function spaces are <u>not</u> uniformly equivalent (or even equivalent).*

It is easiest to see this for ρ_1 and ρ_∞ on $\mathcal{R}[a,b]$ (the space of Riemann integrablefunctions on $[a,b]$). It follows from the example in the prevous section, since $f_n \to f$ with respect to ρ_1 but not with respect to ρ_∞.

[47] Remember that "Riemann integrable" implies "bounded" according to our definitions.

[48] What is here called "uniformly equivalent" is usually called "equivalent". But for consistency, I will keep to the convention in Reed.

15.4. Cauchy sequences and completeness. The reference here is Reed §5.7 to the end of page 204.

The definition of a Cauchy sequence is exactly as for Definition 11.5.1 in $\mathbb{R}^p$.

DEFINITION 15.4.1. A sequence $(\mathbf{x}_n)$ is a *Cauchy sequence* if for any $\epsilon > 0$ there is a corresponding N such that

$$m, n \geq N \quad \Rightarrow \quad \rho(\mathbf{x}_m, \mathbf{x}_n) \leq \epsilon.$$

It follows from the Definition in Reed Page 177 that for a sequence of functions "Cauchy in the sup norm" is the same as "Cauchy with respect to the sup metric".

It follows as for convergence, that a sequence is Cauchy with respect to one metric then it is Cauchy with respect to any uniformly equivalent metric.

The definition of a complete metric space is *extremely important* (Reed page 204).

DEFINITION 15.4.2. A metric space $(\mathcal{M}, \rho)$ is complete if every Cauchy sequence from $\mathcal{M}$ converges to an element in $\mathcal{M}$.

Moreover (Problem 10(a) page 209):

THEOREM 15.4.3. *Suppose ρ and σ are uniformly equivalent metrics on $\mathcal{M}$. Then $(\mathcal{M}, \rho)$ is complete iff $(\mathcal{M}, \sigma)$ is complete.*

PROOF. Suppose $(\mathcal{M}, \rho)$ is complete. Let (x_n) be Cauchy with respect to σ. Then it is Cauchy with respect to ρ and hence converges to x (say) with respect to ρ. By Theorem 15.3.2, $x_n \to x$ with respect to σ. It follows that $(\mathcal{M}, \sigma)$ is complete. □

EXAMPLE 15.4.4. The Cauchy Completeness Axiom says that $(\mathbb{R}, \rho_2)$ is complete, and it follows from Proposition 11.5.3 that $(\mathbb{R}^n, \rho_2)$ is also complete.

It then follows from the previous theorem that $(\mathbb{R}^n, \rho_p)$ *is complete for any* $1 \leq p \leq \infty$.

More generally, *if $M \subseteq \mathbb{R}^n$ is closed, then (M, ρ_p) is complete for any* $1 \leq p \leq \infty$. To see this, consider any Cauchy sequence $(\mathbf{x}_i)$ in (M, ρ_2). Since $(\mathbf{x}_i)$ is also a Cauchy sequence in $\mathbb{R}^n$ it converges to $\mathbf{x}$ (say) in $\mathbb{R}^n$. Since M is closed, $\mathbf{x} \in M$. Hence (M, ρ_2) is complete. It follows from the previous theorem that (M, ρ_p) is complete for any $1 \leq p \leq \infty$.

If $M \subseteq \mathbb{R}^n$ is not closed, then (M, ρ_p) is not complete. To see this, choose a sequence $(\mathbf{x}_i)$ in M such that $\mathbf{x}_i \to \mathbf{x} \notin M$. Since $(\mathbf{x}_i)$ converges in $\mathbb{R}^n$ it is Cauchy in $\mathbb{R}^n$ and hence in M, but it does not converge to a member of M. Hence (M, ρ_p) is not complete.

EXAMPLE 15.4.5. $(\mathcal{C}[a, b], \rho_\infty)$ *is complete* from Reed Theorem 5.3.3 page 178.

However, $(\mathcal{C}[a, b], \rho_1)$ *is not complete.*

To see this, consider the example in Section 15.2. It is fairly easy to see that the sequence (f_n) is Cauchy in the ρ_1 metric. In fact if $m > n$ then

$$\rho_1(f_n, f_m) = \int_0^{1/n} |f_n - f_m| \leq \int_0^{1/n} |f_n| + |f_m| \leq 2/n.$$

Since $f_n \to f$ in the larger space $(\mathcal{R}[-1, 1], \rho_1)$ of Riemann integrable functions on $[-1, 1]$ with the ρ_1 metric, it follows that (f_n) is Cauchy in this space, and hence is also Cauchy in $(\mathcal{C}[a, b], \rho_1)$. But f_n does not converge to a member of this space[49].

[49] If it did, it would converge to the same function in the larger space, contradicting uniqueness of limits in the larger space

15.5. Contraction Mapping Principle. The reference here is Reed §5.7.

The definition of a contraction in Section 11.6 for a function $F : S \to S$ where $S \subseteq \mathbb{R}^p$ applies with S replaced by $\mathcal{M}$ and d replaced by ρ for any metric space $(\mathcal{M}, \rho)$.

The Contraction Mapping Theorem, Theorem 11.6.1 applies with exactly the same proof, for any *complete* metric space $(\mathcal{M}, \rho)$ instead of the closed set S and the Euclidean metric d. Or see Reed Theorem 5.7.1 page 205.

Completeness is needed in the proof in order to show that the Cauchy sequence of iterates actually converges to a member of $\mathcal{M}$.

In Section 16 we will give some important applications of the Contraction Mapping Principle. In particular, we will apply it to the complete metric space $(\mathcal{C}, [a, b], \rho_\infty)$.

15.6. Topological notions in metric spaces*. In this section I will point out that much of what we proved for $(\mathbb{R}^n, d)$ applies to an arbitrary metric space $(\mathcal{M}, \rho)$ with the same proofs. The extensions are straightforward and I include them for completeness and future reference in the Analysis II course. But we will not actually need the material in this course.

References at about the right level are "Primer of Modern Analysis" by Kennan T. Smith, Chapter 7, and "An Introduction to Analysis" (second edition) by William Wade, Chapter 10.

The definitions of *neighbourhood*, *open ball* $B_r(x)$, and *open set* are analogous to those in Section 10.5. Theorem 10.5.2 remains true with $\mathbb{R}^n$ replaced by $\mathcal{M}$.

The definition of *closed set* is analogous to that in Section 10.6 and Theorem 10.6.2 is true with $\mathbb{R}^n$ replaced by $\mathcal{M}$.

The definitions of *boundary point*, *boundary*, *interior point*, *interior*, *exterior point* and *exterior* are analogous to those in Section 10.8. The three dot points there are true for $S \subseteq \mathcal{M}$.

The *product* of two metric spaces $(\mathcal{M}_1, \rho_1)$ and $(\mathcal{M}_2, \rho_2)$ is the set $\mathcal{M}_1 \times \mathcal{M}_2$ with the metric $\rho_1 \times \rho_2$ defined by

$$(\rho_1 \times \rho_2)\big((x_1, y_1), (x_2, y_2)\big) = \sqrt{\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2}.$$

If A_1 is open in $(\mathcal{M}_1, \rho_1)$ and A_2 is open in $(\mathcal{M}_2, \rho_2)$ the $A_1 \times A_2$ is open in $(\mathcal{M}_1 \times \mathcal{M}_2, \rho_1 \times \rho_2)$. The proof is the same as Theorem 10.10.1.

A set is closed iff every convergent sequence from the set has its limit in the set; the proof is the same as for Theorem 11.2.1.

The definition of a *bounded set* is the same as in Section 11.3, tha analogue of Proposition 11.3.2 holds (with the origin replaced by any fixed point $a^* \in \mathcal{M}$), convergent sequences are bounded (Proposition 11.3.4).

Bounded sequences do *not* necessarily have a convergent subsequence (c.f. Theorem 11.4.2). The sequence (f_n) in Section 15.2 is bounded in $\mathcal{C}[-1, 1]$ but no subsequence converges. In fact it is not too hard to check that

$$\rho_\infty(f_n, f_m) \geq \frac{1}{2}$$

if $m \geq 2n$. To see this take $x = \frac{1}{m}$. Then

$$|f_m(x) - f_n(x)| = 1 - \frac{n}{m} \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

It follows that no subsequence can be Cauchy, and in particular no subsequence can converge.

Cauchy sequences are defined as in Definition 11.5.1. Convergent sequences are Cauchy, this is proved easily as for sequences of real numbers. But Cauchy

sequences will not necesarily converge (to a point in the metric space) unless the metric space is complete.

A metric space $(\mathcal{M}, \rho)$ is *sequentially compact* if every sequence from $\mathcal{M}$ has a convergent subsequence (to a point in $\mathcal{M}$). This corresponds to the set A in Definition 12.1.1 being sequentially compact. The metric space $(\mathcal{M}, \rho)$ is *compact* if every open cover (i.e. a cover of $\mathcal{M}$ by subsets of $\mathcal{M}$ that are open in $\mathcal{M}$) has a finite subcover. This corresponds to the set A in Definition 12.2.2 being compact (although A is covered by sets which are open in $\mathbb{R}^p$, the restriction of these sets to A is a cover of A by sets that are open in A).

A subset of a metric space $(\mathcal{M}, \rho)$ is sequentially compact (compact) iff it is sequentially compact (compact) when regarded as a metric space with the induced metric from $\mathcal{M}$.

A metric space $(\mathcal{M}, \rho)$ is *totally bounded* if for every $\varepsilon > 0$ there is a cover of $\mathcal{M}$ by a *finite* number of open balls (in $\mathcal{M}$) of radius ε. Then it can be proved *a metric space is totally bounded iff every sequence has a Cauchy subsequence*; see Smith Theorem 9.9. (Note that what Smith calls "compact" is what is here, and usually, called "sequentially compact".)

A metric space is complete and totally bounded iff it is sequentially compact iff it is compact. This is the analogue of Theorems 12.1.2 and 12.3.1, where the metric space corresponds to A with the metric induced from $\mathbb{R}^p$.

The proof of the first 'iff" is fairly straightforward. First suppose $(\mathcal{M}, \rho)$ is complete and totally bounded. If (x_n) is a sequence from $\mathcal{M}$ then it has a Cauchy subsequence by the result two paragraphs back, and this subsequence converges to a limit in $\mathcal{M}$ by completeness and so $(\mathcal{M}, \rho)$ is sequentially compact. Next suppose $(\mathcal{M}, \rho)$ is sequentially compact. Then it is totally bounded again by the result two paragraphs back. To show it is complete suppose the sequence (x_n) is Cauchy: first by compactness it has a subsquence which converges to a member of $\mathcal{M}$, and second we use the fact (*Exercise*) that if a Cauchy sequence has a convergent subsequence then the Cauchy sequence itself must converge to the same limit as the subsequence.

The proof of the second "iff" is similar to that on Theorem 12.3.1. The proof that "compact" implies "sequentially compact" is esentially identical. The proof in the other direction uses the existence of a countable dense subset of $\mathcal{M}$. (This replaces the idea of considering those points in $\mathbb{R}^p$ all of whose components are rational.) We say a subset D of $\mathcal{M}$ is *dense* if for each $x \in \mathcal{M}$ and each $\varepsilon > 0$ there is an $x^* \in D$ such that $x \in B_\varepsilon(x^*)$. The existence of a countable dense subset of $\mathcal{M}$ follows from sequential compactness.[50]

If $(\mathcal{M}, \rho)$ and $(\mathcal{N}, \sigma)$ are metric spaces and $f : \mathcal{M} \to \mathcal{N}$ then we say f is *continuous* at $a \in \mathcal{M}$ if

$$(x_n) \subseteq \mathcal{M} \text{ and } x_n \to a \Rightarrow f(x_n) \to f(a).$$

It follows that *f is continuous at a iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x, y \in \mathcal{M}$:*

$$\rho(x, y) < \delta \Rightarrow \sigma(f(x), f(y)) < \varepsilon.$$

The proof is the same as for Theorem 7.1.2. One defines *uniform continuity* as in Definition 13.1.3.

A continuous function defined on a compact set is bounded above and below and has a maximum and a minimum value. Moreoveor it is uniformly continuous. The

[50]We have already noted that a sequentially compact set is totally bounded. Let A_k be the finite set of points corresponding to $\varepsilon = \frac{1}{k}$ in the definition of total boundedness. Let $A = \bigcup_{k=1}^{\infty} A_k$. Then A is countable. It is also dense, since if $x \in \mathcal{M}$ then there exist points in A as close to x as we wish.

proof is the same as for Theorem 13.2.1, after using the fact that compactness is equivalent to sequential compactness.

Limit points, *isolated points*, the definition of *limit of a function at a point*, and the equivalence of this definition with the ε–δ characterisation, are completely analogous to Definition 13.3.1 and Theorem 13.3.2.

A function $f : \mathcal{M} \to \mathcal{N}$ is continuous iff the inverse image of every open (closed) set in $\mathcal{N}$ is open (closed) in $\mathcal{M}$. The proof is esentially the same as in Theorem 13.4.3.

The continuous image of a compact set is compact. The proof is the same as in Theorem 13.5.1.

The inverse of a one-to-one continuous function defined on a compact set is continuous. The proof is essentially the same as for Theorem 13.5.2.

16. Some applicatons of the Contraction Mapping Principle

16.1. Markov processes. Recall Theorem 4.0.2:

THEOREM 16.1.1. *If all entries in the probability transition matrix P are greater than 0, and $\mathbf{x}_0$ is a probability vector, then the sequence of vectors*

$$\mathbf{x_0}, P\mathbf{x_0}, P^2\mathbf{x_0}, \dots,$$

converges to a probability vector $\mathbf{x}^*$, and this vector does not depend on $\mathbf{x_0}$.

Moreover, the same results are true even if we just assume P^k has all entries non-zero for some integer $k > 1$.

The vector $\mathbf{x}^*$ is the unique non-zero solution of $(P - I)\mathbf{x}^* = \mathbf{0}$.

We will give a proof that uses the Contraction Mapping Principle. The proof is rather subtle, since P is only a contraction with respect to the ρ_1 metric, and it is also necessary to restrict to the subset of $\mathbb{R}^n$ consisting of those vectors $\mathbf{a}$ such that $\sum a_i = 1$.

LEMMA 16.1.2. *Suppose P is a probability transition matrix all of whose entries are at least ε, for some $\varepsilon > 0$. Then P is a contraction map on*

$$\mathcal{M} = \{\, \mathbf{a} \in \mathbb{R}^n \mid \sum a_i = 1 \,\}.$$

in the ρ_1 metric, with contraction constant $1 - n\varepsilon$.

PROOF. We will prove that

$$\rho_1(P\mathbf{a}, P\mathbf{b}) \le (1 - n\varepsilon)\, \rho_1(\mathbf{a}, \mathbf{b}) \tag{24}$$

for all $\mathbf{a}, \mathbf{b} \in \mathcal{M}$.

Suppose $\mathbf{a}, \mathbf{b} \in \mathcal{M}$. Let $\mathbf{v} = \mathbf{a} - \mathbf{b}$. Then $\mathbf{v} \in H$, where

$$H = \{\, \mathbf{v} \in \mathbb{R}^n \mid \sum v_i = 0 \,\}.$$

Moreover,

$$\rho_1(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_1 = \|\mathbf{v}\|_1,$$
$$\rho_1(P\mathbf{a}, P\mathbf{b}) = \|P\mathbf{a} - P\mathbf{b}\|_1 = \|P\mathbf{v}\|_1,$$

where $\|\mathbf{w}\|_1 = \sum |w_i|$.

Thus in order to prove (24) it is sufficient to show

$$\|P\mathbf{v}\|_1 \le (1 - n\varepsilon)\, \|\mathbf{v}\|_1 \tag{25}$$

for all $\mathbf{v} \in H$.

From the following Lemma, we can write

$$\mathbf{v} = \sum_{ij} \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j), \quad \text{where } \eta_{ij} > 0,\ \|\mathbf{v}\|_1 = 2\sum \eta_{ij}. \tag{26}$$

Then

$$\begin{aligned} \|P\mathbf{v}\|_1 &= \Big\| \sum_{ij} \eta_{ij} P(\mathbf{e}_i - \mathbf{e}_j) \Big\|_1 \\ &\le \sum_{ij} \eta_{ij}\, \|P(\mathbf{e}_i - \mathbf{e}_j)\|_1 \end{aligned}$$

(by the triangle inequality for $\|\cdot\|_1$ and using $\eta_{ij} > 0$)

$$\begin{aligned} &= \sum_{ij} \eta_{ij} \Big\| \sum_k (P_{ki} - P_{kj})\, \mathbf{e}_k \Big\|_1 \\ &= \sum_{ij} \eta_{ij} \sum_k |P_{ki} - P_{kj}| \end{aligned}$$

$$= \sum_{ij} \eta_{ij} \sum_k \big(P_{ki} + P_{kj} - 2\min\{P_{ki}, P_{kj}\}\big)$$

(since $|a-b| = a+b-2\min\{a,b\}$ for any real numbers a and b, as one sees by checking, without loss of generality, the case $b \le a$)

$$= \sum_{ij} \eta_{ij}(2 - 2\varepsilon)$$

(since the columns in P sum to 1 and each entry is $\ge \varepsilon$)

$$= (1 - n\varepsilon)\|\mathbf{v}\|_1 \quad \text{(from (26))}$$

This completes the proof. □

LEMMA 16.1.3. *Suppose* $\mathbf{v} \in \mathbb{R}^n$ *and* $\sum v_i = 0$. *Then* $\mathbf{v}$ *can be written in the form*

$$\mathbf{v} = \sum_{i,j} \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j), \quad \textit{where } \eta_{ij} > 0,\ \|\mathbf{v}\|_1 = 2\sum \eta_{ij}.$$

PROOF. If $\mathbf{v}$ is in the span of two basis vectors, then after renumbering we have

$$\begin{aligned} \mathbf{v} &= v_1\mathbf{e}_1 + v_2\mathbf{e}_2 \quad (\text{where } \ v_1 + v_2 = 0) \\ &= v_1(\mathbf{e}_1 - \mathbf{e}_2) \\ &= v_2(\mathbf{e}_2 - \mathbf{e}_1). \end{aligned}$$

Since either $v_1 \ge 0$ and then $\|\mathbf{v}\|_1 = 2v_1$, or $v_2 \ge 0$ and then $\|\mathbf{v}\|_1 = 2v_2$, this proves the result in this case.

Suppose the claim is true for any $\mathbf{v} \in H$ which is in the span of k basis vectors. Assume that $\mathbf{v}$ is spanned by $k+1$ basis vectors, and write (after renumbering if necessary)

$$\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_{k+1}\mathbf{e}_{k+1}.$$

where $v_1 + \cdots + v_{k+1} = 0$.

Choose p so

$$|v_p| = \max\{\,|v_1|, \ldots, |v_{k+1}|\,\}.$$

Choose q so v_q has the opposite sign to v_p (this is possible since $\sum v_i = 0$). Note that

$$|v_p + v_q| = |v_p| - |v_q| \tag{27}$$

since $|v_q| \le |v_p|$ and since v_q has the opposite sign to v_p.

Write (where $\widehat{\ }$ indicates that the relevant term is missing from the sum)

$$\begin{aligned} \mathbf{v} &= \Big(v_1\mathbf{e}_1 + \cdots + \widehat{v_q\mathbf{e}_q} + \cdots + (v_p + v_q)\mathbf{e}_p + \cdots + v_{kn+1}\mathbf{e}_{k+1}\Big) + v_q(\mathbf{e}_q - \mathbf{e}_p) \\ &= \mathbf{v}^* + v_q(\mathbf{e}_q - \mathbf{e}_p). \end{aligned}$$

Since $\mathbf{v}^*$ has no $\mathbf{e}_q$ component, and the sum of the coefficients of $\mathbf{v}^*$ is $\sum v_i = 0$, we can apply the inductive hypothesis to $\mathbf{v}^*$ to write

$$\mathbf{v}^* = \sum \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j), \quad \text{where } \eta_{ij} > 0, \|\mathbf{v}^*\|_1 = 2\sum \eta_{ij}.$$

In particular,

$$\mathbf{v} = \begin{cases} \sum \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j) + |v_q|(\mathbf{e}_q - \mathbf{e}_p) & v_q \ge 0 \\ \sum \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j) + |v_q|(\mathbf{e}_p - \mathbf{e}_q) & v_q \le 0 \end{cases}$$

All that remains to be proved is

$$\|\mathbf{v}\|_1 = 2\sum \eta_{ij} + 2|v_q|,$$

i.e.

$$\|\mathbf{v}\|_1 = \|\mathbf{v}^*\|_1 + 2|v_q|. \tag{28}$$

But

$$\begin{aligned}\|\mathbf{v}^*\|_1 &= |v_1| + \cdots + \widehat{|v_q|} + \cdots + |v_p + v_q| + \cdots + |v_{k+1}| \\ &= (|v_1| + \cdots + |v_{k+1}|) - |v_q| - |v_p| + |v_p + v_q| \\ &= (|v_1| + \cdots + |v_{k+1}|) - 2|v_q| \quad \text{by (27)} \\ &= \|\mathbf{v}\|_1 - 2|v_q|.\end{aligned}$$

This proves (28) and hence the Lemma. □

Proof of Theorem 16.1.1. The first paragraph of the theorem follows from the Contraction Mapping Principle and Lemma 16.1.2.

The last paragraph was proved after the statement of Theorem 4.0.2.

For the second paragraph we consider a sequence $P^i(\mathbf{x}_0)$ with $i \to \infty$. From Lemma 16.1.2, P^k is a contraction map with contraction constant λ (say). For any natural number i we can write $i = mk + j$, where $0 \le j < k$.

Then

$$\begin{aligned}\rho_1(P^i\mathbf{x}_0, \mathbf{x}^*) &= \rho_1(P^{mk+j}\mathbf{x}_0, \mathbf{x}^*) \\ &\le \rho_1(P^{mk+j}\mathbf{x}_0, P^{mk}\mathbf{x}_0) + \rho_1(P^{mk}\mathbf{x}_0, \mathbf{x}^*) \\ &\le \rho_1\big((P^k)^m P^j\mathbf{x}_0, (P^k)^m\mathbf{x}_0\big) + \rho_1((P^k)^m\mathbf{x}_0, \mathbf{x}^*) \\ &\le \lambda^m \rho_1\big(P^j\mathbf{x}_0, \mathbf{x}_0\big) + \rho_1((P^k)^m\mathbf{x}_0, \mathbf{x}^*) \\ &\le \lambda^m \max_{0\le j<k} \rho_1\big(P^j\mathbf{x}_0, \mathbf{x}_0\big) + \rho_1((P^k)^m\mathbf{x}_0, \mathbf{x}^*).\end{aligned}$$

(Note that $m \to \infty$ as $i \to \infty$). The first term approaches 0 since $\lambda^m \to 0$ and the second approaches 0 by the Contraction Mapping Principle applied to the contraction map P^k. □

Remark 16.1.1. The fact we only needed to assume some power of P was a contraction map can easily be generalised.

That is, in the Contraction Mapping Principle, Theorem 11.6.1, if we assume $\overset{1}{F} \circ \cdots \circ \overset{k}{F}$ is a contraction map for some k then there is still a unique fixed point. The proof is similar to that above for P.

16.2. Integral Equations. See Reed Section 5.4 for discussion.

Instead of the long proof in Reed of the main theorem, Theorem 5.4.1, we see here that it is a consequence of the Contraction Mapping Theorem.

Theorem 16.2.1. *Let $F(x)$ be a continuous function defined on $[a, b]$. Suppose $K(x, y)$ is continuous on $[a, b] \times [a, b]$ and*

$$M = \max\{|K(x, y)| \mid a \le x \le b,\ a \le y \le b\}.$$

Then there is a unique continuous function $\psi(x)$ defined on $[a, b]$ such that

$$\psi(x) = f(x) + \lambda \int_a^b K(x, y)\,\psi(y)\,dy, \tag{29}$$

provided $|\lambda| < \frac{1}{M(b-a)}$.

PROOF. We will use the contraction mapping principle on the complete metric space $(\mathcal{C}[a,b], \rho_\infty)$.

Define

$$T : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$$

by

$$(T\psi)(x) = f(x) + \lambda \int_a^b K(x,y)\,\psi(y)\,dy \tag{30}$$

Note that if $\psi \in \mathcal{C}[a,b]$, then $T\psi$ is certainly a *function* defined on $[a,b]$ — the value of $T\psi$ at $x \in [a,b]$ is obtained by evaluating the right side of (30). For fixed x, the integrand in (30) is a continuous function of y, and so the integral exists.

We next claim that $T\psi$ is *continuous*, and so indeed $T : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$.

The first term on the right side of (30) is continuous, being f.

The second term is also continuous. To see this let $G(x,y) = K(x,y)\,\psi(y)$ and let $g(x) = \int_a^b G(x,y)\,dy$. Then $T\psi = f + \lambda g$. To see that g is continuous on $[a,b]$, first note that

$$|g(x_1) - g(x_2)| = \Big| \int_a^b G(x_1,y) - G(x_2,y)\,dy \Big| \le \int_a^b |G(x_1,y) - G(x_2,y)|\,dy. \tag{31}$$

Now suppose $\epsilon > 0$. Since $G(x,y)$ is continuous on the closed bounded set $[a,b] \times [a,b]$, it is uniformly continuous there (by Theorem 13.2.1). Hence there is a $\delta > 0$ such that

$$|(x_1,y_1) - (x_2,y_2)| < \delta \quad \Rightarrow \quad |G(x_1,y_1) - G(x_2,y_2)| < \epsilon.$$

Using this it follows from (31) that

$$|x_1 - x_2| < \delta \quad \Rightarrow \quad |g(x_1) - g(x_2)| < \epsilon(b-a).$$

Hence g is uniformly continuous on $[a,b]$.

This completes the proof that $T\psi$ is continuous (in fact uniformly) on $[a,b]$. Hence

$$T : \mathcal{C}[a,b] \to \mathcal{C}[a,b].$$

Moreover, *ψ is a fixed point of T iff ψ solves* (29).

We next claim that T is *contraction map* on $\mathcal{C}[a,b]$. To see this we estimate

$$\begin{aligned} |T\psi_1(x) - T\psi_2(x)| &= |\lambda| \,\Big| \int_a^b K(x,y)\,(\psi_1(y) - \psi_2(y))\,dy \Big| \\ &\le |\lambda| \int_a^b |K(x,y)|\,|\psi_1(y) - \psi_2(y)|\,dy \\ &\le |\lambda|\,M(b-a)\,\rho_\infty(\psi_1,\psi_2). \end{aligned}$$

Since this is true for every $x \in [a,b]$, it follows that

$$\rho_\infty(T\psi_1, T\psi_2) \le |\lambda|\,M(b-a)\,\rho_\infty(\psi_1,\psi_2).$$

By the assumption on λ, it follows that T is a contraction map with contraction ratio $|\lambda|\,M(b-a)$, which is < 1.

Because $(\mathcal{C}[a,b], \rho_\infty)$ is a *complete* metric space, it follows that T has a unique fixed point, and so there is a unique continuous function ψ solving (29). □

REMARK 16.2.1. Note that we also have a way of approximating the solution. We can begin with some function such as

$$\psi_0(x) = 0 \quad \text{for } a \le x \le b,$$

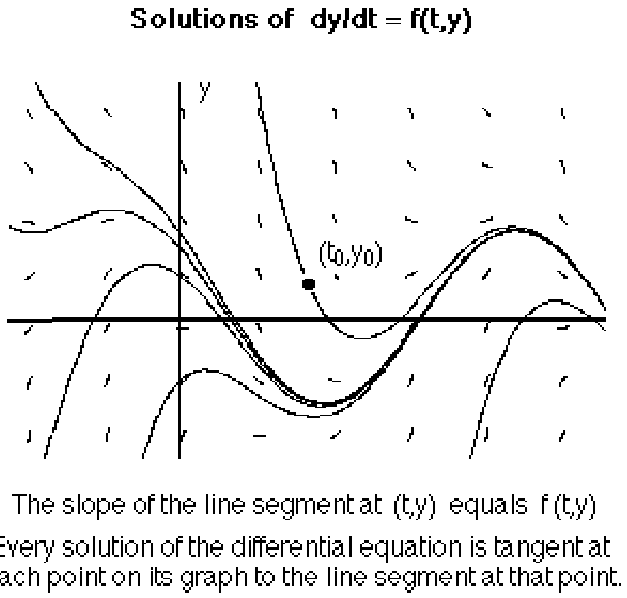
and then successively apply T to find better and better approximations to the solution ψ.

In practice, we apply some type of numerical integration to find approximations to $T\psi_0$, $T^2\psi_0$, $T^3\psi_0, \ldots$.

16.3. Differential Equations. In this section we will prove the *Fundamental Existence and Uniqueness Theorem for Differential Equations*, Theorem 7.1.1 of Reed.

You should first read Chapter 7 of Reed up to the beginning of Theorem 7.1.1.

We will prove Theorem 7.1.1 here, but with a simpler proof using the Contraction Mapping Principle.



THEOREM 16.3.1. *Let f be continuously differentiable on the square*

$$S = [t_0 - \delta, t_0 + \delta] \times [y_0 - \delta, y_0 + \delta].$$

Then there is a $T \le \delta$, and a unique continuously differentiable function $y(t)$ defined on $[t_0 - T, t_0 + T]$, such that

$$\begin{aligned} \frac{dy}{dt} &= f(t,y) \quad \text{on } [t_0 - T, t_0 + T], \\ y(t_0) &= y_0. \end{aligned} \tag{32}$$

REMARK 16.3.1.

1. The diagram shows the square S centred at (t_0, y_0) and illustrates the situation. We are looking for the unique curve through the point (t_0, y_0) which at every point is tangent to the line segment of slope $f(t, y)$.

   Note that the solution through the point (t_0, y_0) "escapes" through the top of the square S, and this is why we may need to take $T < \delta$.

   From the diagram, the absolute value of the slope of the solution will be $\le M$, where $M = \max\{\, |f(t,y)| \mid (t,y) \in S \,\}$. For this reason, we will require that $MT \le \delta$, i.e. $T \le \delta/M$.

2. The function $y(t)$ must be differentiable for the differential equation to be defined. But it then follows from the differential equation that dy/dt is in fact continuous.

   * In fact more is true. The right side of the differential equation is differentiable (*why*?) and its derivative is even continuous (*why*?). This shows that dy/dt is differentiable with continuous derivative, i.e. $y(t)$ is in fact twice continuously differentiable. If f is infinitely differentiable, it can similarly be shown that $y(t)$ is also infinitely differentiable.

Proof of Theorem.

Step 1: We first reduce the problem to an equivalent integral equation (not the same one as in the last section, however).

It follows by integrating (32) from t_0 to t that any solution $y(t)$ of (32) satisfies

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds. \tag{33}$$

Conversely, any continuous function $y(t)$ satisfying (33) is differentiable (by the Fundamental Theorem of Calculus) and its derivative satisfies the differential equation in (32). It also satisfies the initial condition $y(t_0) = y_0$ (*why?*).

Step 2: We next show that (33) is the same as finding a certain "'fixed point".

If $y(t)$ is a continuous function defined on $[t_0 - \delta, t_0 + \delta]$, let Fy (often written $F(y)$) be the function defined by

$$(Fy)(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds. \tag{34}$$

For the integral to be defined, we require that $(s, y(s))$ belong to the square S, and for this reason we will need to restrict to functions $y(t)$ defined on some sufficiently small interval $[t_0 - T, t_0 + T]$.

More precisely, since f is continuous and hence bounded on S, it follows that there is a constant M such that

$$(t, y) \in S \quad \Rightarrow \quad |f(t, y)| \le M.$$

This implies from (34) that

$$|(Fy)(t) - y_0| \le M\,|t - t_0|.$$

It follows that if the graph of $y(t)$ is in S, then so is the graph of $(Fy)(t)$, provided t satisfies $M\,|t - t_0| \le \delta$.

For this reason, we impose the restriction on T that $MT \le \delta$, i.e.

$$T \le \frac{\delta}{M}. \tag{35}$$

This ensures that $(Fy)(t)$ is defined for $t \in [t_0 - T, t_0 + T]$.

Since $f(s, y(s))$ is continuous, being a composition of continuous functions, it follows from the Fundamental Theorem of Calculus that $(Fy)(t)$ is differentiable, and in particular is continuous. Hence

$$F : \mathcal{C}[t_0 - T, t_0 + T] \to \mathcal{C}[t_0 - T, t_0 + T].$$

Moreover, it follows from the definition (34) that $y(t)$ *is a fixed point of* F *iff* $y(t)$ *is a solution of* (33) *on* $[t_0 - T, t_0 + T]$ *and hence iff* $y(t)$ *is a solution of* (32) *on* $[t_0 - T, t_0 + T]$.

Step 2: We next impose a further restriction on T in order that F be a contraction map. For this we need the fact that since $\frac{\partial f}{\partial t}$ is continuous on S, there is a constant

K such that

$$(t, y) \in S \quad \Rightarrow \quad \left|\frac{\partial f}{\partial y}\right| \leq K.$$

We now compute for any $t \in [t_0 - T, t_0 + T]$ and any $y_1, y_2 \in \mathcal{C}[t_0 - T, t_0 + T]$, that

$$\begin{aligned} |Fy_1(t) - Fy_2(t)| &= \left| \int_{t_0}^{t} f(s, y_1(s)) - f(s, y_2(s))\, ds \right| \\ &\leq \int_{t_0}^{t} \left| f(s, y_1(s)) - f(s, y_1(s)) \right| ds \\ &\leq \int_{t_0}^{t} K\, |y_1(s) - y_2(s)|\, ds \quad \text{by the Mean Value Theorem} \\ &\leq KT\, \rho_\infty(y_1, y_2) \end{aligned}$$

since $|t - t_0| \leq T$ and since $|y_1(s) - y_2(s)| \leq \rho_\infty(y_1, y_2)$.

Since t is any point in $[t_0 - T, t_0 + T]$, it follows that

$$\rho_\infty(Fy_1, Fy_2) \leq KT\, \rho_\infty(y_1, y_2).$$

We now make the second restriction on T that

$$T \leq \frac{1}{2K}. \tag{36}$$

This guarantees that F is a contraction map on $\mathcal{C}[t_0 - T, t_0 + T]$ with contraction constant $\frac{1}{2}$. It follows that F has a unique fixed point, and that this is then the unique solution of (32). □

Remark 16.3.2. It appears from the diagram that we do not really need the second restriction (36) on T. In other words, we should only need (35). This is indeed the case. One can show by repeatedly applying the previous theorem that the solution can be continued until it either escapes through the top, bottom or sides of S. In fact this works for much more general sets S.

In particular, if the function $f(t, y)$ and the differential equation are defined for all $(t, y) \in \mathbb{R}^2$, then the solution will either approach $+\infty$ or $-\infty$ at some finite time t^*, or will exist for all time. For more discussion see Reed Section 7.2.

Remark 16.3.3. The same proof, with only notational changes, works for general first order systems of differential equations of the form

$$\begin{aligned} \frac{dy_1}{dt} &= f_1(t, y_1, y_2, \ldots, y_n) \\ \frac{dy_2}{dt} &= f_2(t, y_1, y_2, \ldots, y_n) \\ &\vdots \\ \frac{dy_n}{dt} &= f_n(t, y_1, y_2, \ldots, y_n). \end{aligned}$$

Remark 16.3.4. Second and higher order differential equations can be reduced to first order systems by introducing new variables for the lower order derivates.

For example, the second order differential equation

$$y'' = F(t, y, y')$$

is equivalent to the first order system of differential equations

$$y_1' = y_2, \quad y_2' = F(t, y_1, y_2).$$

To see this, suppose $y(t)$ is a solution of the given differential equation and let $y_1 = y,\ y_2 = y'$. Then clearly $y_1(t),\ y_2(t)$ solve the first order system.

Conversely, if $y_1(t),\ y_2(t)$ solve the first order system let $y = y_1$. Then $y' = y_2$ and $y(t)$ solves the second order differential equation.

It follows from the previous remark that a similar Existence and Uniqueness Theorem applies to second order differential equations, provided we specify *both* $y(t_0)$ and $y'(t_0)$.

A similar remark applies to nth order differential equations, except that one must specify the first $n-1$ derivatives at t_0.

Finally, similar remarks also apply to higher order systems of differential equations. There are no new ideas involved, just notation!

17. Differentiation of Real-Valued Functions

I have included quite a lot of additional material, for future reference.

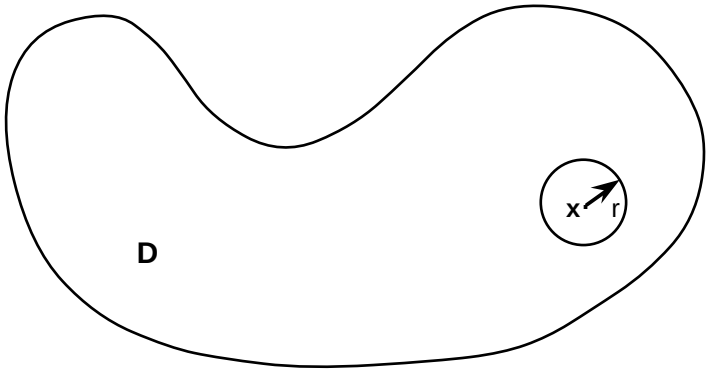
You need only consider Subsections 1, 3, 4, 5, 7, 8, 10. What is required is just a basic understanding of the ideas and application to straightforward examples.

Note that the definition of "differentiable" on Reed page 155 is completely non-standard. It is the same as what is here, and elsewhere, called "continuously differentiable".

17.1. Introduction. In this section we discuss the notion of *derivative* (i.e. *differential*) for functions $f : D\,(\subset \mathbb{R}^n) \to \mathbb{R}$. In the next chapter we consider the case for functions $f : D\,(\subset \mathbb{R}^n) \to \mathbb{R}^m$.

If $m = 1$ and $n = 1$ or 2, we can sometimes represent such a function by drawing its graph. In case $n = 2$ (or perhaps $n = 3$) we can draw the level sets, as is done in Section 17.6.

Convention Unless stated otherwise, we consider functions $f : D(\subset \mathbb{R}^n) \to \mathbb{R}$ where the domain D is *open*. This implies that for any $\mathbf{x} \in D$ there exists $r > 0$ such that $B_r(\mathbf{x}) \subset D$.



Most of the following applies to more general domains D by taking one-sided, or otherwise restricted, limits. No essentially new ideas are involved.

17.2. Algebraic Preliminaries. The *inner product* in $\mathbb{R}^n$ is represented by

$$\mathbf{y} \cdot \mathbf{x} = y^1 x^1 + \ldots + y^n x^n$$

where $\mathbf{y} = (y^1, \ldots, y^n)$ and $\mathbf{x} = (x^1, \ldots, x^n)$.

For each *fixed* $\mathbf{y} \in \mathbb{R}^n$ the inner product enables us to define a *linear function*

$$L_{\mathbf{y}} = L : \mathbb{R}^n \to \mathbb{R}$$

given by

$$L(\mathbf{x}) = \mathbf{y} \cdot \mathbf{x}.$$

Conversely, we have the following.

PROPOSITION 17.2.1. *For any linear function*

$$L : \mathbb{R}^n \to \mathbb{R}$$

there exists a **unique** $\mathbf{y} \in \mathbb{R}^n$ *such that*

$$L(\mathbf{x}) = \mathbf{y} \cdot \mathbf{x} \quad \forall \mathbf{x} \in \mathbb{R}^n. \tag{37}$$

The components of $\mathbf{y}$ *are given by* $y^i = L(\mathbf{e_i})$.

PROOF. Suppose $L:\mathbb{R}^n \to \mathbb{R}$ is linear. *Define* $\mathbf{y} = (y^1, \dots, y^n)$ by

$$y^i = L(\mathbf{e_i}) \quad i = 1, \dots, n.$$

Then

$$\begin{aligned} L(\mathbf{x}) &= L(x^1\mathbf{e_1} + \cdots + x^n\mathbf{e_n}) \\ &= x^1 L(\mathbf{e_1}) + \cdots + x^n L(\mathbf{e_n}) \\ &= x^1 y^1 + \cdots + x^n y^n \\ &= \mathbf{y} \cdot \mathbf{x}. \end{aligned}$$

This proves the *existence* of $\mathbf{y}$ satisfying (37).

The *uniqueness* of $\mathbf{y}$ follows from the fact that if (37) is true for some $\mathbf{y}$, then on choosing $\mathbf{x} = \mathbf{e_i}$ it follows we *must* have

$$L(\mathbf{e_i}) = y^i \quad i = 1, \dots, n.$$

□

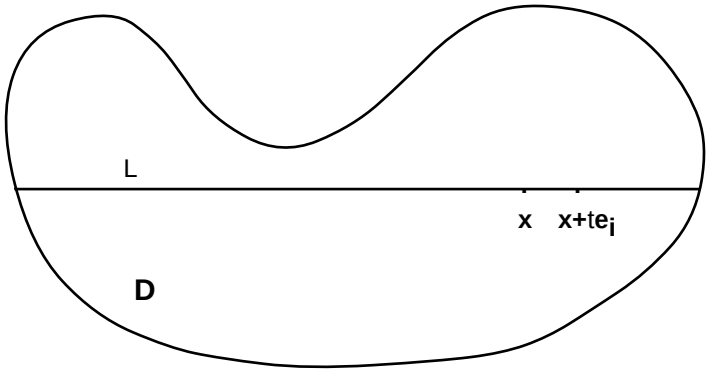
Note that if L is the *zero operator*, i.e. if $L(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^n$, then the vector $\mathbf{y}$ corresponding to L is the zero vector.

17.3. Partial Derivatives.

DEFINITION 17.3.1. The ith *partial derivative* of f at $\mathbf{x}$ is defined by

$$\begin{aligned} \frac{\partial f}{\partial x^i}(\mathbf{x}) &= \lim_{t\to 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t} \\ &= \lim_{t\to 0} \frac{f(x^1, \dots, x^i + t, \dots, x^n) - f(x^1, \dots, x^i, \dots, x^n)}{t}, \end{aligned} \tag{38}$$

provided the limit exists. The notation $D_i f(\mathbf{x})$ is also used.



Thus $\dfrac{\partial f}{\partial x^i}(\mathbf{x})$ is just the usual derivative at $t = 0$ of the *real-valued* function g defined by $g(t) = f(x^1, \dots, x^i + t, \dots, x^n)$. Think of g as being defined along the line L, with $t = 0$ corresponding to the point $\mathbf{x}$.

17.4. Directional Derivatives.

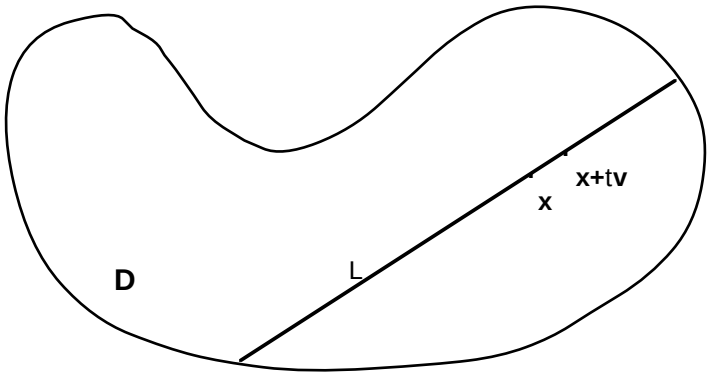
DEFINITION 17.4.1. The *directional derivative* of f at $\mathbf{x}$ in the direction $\mathbf{v} \neq \mathbf{0}$ is defined by

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{t\to 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}, \tag{39}$$

provided the limit exists.

It follows immediately from the definitions that

$$\frac{\partial f}{\partial x^i}(\mathbf{x}) = D_{\mathbf{e}_i} f(\mathbf{x}). \tag{40}$$



Note that $D_{\mathbf{v}} f(\mathbf{x})$ is just the usual derivative at $t = 0$ of the *real-valued* function g defined by $g(t) = f(\mathbf{x} + t\mathbf{v})$. As before, think of the function g as being defined along the line L in the previous diagram.

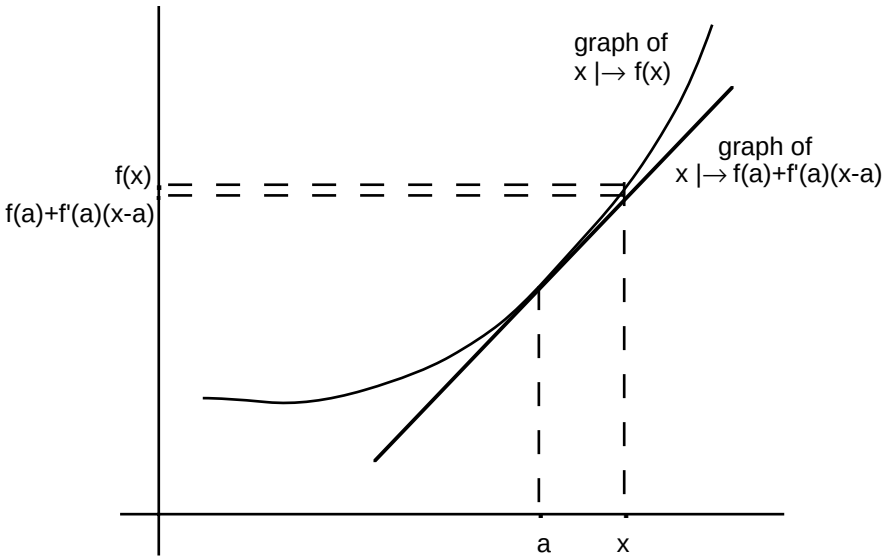
Thus we interpret $D_{\mathbf{v}} f(\mathbf{x})$ as the rate of change of f at $\mathbf{x}$ in the direction $\mathbf{v}$; at least in the case $\mathbf{v}$ is a unit vector.

Exercise: Show that $D_{\alpha\mathbf{v}} f(\mathbf{x}) = \alpha D_{\mathbf{v}} f(\mathbf{x})$ for any real number α.

17.5. The Differential (or Derivative).

Motivation Suppose $f : I\,(\subset \mathbb{R}) \to \mathbb{R}$ is differentiable at $a \in I$. Then $f'(a)$ can be used to define the *best linear approximation* to $f(x)$ for x near a. Namely:

$$f(x) \approx f(a) + f'(a)(x - a). \tag{41}$$



Note that the right-hand side of (41) is linear in x. (More precisely, the right side is a polynomial in x of degree one.)

The *error*, or difference between the two sides of (41), approaches zero as $x \to a$, *faster* than $|x - a| \to 0$. More precisely

$$\frac{\Big|f(x) - \Big(f(a) + f'(a)(x-a)\Big)\Big|}{|x-a|} \quad = \quad \left|\frac{f(x) - \Big(f(a) + f'(a)(x-a)\Big)}{x-a}\right|$$

$$= \left| \frac{f(x) - f(a)}{x - a} - f'(a) \right|$$

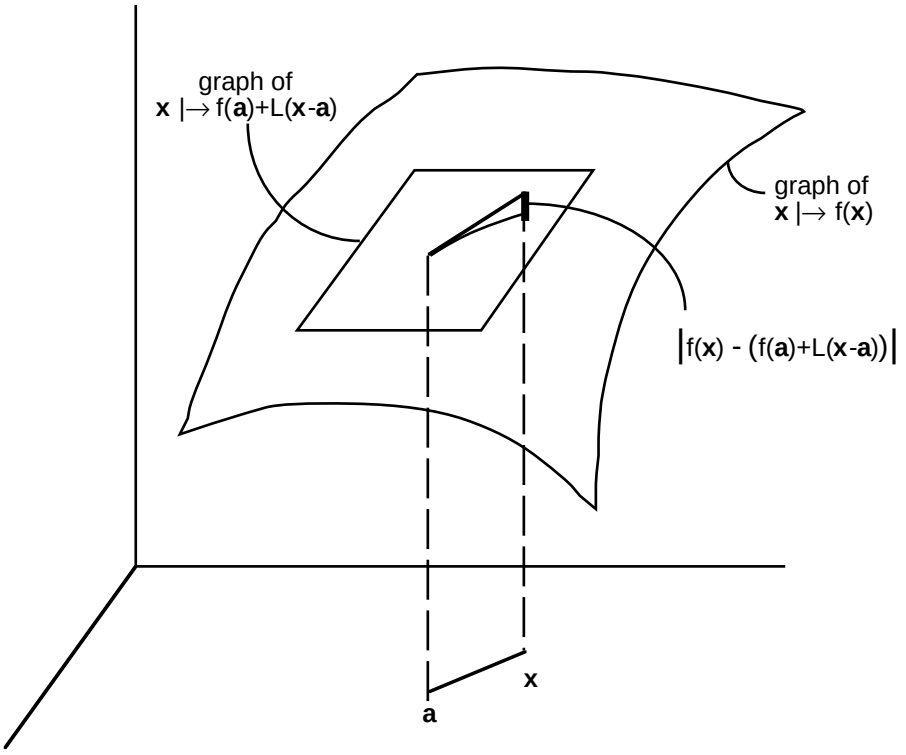
$$\to 0 \quad \text{as } x \to a. \tag{42}$$

We make this the basis for the next definition in the case $n > 1$.

DEFINITION 17.5.1. Suppose $f : D\,(\subset \mathbb{R}^n) \to \mathbb{R}$. Then f is *differentiable* at $\mathbf{a} \in D$ if there is a linear function $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$\frac{\left| f(\mathbf{x}) - \Big(f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a}) \Big) \right|}{|\mathbf{x} - \mathbf{a}|} \to 0 \quad \text{as } \mathbf{x} \to \mathbf{a}. \tag{43}$$

The linear function L is denoted by $f'(\mathbf{a})$ or $df(\mathbf{a})$ and is called the *derivative* or *differential* of f at $\mathbf{a}$. (We will see in Proposition 17.5.2 that if L exists, it is uniquely determined by this definition.)



The idea is that the graph of $\mathbf{x} \mapsto f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is "tangent" to the graph of $f(\mathbf{x})$ at the point $\Big(\mathbf{a}, f(\mathbf{a})\Big)$.

Notation: We write $\langle df(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle$ for $L(\mathbf{x} - \mathbf{a})$, and read this as "$df$ at $\mathbf{a}$ applied to $\mathbf{x} - \mathbf{a}$". We think of $df(\mathbf{a})$ as a linear transformation (or function) which operates on vectors $\mathbf{x} - \mathbf{a}$ whose "base" is at $\mathbf{a}$.

The next proposition gives the connection between the differential operating on a vector $\mathbf{v}$, and the directional derivative in the direction corresponding to $\mathbf{v}$. In particular, it shows that the differential is *uniquely defined* by Definition 17.5.1.

Temporarily, we let $df(\mathbf{a})$ be *any* linear map satisfying the definition for the differential of f at $\mathbf{a}$.

PROPOSITION 17.5.2. *Let $\mathbf{v} \in \mathbb{R}^n$ and suppose f is differentiable at $\mathbf{a}$.*
Then $D_{\mathbf{v}} f(\mathbf{a})$ exists and

$$\langle df(\mathbf{a}), \mathbf{v} \rangle = D_{\mathbf{v}} f(\mathbf{a}).$$

In particular, the differential is unique.

PROOF. Let $\mathbf{x} = \mathbf{a} + t\mathbf{v}$ in (43). Then

$$\lim_{t\to 0} \frac{\left| f(\mathbf{a}+t\mathbf{v}) - \Big(f(\mathbf{a}) + \langle df(\mathbf{a}), t\mathbf{v}\rangle\Big)\right|}{t} = 0.$$

Hence

$$\lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t} - \langle df(\mathbf{a}), \mathbf{v}\rangle = \mathbf{0}.$$

Thus

$$D_{\mathbf{v}} f(\mathbf{a}) = \langle df(\mathbf{a}), \mathbf{v}\rangle$$

as required. □

Thus $\langle df(\mathbf{a}), \mathbf{v}\rangle$ is just the directional derivative at $\mathbf{a}$ in the direction $\mathbf{v}$.

The next result shows $df(\mathbf{a})$ is the linear map given by the row vector of partial derivatives of f at $\mathbf{a}$.

COROLLARY 17.5.3. *Suppose f is differentiable at $\mathbf{a}$. Then for* any *vector $\mathbf{v}$,*

$$\langle df(\mathbf{a}), \mathbf{v}\rangle = \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x^i}(\mathbf{a}).$$

That is, $df(\mathbf{a})$ is the row vector $\left[\frac{\partial f}{\partial x^1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x^n}(\mathbf{a})\right]$.

PROOF.

$$\begin{aligned}
\langle df(\mathbf{a}), \mathbf{v}\rangle &= \langle df(\mathbf{a}),\ v^1\mathbf{e}_1 + \cdots + v^n\mathbf{e}_n\rangle \\
&= v^1\langle df(\mathbf{a}), \mathbf{e}_1\rangle + \cdots + v^n\langle df(\mathbf{a}), \mathbf{e}_n\rangle \\
&= v^1 D_{\mathbf{e}_1} f(\mathbf{a}) + \cdots + v^n D_{\mathbf{e}_n} f(\mathbf{a}) \\
&= v^1 \frac{\partial f}{\partial x^1}(a) + \cdots + v^n \frac{\partial f}{\partial x^n}(\mathbf{a}).
\end{aligned}$$

□

EXAMPLE 17.5.4. Let $f(x, y, z) = x^2 + 3xy^2 + y^3 z + z$.

Then

$$\begin{aligned}
\langle df(\mathbf{a}), \mathbf{v}\rangle &= v_1 \frac{\partial f}{\partial x}(\mathbf{a}) + v_2 \frac{\partial f}{\partial y}(\mathbf{a}) + v_3 \frac{\partial f}{\partial z}(\mathbf{a}) \\
&= v_1(2a_1 + 3{a_2}^2) + v_2(6a_1a_2 + 3{a_2}^2 a_3) + v_3({a_2}^3 + 1).
\end{aligned}$$

Thus $df(\mathbf{a})$ is the linear map corresponding to the row vector $(2a_1 + 3{a_2}^2, 6a_1a_2 + 3{a_2}^2a_3, {a_2}^3 + 1)$.

If $\mathbf{a} = (1, 0, 1)$ then $\langle df(\mathbf{a}), \mathbf{v}\rangle = 2v_1 + v_3$. Thus $df(\mathbf{a})$ is the linear map corresponding to the row vector $(2, 0, 1)$.

If $\mathbf{a} = (1, 0, 1)$ and $\mathbf{v} = \mathbf{e}_1$ then $\langle df(1, 0, 1), \mathbf{e}_1\rangle = \dfrac{\partial f}{\partial x}(1, 0, 1) = 2$.

DEFINITION 17.5.5. (Rates of Convergence) If a function $\psi(\mathbf{x})$ has the property that

$$\frac{|\psi(\mathbf{x})|}{|\mathbf{x} - \mathbf{a}|} \to 0 \text{ as } \mathbf{x} \to \mathbf{a},$$

then we say "$|\psi(\mathbf{x})| \to 0$ as $\mathbf{x} \to \mathbf{a}$, faster than $|\mathbf{x} - \mathbf{a}| \to 0$". We write $o(|\mathbf{x} - \mathbf{a}|)$ for $\psi(\mathbf{x})$, and read this as "little *oh* of $|\mathbf{x} - \mathbf{a}|$".

If

$$\frac{|\psi(\mathbf{x})|}{|\mathbf{x} - \mathbf{a}|} \leq M \quad \forall |\mathbf{x} - \mathbf{a}| < \epsilon,$$

for some M and some $\epsilon > 0$, i.e. if $\dfrac{|\psi(\mathbf{x})|}{|\mathbf{x}-\mathbf{a}|}$ is bounded as $\mathbf{x} \to \mathbf{a}$, then we say "$|\psi(\mathbf{x})| \to 0$ as $\mathbf{x} \to \mathbf{a}$, at least as fast as $|\mathbf{x}-\mathbf{a}| \to 0$". We write $O(|\mathbf{x}-\mathbf{a}|)$ for $\psi(\mathbf{x})$, and read this as "big *oh* of $|\mathbf{x}-\mathbf{a}|$".

EXAMPLE 17.5.6. we can write

$$o(|x-a|) \text{ for } |x-a|^{3/2},$$

and

$$O(|x-a|) \text{ for } \sin(x-a).$$

Clearly, if $\psi(\mathbf{x})$ can be written as $o(|\mathbf{x}-\mathbf{a}|)$ then it can also be written as $O(|\mathbf{x}-\mathbf{a}|)$, but the converse may not be true as the above example shows.

The next proposition gives an equivalent definition for the differential of a function.

PROPOSITION 17.5.7. *If f is differentiable at $\mathbf{a}$ then*

$$f(\mathbf{x}) = f(\mathbf{a}) + \langle df(\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle + \psi(\mathbf{x}),$$

where $\psi(\mathbf{x}) = o(|\mathbf{x}-\mathbf{a}|)$.

Conversely, suppose

$$f(\mathbf{x}) = f(\mathbf{a}) + L(\mathbf{x}-\mathbf{a}) + \psi(\mathbf{x}),$$

where $L\colon \mathbb{R}^n \to \mathbb{R}$ is linear and $\psi(\mathbf{x}) = o(|\mathbf{x}-\mathbf{a}|)$. Then f is differentiable at $\mathbf{a}$ and $df(\mathbf{a}) = L$.

PROOF. Suppose f is differentiable at $\mathbf{a}$. Let

$$\psi(\mathbf{x}) = f(\mathbf{x}) - \Big(f(\mathbf{a}) + \langle df(\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle\Big).$$

Then

$$f(\mathbf{x}) = f(\mathbf{a}) + \langle df(\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle + \psi(\mathbf{x}),$$

and $\psi(\mathbf{x}) = o(|\mathbf{x}-\mathbf{a}|)$ from Definition 17.5.1.

Conversely, suppose

$$f(\mathbf{x}) = f(\mathbf{a}) + L(\mathbf{x}-\mathbf{a}) + \psi(\mathbf{x}),$$

where $L\colon \mathbb{R}^n \to \mathbb{R}$ is linear and $\psi(\mathbf{x}) = o(|\mathbf{x}-\mathbf{a}|)$. Then

$$\frac{f(\mathbf{x}) - \Big(f(\mathbf{a}) + L(\mathbf{x}-\mathbf{a})\Big)}{|\mathbf{x}-\mathbf{a}|} = \frac{\psi(\mathbf{x})}{|\mathbf{x}-\mathbf{a}|} \to 0 \quad \text{as } \mathbf{x} \to \mathbf{a},$$

and so f is differentiable at $\mathbf{a}$ and $df(\mathbf{a}) = L$. □

Finally we have:

PROPOSITION 17.5.8. *If $f, g\colon D\,(\subset \mathbb{R}^n) \to \mathbb{R}$ are differentiable at $\mathbf{a} \in D$, then so are αf and $f+g$. Moreover,*

$$\begin{aligned} d(\alpha f)(\mathbf{a}) &= \alpha df(\mathbf{a}), \\ d(f+g)(\mathbf{a}) &= df(\mathbf{a}) + dg(\mathbf{a}). \end{aligned}$$

PROOF. This is straightforward (*exercise*) from Proposition 17.5.7. □

The previous proposition corresponds to the fact that the partial derivatives for $f+g$ are the sum of the partial derivatives corresponding to f and g respectively. Similarly for αf [51].

[51] We cannot establish the differentiability of $f+g$ (or αf) this way, since the existence of the partial derivatives does not imply differentiability.

17.6. The Gradient. Strictly speaking, $df(\mathbf{a})$ is a *linear operator* on vectors in $\mathbb{R}^n$ (where, for convenience, we think of these vectors as having their "base at $\mathbf{a}$").

We saw in Section 17.2 that every linear operator from $\mathbb{R}^n$ to $\mathbb{R}$ corresponds to a unique vector in $\mathbb{R}^n$. In particular, the vector corresponding to the differential at $\mathbf{a}$ is called the *gradient* at $\mathbf{a}$.

DEFINITION 17.6.1. Suppose f is differentiable at $\mathbf{a}$. The *vector* $\nabla f(\mathbf{a}) \in \mathbb{R}^n$ (uniquely) determined by

$$\nabla f(\mathbf{a}) \cdot \mathbf{v} = \langle df(\mathbf{a}), \mathbf{v} \rangle \quad \forall \mathbf{v} \in \mathbb{R}^n,$$

is called the *gradient* of f at $\mathbf{a}$.

PROPOSITION 17.6.2. *If f is differentiable at $\mathbf{a}$, then*

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x^1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x^n}(\mathbf{a}) \right).$$

PROOF. It follows from Proposition 17.2.1 that the components of $\nabla f(\mathbf{a})$ are $\langle df(\mathbf{a}), \mathbf{e}_i \rangle$, i.e. $\dfrac{\partial f}{\partial x^i}(\mathbf{a})$. □

EXAMPLE 17.6.3. For the example in Section 17.5 we have

$$\begin{aligned} \nabla f(\mathbf{a}) &= (2a_1 + 3{a_2}^2,\ 6a_1a_2 + 3{a_2}^2a_3,\ a_2^3 + 1), \\ \nabla f(1,0,1) &= (2,0,1). \end{aligned}$$

PROPOSITION 17.6.4. *Suppose f is differentiable at $\mathbf{x}$. Then the directional derivatives at $\mathbf{x}$ are given by*

$$D_{\mathbf{v}} f(\mathbf{x}) = \mathbf{v} \cdot \nabla f(\mathbf{x}).$$

The unit vector $\mathbf{v}$ for which this is a maximum is $\mathbf{v} = \nabla f(\mathbf{x})/|\nabla f(\mathbf{x})|$ (assuming $|\nabla f(\mathbf{x})| \neq 0$), and the directional derivative in this direction is $|\nabla f(\mathbf{x})|$.

PROOF. From Definition 17.6.1 and Proposition 17.5.2 it follows that

$$\nabla f(\mathbf{x}) \cdot \mathbf{v} = \langle df(\mathbf{x}), \mathbf{v} \rangle = D_{\mathbf{v}} f(\mathbf{x})$$

This proves the first claim.

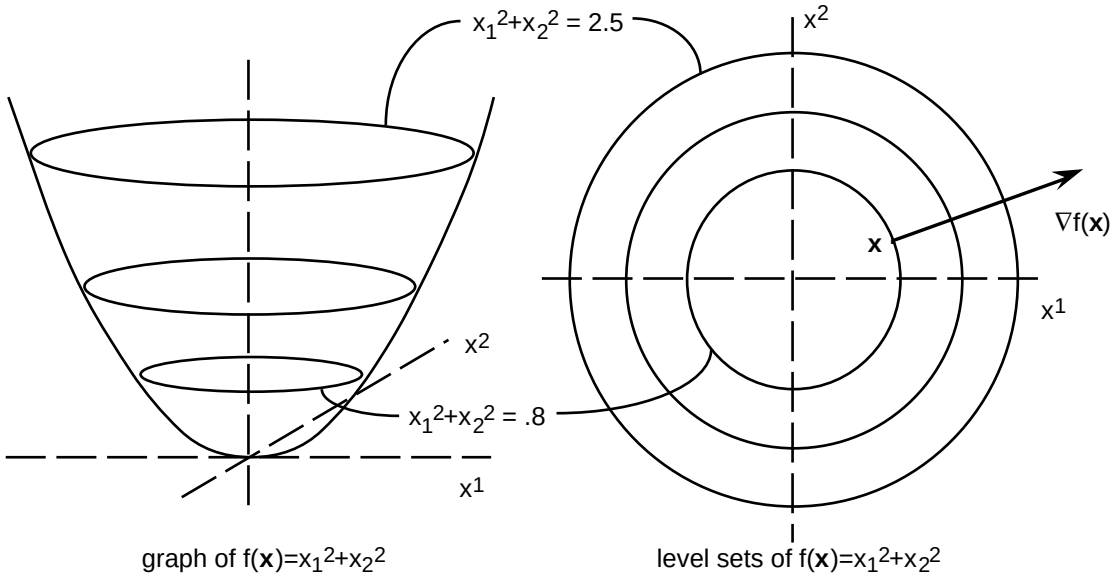
Now suppose $\mathbf{v}$ is a unit vector. From the Cauchy-Schwartz Inequality we have

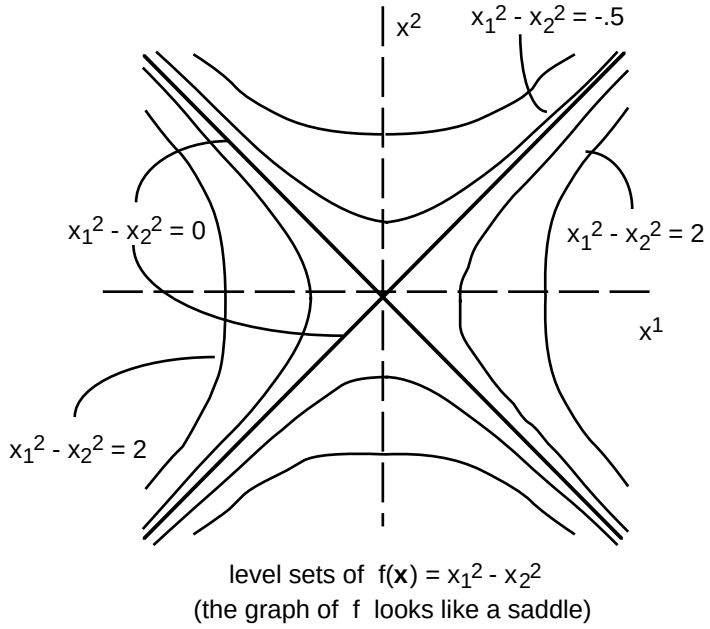
$$\nabla f(\mathbf{x}) \cdot \mathbf{v} \leq |\nabla f(\mathbf{x})|. \tag{44}$$

By the condition for equality in the Cauchy-Schwartz Inequality, equality holds in (44) iff $\mathbf{v}$ is a *positive* multiple of $\nabla f(\mathbf{x})$. Since $\mathbf{v}$ is a unit vector, this is equivalent to $\mathbf{v} = \nabla f(\mathbf{x})/|\nabla f(\mathbf{x})|$. The left side of (44) is then $|\nabla f(\mathbf{x})|$. □

DEFINITION 17.6.5. If $f : \mathbb{R}^n \to \mathbb{R}$ then the *level set* through $\mathbf{x}$ is $\{\mathbf{y}\colon\ \mathrm{f}(\mathbf{y}) = \mathrm{f}(\mathbf{x})\ \}$.

For example, the contour lines on a map are the level sets of the height function.





DEFINITION 17.6.6. A vector $\mathbf{v}$ is *tangent* at $\mathbf{x}$ to the level set S through $\mathbf{x}$ if

$$D_{\mathbf{v}} f(\mathbf{x}) = 0.$$

This is a reasonable definition, since f is *constant* on S, and so the rate of change of f in any direction tangent to S should be zero.

PROPOSITION 17.6.7. *Suppose f is differentiable at $\mathbf{x}$. Then $\nabla f(\mathbf{x})$ is orthogonal to all vectors which are tangent at $\mathbf{x}$ to the level set through $\mathbf{x}$.*

PROOF. This is immediate from the previous Definition and Proposition 17.6.4. □

In the previous proposition, we say $\nabla f(\mathbf{x})$ *is orthogonal to the level set through* $\mathbf{x}$.

17.7. Some Interesting Examples.

EXAMPLE 17.7.1. *An example where the partial derivatives exist but the other directional derivatives do not exist.*

Let

$$f(x,y) = (xy)^{1/3}.$$

Then

1. $\dfrac{\partial f}{\partial x}(0,0) = 0$ since $f = 0$ on the x-axis;
2. $\dfrac{\partial f}{\partial y}(0,0) = 0$ since $f = 0$ on the y-axis;
3. Let $\mathbf{v}$ be *any* vector. Then

$$\begin{aligned} D_{\mathbf{v}} f(0,0) &= \lim_{t\to 0} \frac{f(t\mathbf{v}) - f(0,0)}{t} \\ &= \lim_{t\to 0} \frac{t^{2/3}(v_1 v_2)^{1/3}}{t} \\ &= \lim_{t\to 0} \frac{(v_1 v_2)^{1/3}}{t^{1/3}}. \end{aligned}$$

This limit does *not* exist, unless $v_1 = 0$ or $v_2 = 0$.

EXAMPLE 17.7.2. *An example where the directional derivatives at some point all exist, but the function is not differentiable at the point.*

Let

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

Let $\mathbf{v} = (v_1, v_2)$ be any non-zero vector. Then

$$\begin{aligned} D_{\mathbf{v}} f(0,0) &= \lim_{t\to 0} \frac{f(t\mathbf{v}) - f(0,0)}{t} \\ &= \lim_{t\to 0} \frac{\dfrac{t^3 v_1 {v_2}^2}{t^2 {v_1}^2 + t^4 {v_2}^4} - 0}{t} \\ &= \lim_{t\to 0} \frac{v_1 {v_2}^2}{{v_1}^2 + t^2 {v_2}^4} \\ &= \begin{cases} {v_2}^2/v_1 & v_1 \neq 0 \\ 0 & v_1 = 0 \end{cases} \end{aligned} \tag{45}$$

Thus the directional derivatives $D_{\mathbf{v}} f(0,0)$ exist for all $\mathbf{v}$, and are given by (45).

In particular

$$\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0. \tag{46}$$

But if f were differentiable at $(0,0)$, then we could compute any directional derivative from the partial drivatives. Thus for any vector $\mathbf{v}$ we would have

$$\begin{aligned} D_{\mathbf{v}} f(0,0) &= \langle df(0,0), \mathbf{v} \rangle \\ &= v_1 \frac{\partial f}{\partial x}(0,0) + v_2 \frac{\partial f}{\partial y}(0,0) \\ &= 0 \quad \text{from (46)} \end{aligned}$$

This contradicts (45).

EXAMPLE 17.7.3. *An Example where the directional derivatives at a point all exist, but the function is not continuous at the point*

Take the same example as in Example 17.7.2. Approach the origin along the curve $x = \lambda^2$, $y = \lambda$. Then

$$\lim_{\lambda \to 0} f(\lambda^2, \lambda) = \lim_{\lambda \to 0} \frac{\lambda^4}{2\lambda^4} = \frac{1}{2}.$$

But if we approach the origin along any straight line of the form $(\lambda v_1, \lambda v_2)$, then we can check that the corresponding limit is 0.

Thus it is impossible to define f at $(0, 0)$ in order to make f continuous there.

17.8. Differentiability Implies Continuity. Despite Example 17.7.3 we have the following result.

PROPOSITION 17.8.1. *If f is differentiable at $\mathbf{a}$, then it is continuous at $\mathbf{a}$.*

PROOF. Suppose f is differentiable at $\mathbf{a}$. Then

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\mathbf{a})(x_i - a^i) + o(|\mathbf{x} - \mathbf{a}|).$$

Since $x^i - a^i \to 0$ and $o(|\mathbf{x} - \mathbf{a}|) \to 0$ as $\mathbf{x} \to \mathbf{a}$, it follows that $f(\mathbf{x}) \to f(\mathbf{a})$ as $\mathbf{x} \to \mathbf{a}$. That is, f is continuous at $\mathbf{a}$. □

17.9. Mean Value Theorem and Consequences.

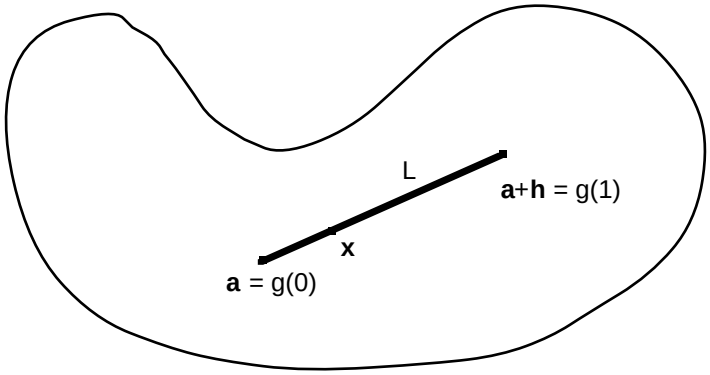
THEOREM 17.9.1. *Suppose f is continuous at all points on the line segment L joining $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$; and is differentiable at all points on L, except possibly at the end points.*

Then

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = \langle df(\mathbf{x}), \mathbf{h} \rangle \tag{47}$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\mathbf{x})\, h^i \tag{48}$$

for some $\mathbf{x} \in L$, $\mathbf{x}$ not an endpoint of L.



PROOF. Note that (48) follows immediately from (47) by Corollary 17.5.3.

Define the *one variable* function g by

$$g(t) = f(\mathbf{a} + t\mathbf{h}).$$

Then g is continuous on [0,1] (being the composition of the continuous functions $t \mapsto \mathbf{a} + t\mathbf{h}$ and $\mathbf{x} \mapsto f(\mathbf{x})$). Moreover,

$$g(0) = f(\mathbf{a}),\ g(1) = f(\mathbf{a} + \mathbf{h}). \tag{49}$$

We next show that g is differentiable and compute its derivative.

If $0 < t < 1$, then f is differentiable at $\mathbf{a} + t\mathbf{h}$, and so

$$0 = \lim_{|\mathbf{w}|\to 0} \frac{f(\mathbf{a}+t\mathbf{h}+\mathbf{w}) - f(\mathbf{a}+t\mathbf{h}) - \langle df(\mathbf{a}+t\mathbf{h}), \mathbf{w}\rangle}{|\mathbf{w}|}. \tag{50}$$

Let $\mathbf{w} = s\mathbf{h}$ where s is a small real number, positive or negative. Since $|\mathbf{w}| = \pm s|\mathbf{h}|$, and since we may assume $\mathbf{h} \neq \mathbf{0}$ (as otherwise (47) is trivial), we see from (50) that

$$\begin{aligned} 0 &= \lim_{s\to 0} \frac{f\Big((\mathbf{a}+(t+s)\mathbf{h}\Big) - f(\mathbf{a}+t\mathbf{h}) - \langle df(\mathbf{a}+t\mathbf{h}), s\mathbf{h}\rangle}{s} \\ &= \lim_{s\to 0} \left(\frac{g(t+s)-g(t)}{s} - \langle df(\mathbf{a}+t\mathbf{h}), \mathbf{h}\rangle \right), \end{aligned}$$

using the linearity of $df(\mathbf{a}+t\mathbf{h})$.

Hence $g'(t)$ exists for $0 < t < 1$, and moreover

$$g'(t) = \langle df(\mathbf{a}+t\mathbf{h}), \mathbf{h}\rangle. \tag{51}$$

By the usual Mean Value Theorem for a function of one variable, applied to g, we have

$$g(1) - g(0) = g'(t) \tag{52}$$

for some $t \in (0,1)$.

Substituting (49) and (51) in (52), the required result (47) follows. □

If the norm of the gradient vector of f is bounded by M, then it is not surprising that the difference in value between $f(\mathbf{a})$ and $f(\mathbf{a}+\mathbf{h})$ is bounded by $M|\mathbf{h}|$. More precisely.

COROLLARY 17.9.2. *Assume the hypotheses of the previous theorem and suppose* $|\nabla f(\mathbf{x})| \leq M$ *for all* $\mathbf{x} \in L$. *Then*

$$|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})| \leq M|\mathbf{h}|$$

PROOF. From the previous theorem

$$\begin{aligned} |f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})| &= |\langle df(\mathbf{x}), \mathbf{h}\rangle| \quad \text{for some } \mathbf{x} \in L \\ &= |\nabla f(\mathbf{x}) \cdot \mathbf{h}| \\ &\leq |\nabla f(\mathbf{x})|\,|\mathbf{h}| \\ &\leq M|\mathbf{h}|. \end{aligned}$$

□

COROLLARY 17.9.3. *Suppose* $\Omega \subset \mathbb{R}^n$ *is open and* **connected** *and* $f : \Omega \to \mathbb{R}$. *Suppose* f *is differentiable in* Ω *and* $df(\mathbf{x}) = 0$ *for all* $\mathbf{x} \in \Omega$[52].

Then f *is constant on* Ω.

PROOF. Choose any $\mathbf{a} \in \Omega$ and suppose $f(\mathbf{a}) = \alpha$. Let

$$E = \{\mathbf{x} \in \Omega : f(\mathbf{x}) = \alpha\}.$$

Then E is non-empty (as $\mathbf{a} \in E$). We will prove E is both open and closed *in* Ω. Since Ω is connected, this will imply that E is *all* of Ω. This establishes the result.

To see E is *open*[53], suppose $\mathbf{x} \in E$ and choose $r > 0$ so that $B_r(\mathbf{x}) \subset \Omega$.

If $\mathbf{y} \in B_r(\mathbf{x})$, then from (47) for some $\mathbf{u}$ between $\mathbf{x}$ and $\mathbf{y}$,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= \langle df(\mathbf{u}), \mathbf{y}-\mathbf{x}\rangle \\ &= 0, \quad \text{by hypothesis.} \end{aligned}$$

[52]Equivalently, $\nabla f(\mathbf{x}) = \mathbf{0}$ in Ω.

[53]Being open in Ω and being open in $\mathbb{R}^n$ is the same for subsets of Ω, since we are assuming Ω is itself open in $\mathbb{R}^n$.

Thus $f(\mathbf{y}) = f(\mathbf{x})\,(= \alpha)$, and so $\mathbf{y} \in E$.

Hence $B_r(\mathbf{x}) \subset E$ and so E is open.

To show that E is *closed* in Ω, it is sufficient to show that $E^c = \{\mathbf{y} : f(\mathbf{x}) \neq \alpha\}$ is open in Ω.

From Proposition 17.8.1 we know that f is continuous. Since we have $E^c = f^{-1}[\mathbb{R} \setminus \{\alpha\}]$ and $\mathbb{R} \setminus \{\alpha\}$ is open, it follows that E^c is open in Ω. Hence E is closed in Ω, as required.

Since $E \neq \emptyset$, and E is both open and closed in Ω, it follows $E = \Omega$ (as Ω is connected).

In other words, f is constant $(= \alpha)$ on Ω. □

17.10. Continuously Differentiable Functions. We saw in Section 17.7, Example (2), that the partial derivatives (and even all the directional derivatives) of a function can exist without the function being differentiable.

However, we do have the following important theorem:

THEOREM 17.10.1. *Suppose $f : \Omega\,(\subset \mathbb{R}^n) \to \mathbb{R}$ where Ω is open. If the partial derivatives of f exist and are continuous at every point in Ω, then f is differentiable everywhere in Ω.*

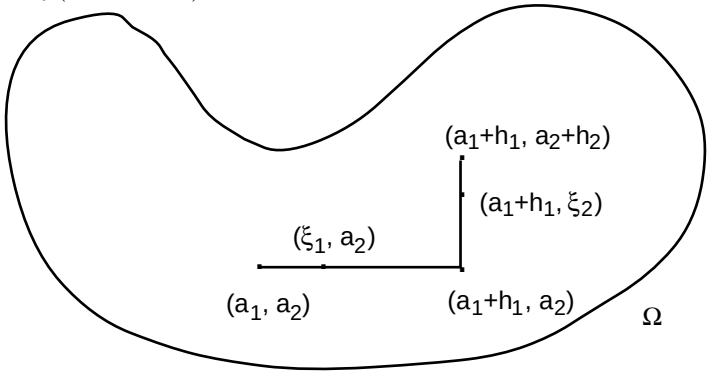
REMARK 17.10.1. If the partial derivatives of f exist in some neighbourhood of, and are continuous at, a *single point*, it does not necessarily follow that f is differentiable at that point. The hypotheses of the theorem need to hold at *all* points in some *open* set Ω.

PROOF OF THEOREM. We prove the theorem in case $n = 2$ (the proof for $n > 2$ is only notationally more complicated).

Suppose that the partial derivatives of f exist and are continuous in Ω. Then if $\mathbf{a} \in \Omega$ and $\mathbf{a} + \mathbf{h}$ is sufficiently close to $\mathbf{a}$,

$$\begin{aligned} f(a^1 + h^1, a^2 + h^2) &= f(a^1, a^2) \\ &\quad + f(a^1 + h^1, a^2) - f(a^1, a^2) \\ &\quad + f(a^1 + h^1, a^2 + h^2) - f(a^1 + h^1, a^2) \\ &= f(a^1, a^2) + \frac{\partial f}{\partial x^1}(\xi^1, a^2)h^1 + \frac{\partial f}{\partial x^2}(a^1 + h^1, \xi^2)h^2, \end{aligned}$$

for some ξ^1 between a^1 and $a^1 + h^1$, and some ξ^2 between a^2 and $a^2 + h^2$. The first partial derivative comes from applying the usual Mean Value Theorem, for a function of *one* variable, to the function $f(x^1, a^2)$ obtained by fixing a^2 and taking x^1 as a variable. The second partial derivative is similarly obtained by considering the function $f(a^1 + h^1, x^2)$, where $a^1 + h^1$ is fixed and x^2 is variable.



Hence

$$f(a^1 + h^1, a^2 + h^2) = f(a^1, a^2) + \frac{\partial f}{\partial x^1}(a^1, a^2)h^1 + \frac{\partial f}{\partial x^2}(a^1, a^2)h^2$$

$$
\begin{aligned}
&+\left(\frac{\partial f}{\partial x^1}(\xi^1,a^2)-\frac{\partial f}{\partial x^1}(a^1,a^2)\right)h^1\\
&+\left(\frac{\partial f}{\partial x^2}(a^1+h^1,\xi^2)-\frac{\partial f}{\partial x^2}(a^1,a^2)\right)h^2\\
=\;& f(a^1,a^2)+L(\mathbf{h})+\psi(\mathbf{h}),\ \text{say.}
\end{aligned}
$$

Here L is the linear map defined by

$$
\begin{aligned}
L(\mathbf{h}) &= \frac{\partial f}{\partial x^1}(a^1,a^2)h^1+\frac{\partial f}{\partial x^2}(a^1,a^2)h^2\\
&= \begin{bmatrix} \dfrac{\partial f}{\partial x^1}(a^1,a^2) & \dfrac{\partial f}{\partial x^2}(a^1,a^2) \end{bmatrix}\begin{bmatrix} h^1\\ h^2 \end{bmatrix}.
\end{aligned}
$$

Thus L is represented by the previous 1×2 matrix.

We claim that the *error term*

$$
\psi(\mathbf{h})=\left(\frac{\partial f}{\partial x^1}(\xi^1,a^2)-\frac{\partial f}{\partial x^1}(a^1,a^2)\right)h^1+\left(\frac{\partial f}{\partial x^2}(a^1+h^1,\xi^2)-\frac{\partial f}{\partial x^2}(a^1,a^2)\right)h^2
$$

can be written as $o(|\mathbf{h}|)$

This follows from the facts:

1. $\dfrac{\partial f}{\partial x^1}(\xi^1,a^2)\to\dfrac{\partial f}{\partial x^1}(a^1,a^2)$ as $\mathbf{h}\to\mathbf{0}$ (by continuity of the partial derivatives),
2. $\dfrac{\partial f}{\partial x^2}(a^1+h^1,\xi^2)\to\dfrac{\partial f}{\partial x^2}(a^1,a^2)$ as $\mathbf{h}\to\mathbf{0}$ (again by continuity of the partial derivatives),
3. $|h^1|\le|\mathbf{h}|$, $|h^2|\le|\mathbf{h}|$.

It now follows from Proposition 17.5.7 that f is differentiable at a, and the differential of f is given by the previous 1×2 matrix of partial derivatives.

Since $\mathbf{a}\in\Omega$ is arbitrary, this completes the proof. □

Definition 17.10.2. If the partial derivatives of f exist and are continuous in the open set Ω, we say f is a C^1 (or *continuously differentiable*) function on Ω. One writes $f\in C^1(\Omega)$.

It follows from the previous Theorem that if $f\in C^1(\Omega)$ then f is indeed differentiable in Ω. *Exercise:* The converse may not be true, give a simple counterexample in $\mathbb{R}$.

17.11. Higher-Order Partial Derivatives. Suppose $f:\Omega\,(\subset\mathbb{R}^n)\to\mathbb{R}$. The partial derivatives $\dfrac{\partial f}{\partial x^1},\dots,\dfrac{\partial f}{\partial x^n}$, if they exist, are also functions from Ω to $\mathbb{R}$, and may themselves have partial derivatives.

The jth partial derivative of $\dfrac{\partial f}{\partial x_i}$ is denoted by

$$
\frac{\partial^2 f}{\partial x_j\partial x_i}\ \text{or}\ f_{ij}\ \text{or}\ D_{ij}f.
$$

If all first and second partial derivatives of f exist and are *continuous* in Ω [54] we write

$$
f\in C^2(\Omega).
$$

Similar remarks apply to higher order derivatives, and we similarly define $C^q(\Omega)$ for any integer $q\ge 0$.

[54] In fact, it is sufficient to assume just that the *second* partial derivatives are continuous. For under this assumption, each $\partial f/\partial x^i$ must be differentiable by Theorem 17.10.1 applied to $\partial f/\partial x^i$. From Proposition 17.8.1 applied to $\partial f/\partial x^i$ it then follows that $\partial f/\partial x^i$ is continuous.

Note that

$$C^0(\Omega) \supset C^1(\Omega) \supset C^2(\Omega) \supset \dots$$

The usual rules for differentiating a sum, product or quotient of functions of a single variable apply to partial derivatives. It follows that $C^k(\Omega)$ is closed under addition, products and quotients (if the denominator is non-zero).

The next theorem shows that for higher order derivatives, the actual order of differentiation does not matter, only the number of derivatives with respect to each variable is important. Thus

$$\frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i},$$

and so

$$\frac{\partial^3 f}{\partial x^i \partial x^j \partial x^k} = \frac{\partial^3 f}{\partial x^j \partial x^i \partial x^k} = \frac{\partial^3 f}{\partial x^j \partial x^k \partial x^i}, \text{ etc.}$$

THEOREM 17.11.1. *If $f \in C^1(\Omega)$[55] and both f_{ij} and f_{ji} exist and are continuous (for some $i \neq j$) in Ω, then $f_{ij} = f_{ji}$ in Ω.*

In particular, if $f \in C^2(\Omega)$ then $f_{ij} = f_{ji}$ for all $i \neq j$.

PROOF. For notational simplicity we take $n = 2$. The proof for $n > 2$ is very similar.

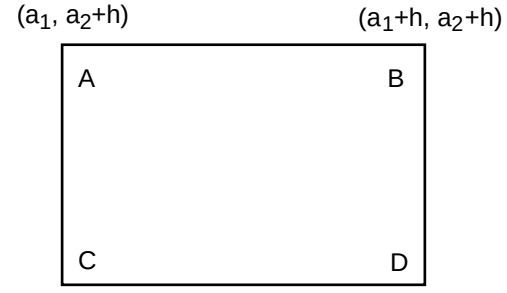
Suppose $\mathbf{a} \in \Omega$ and suppose $h > 0$ is some sufficiently small real number.

Consider the *second difference quotient* defined by

$$\begin{aligned} A(h) &= \frac{1}{h^2}\Big(\Big(f(a^1+h, a^2+h) - f(a^1, a^2+h)\Big) \\ &\qquad -\Big(f(a^1+h, a^2) - f(a^1, a^2)\Big)\Big) \qquad (53) \\ &= \frac{1}{h^2}\Big(g(a^2+h) - g(a^2)\Big), \qquad (54) \end{aligned}$$

where

$$g(x^2) = f(a^1+h, x^2) - f(a^1, x^2).$$



$\mathbf{a} = (a_1, a_2)$ $\qquad (a_1+h, a_2)$

$A(h) = \big(\big(f(B) - f(A) \big) - \big(f(D) - f(C) \big) \big) / h^2$

$= \big(\big(f(B) - f(D) \big) - \big(f(A) - f(C) \big) \big) / h^2$

From the definition of partial differentiation, $g'(x^2)$ exists and

$$g'(x^2) = \frac{\partial f}{\partial x^2}(a^1+h, x^2) - \frac{\partial f}{\partial x^2}(a^1, x^2) \qquad (55)$$

for $a^2 \leq x \leq a^2 + h$.

[55] As usual, Ω is assumed to be open.

Applying the mean value theorem for a function of a single variable to (54), we see from (55) that

$$\begin{aligned} A(h) &= \frac{1}{h}g'(\xi^2) \quad \text{some } \xi^2 \in (a^2, a^2+h) \\ &= \frac{1}{h}\left(\frac{\partial f}{\partial x^2}(a^1+h,\xi^2) - \frac{\partial f}{\partial x^2}(a^1,\xi^2)\right). \end{aligned} \tag{56}$$

Applying the mean value theorem again to the function $\dfrac{\partial f}{\partial x^2}(x^1,\xi^2)$, with ξ^2 fixed, we see

$$A(h) = \frac{\partial^2 f}{\partial x^1 \partial x^2}(\xi^1,\xi^2) \quad \text{some } \xi^1 \in (a^1, a^1+h). \tag{57}$$

If we now rewrite (53) as

$$\begin{aligned} A(h) &= \frac{1}{h^2}\Big(\Big(f(a^1+h,a^2+h) - f(a^1+h,a^2)\Big) \\ &\qquad -\Big(f(a^1,a^2+h) - f(a^1+a^2)\Big)\Big) \end{aligned} \tag{58}$$

and interchange the roles of x^1 and x^2 in the previous argument, we obtain

$$A(h) = \frac{\partial^2 f}{\partial x^2 \partial x^1}(\eta^1,\eta^2) \tag{59}$$

for some $\eta^1 \in (a^1, a^1+h)$, $\eta^2 \in (a^2, a^2+h)$.

If we let $h \to 0$ then (ξ^1,ξ^2) and $(\eta^1,\eta^2) \to (a^1,a^2)$, and so from (57), (59) and the continuity of f_{12} and f_{21} at $\mathbf{a}$, it follows that

$$f_{12}(\mathbf{a}) = f_{21}(\mathbf{a}).$$

This completes the proof. □

17.12. Taylor's Theorem. If $g \in C^1[a,b]$, then we know

$$g(b) = g(a) + \int_a^b g'(t)\,dt$$

This is the case $k = 1$ of the following version of Taylor's Theorem for a function of *one* variable.

THEOREM 17.12.1 (Single Variable, Integral form of the Remainder). *Suppose $g \in C^k[a,b]$. Then*

$$\begin{aligned} g(b) &= g(a) + g'(a)(b-a) + \frac{1}{2!}g''(a)(b-a)^2 + \cdots \\ &\quad + \frac{1}{(k-1)!}g^{(k-1)}(a)(b-a)^{k-1} + \int_a^b \frac{(b-t)^{k-1}}{(k-1)!}g^{(k)}(t)\,dt. \end{aligned} \tag{60}$$

PROOF. An elegant (but not obvious) proof is to begin by computing:

$$\begin{aligned} &\frac{d}{dt}\left(g\varphi^{(k-1)} - g'\varphi^{(k-2)} + g''\varphi^{(k-3)} - \cdots + (-1)^{k-1}g^{(k-1)}\varphi\right) \\ &\quad = \left(g\varphi^{(k)} + g'\varphi^{(k-1)}\right) - \left(g'\varphi^{(k-1)} + g''\varphi^{(k-2)}\right) + \left(g''\varphi^{(k-2)} + g'''\varphi^{(k-3)}\right) - \\ &\qquad \cdots + (-1)^{k-1}\left(g^{(k-1)}\varphi' + g^{(k)}\varphi\right) \\ &= g\varphi^{(k)} + (-1)^{k-1}g^{(k)}\varphi. \end{aligned} \tag{61}$$

Now choose

$$\varphi(t) = \frac{(b-t)^{k-1}}{(k-1)!}.$$

Then

$$\begin{aligned}
\varphi'(t) &= (-1)\frac{(b-t)^{k-2}}{(k-2)!} \\
\varphi''(t) &= (-1)^2\frac{(b-t)^{k-3}}{(k-3)!} \\
&\vdots \\
\varphi^{(k-3)}(t) &= (-1)^{k-3}\frac{(b-t)^2}{2!} \\
\varphi^{(k-2)}(t) &= (-1)^{k-2}(b-t) \\
\varphi^{(k-1)}(t) &= (-1)^{k-1} \\
\varphi^{k}(t) &= 0. \qquad (62)
\end{aligned}$$

Hence from (61) we have

$$\begin{aligned}
&(-1)^{k-1}\frac{d}{dt}\left(g(t)+g'(t)(b-t)+g''(t)\frac{(b-t)^2}{2!}+\cdots+g^{k-1}(t)\frac{(b-t)^{k-1}}{(k-1)!}\right) \\
&\quad = (-1)^{k-1}g^{(k)}(t)\frac{(b-t)^{k-1}}{(k-1)!}.
\end{aligned}$$

Dividing by $(-1)^{k-1}$ and integrating both sides from a to b, we get

$$\begin{aligned}
&g(b)-\left(g(a)+g'(a)(b-a)+g''(a)\frac{(b-a)^2}{2!}+\cdots+g^{(k-1)}(a)\frac{(b-a)^{k-1}}{(k-1)!}\right) \\
&\quad = \int_a^b g^{(k)}(t)\frac{(b-t)^{k-1}}{(k-1)!}\,dt.
\end{aligned}$$

This gives formula (60). □

THEOREM 17.12.2 (Single Variable, Second Version). *Suppose $g \in C^k[a,b]$. Then*

$$\begin{aligned}
g(b) &= g(a)+g'(a)(b-a)+\frac{1}{2!}g''(a)(b-a)^2+\cdots \\
&\quad +\frac{1}{(k-1)!}g^{(k-1)}(a)(b-a)^{k-1}+\frac{1}{k!}g^{(k)}(\xi)(b-a)^k \qquad (63)
\end{aligned}$$

for some $\xi \in (a,b)$.

PROOF. We establish (63) from (60).

Since $g^{(k)}$ is continuous in $[a,b]$, it has a minimum value m, and a maximum value M, say.

By elementary properties of integrals, it follows that

$$\int_a^b m\frac{(b-t)^{k-1}}{(k-1)!}\,dt \le \int_a^b g^{(k)}(t)\frac{(b-t)^{k-1}}{(k-1)!}\,dt \le \int_a^b M\frac{(b-t)^{k-1}}{(k-1)!}\,dt,$$

i.e.

$$m \le \frac{\displaystyle\int_a^b g^{(k)}(t)\frac{(b-t)^{k-1}}{(k-1)!}dt}{\displaystyle\int_a^b \frac{(b-t)^{k-1}}{(k-1)!}dt} \le M.$$

By the Intermediate Value Theorem, $g^{(k)}$ takes all values in the range $[m,M]$, and so the middle term in the previous inequality must equal $g^{(k)}(\xi)$ for some $\xi \in (a,b)$. Since

$$\int_a^b \frac{(b-t)^{k-1}}{(k-1)!}dt = \frac{(b-a)^k}{k!},$$

it follows

$$\int_a^b g^{(k)}(t)\frac{(b-t)^{k-1}}{(k-1)!}dt = \frac{(b-a)^k}{k!}g^{(k)}(\xi).$$

Formula (63) now follows from (60). □

Taylor's Theorem generalises easily to functions of more than one variable.

THEOREM 17.12.3 (Taylor's Formula; Several Variables).
Suppose $f \in C^k(\Omega)$ where $\Omega \subset \mathbb{R}^n$, and the line segment joining $\mathbf{a}$ and $\mathbf{a}+\mathbf{h}$ is a subset of Ω.

Then

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= f(\mathbf{a}) + \sum_{i-1}^{n} D_i f(\mathbf{a})\, h^i + \frac{1}{2!}\sum_{i,j=1}^{n} D_{ij} f(\mathbf{a})\, h^i h^j + \cdots \\ &\quad + \frac{1}{(k-1)!}\sum_{i_1,\cdots,i_{k-1}=1}^{n} D_{i_1\ldots i_{k-1}} f(\mathbf{a})\, h^{i_1}\cdot\ldots\cdot h^{i_{k-1}} + R_k(\mathbf{a},\mathbf{h}) \end{aligned}$$

where

$$\begin{aligned} R_k(\mathbf{a},\mathbf{h}) &= \frac{1}{(k-1)!}\sum_{i_1,\ldots,i_k=1}^{n}\int_0^1 (1-t)^{k-1} D_{i_1\ldots i_k} f(\mathbf{a}+t\mathbf{h})\, dt \\ &= \frac{1}{k!}\sum_{i_1,\ldots,i_k=1}^{n} D_{i_1,\ldots,i_k} f(\mathbf{a}+s\mathbf{h})\, h^{i_1}\cdot\ldots\cdot h^{i_k} \quad \textit{for some } s\in(0,1). \end{aligned}$$

PROOF. First note that for any differentiable function $F : D\,(\subset \mathbb{R}^n) \to \mathbb{R}$ we have

$$\frac{d}{dt}F(\mathbf{a}+t\mathbf{h}) = \sum_{i=1}^{n} D_i F(\mathbf{a}+t\mathbf{h})\, h^i. \tag{64}$$

This is just a particular case of the *chain rule*, which we will discuss later. This particular version follows from (51) and Corollary 17.5.3 (with f there replaced by F).

Let

$$g(t) = f(a+th).$$

Then $g : [0,1] \to \mathbb{R}$. We will apply Taylor's Theorem for a function of one variable to g.

From (64) we have

$$g'(t) = \sum_{i-1}^{n} D_i f(\mathbf{a}+t\mathbf{h})\, h^i. \tag{65}$$

Differentiating again, and applying (64) to $D_i F$, we obtain

$$\begin{aligned} g''(t) &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} D_{ij} f(\mathbf{a}+t\mathbf{h})\, h^j\right) h^i \\ &= \sum_{i,j=1}^{n} D_{ij} f(\mathbf{a}+t\mathbf{h})\, h^i h^j. \end{aligned} \tag{66}$$

Similarly

$$g'''(t) = \sum_{i,j,k=1}^{n} D_{ijk} f(\mathbf{a}+t\mathbf{h})\, h^i h^j h^k, \tag{67}$$

etc. In this way, we see $g \in C^k[0,1]$ and obtain formulae for the derivatives of g.

But from (60) and (63) we have

$$\begin{aligned} g(1) &= g(0) + g'(0) + \frac{1}{2!}g''(0) + \cdots + \frac{1}{(k-1)!}g^{(k-1)}(0) \\ &+ \begin{cases} \dfrac{1}{(k-1)!}\displaystyle\int_0^1 (1-t)^{k-1} g^{(k)}(t)\,dt \\ \qquad\qquad or \\ \dfrac{1}{k!}g^{(k)}(s) \quad \text{some } s \in (0,1). \end{cases} \end{aligned}$$

If we substitute (65), (66), (67) etc. into this, we obtain the required results. □

REMARK 17.12.1. The first two terms of Taylor's Formula give the best *first order approximation*[56] in $\mathbf{h}$ to $f(\mathbf{a}+\mathbf{h})$ for $\mathbf{h}$ near $\mathbf{0}$. The first three terms give the best *second order approximation*[57] in $\mathbf{h}$, the first four terms give the best *third order approximation*, etc.

Note that the *remainder term* $R_k(\mathbf{a},\mathbf{h})$ in Theorem 17.12.3 can be written as $O(|\mathbf{h}|^k)$ (see the Remarks on rates of convergence in Section 17.5), i.e.

$$\frac{R_k(\mathbf{a},\mathbf{h})}{|\mathbf{h}|^k} \text{ is bounded as } \mathbf{h} \to 0.$$

This follows from the second version for the remainder in Theorem 17.12.3 and the facts:

1. $D_{i_1\ldots i_k} f(\mathbf{x})$ is continuous, and hence bounded on compact sets,
2. $|h^{i_1} \cdot \ldots \cdot h^{i_k}| \le |\mathbf{h}|^k$.

EXAMPLE 17.12.4. Let

$$f(x,y) = (1+y^2)^{1/2}\cos x.$$

One finds the best second order approximation to f for (x,y) near $(0,1)$ as follows.

First note that

$$f(0,1) = 2^{1/2}.$$

Moreover,

$$\begin{array}{lllll} f_1 &=& -(1+y^2)^{1/2}\sin x; &=& 0 \quad \text{at } (0,1) \\ f_2 &=& y(1+y^2)^{-1/2}\cos x; &=& 2^{-1/2} \quad \text{at } (0,1) \\ f_{11} &=& -(1+y^2)^{1/2}\cos x; &=& -2^{1/2} \quad \text{at } (0,1) \\ f_{12} &=& -y(1+y^2)^{-1/2}\sin x; &=& 0 \quad \text{at } (0,1) \\ f_{22} &=& (1+y^2)^{-3/2}\cos x; &=& 2^{-3/2} \quad \text{at } (0,1). \end{array}$$

Hence

$$f(x,y) = 2^{1/2} + 2^{-1/2}(y-1) - 2^{1/2}x^2 + 2^{-3/2}(y-1)^2 + R_3\Big((0,1),(x,y)\Big),$$

where

$$R_3\Big((0,1),(x,y)\Big) = O\left(|(x,y)-(0,1)|^3\right) = O\left(\left(x^2+(y-1)^2\right)^{3/2}\right).$$

[56] I.e. constant plus linear term.

[57] I.e. constant plus linear term plus quadratic term.

18. Differentiation of Vector-Valued Functions

You need only consider Subsections 1, 2 (to the beginning of Definition 19.2.7), 3, 4, 5 (the last is not examinable). What is required is just a basic understanding of the ideas, and application to straightforward examples.

18.1. Introduction. In this chapter we consider functions

$$\mathbf{f}: D\,(\subset \mathbb{R}^n) \to \mathbb{R}^m,$$

with $m \geq 1$.

We write

$$\mathbf{f}(x^1, \ldots, x^n) = \Big(f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)\Big)$$

where

$$f^i: D \to \mathbb{R} \quad i = 1, \ldots, m$$

are *real*-valued functions.

Example 18.1.1. Let

$$\mathbf{f}(x, y, z) = (x^2 - y^2, 2xz + 1).$$

Then $f^1(x, y, z) = x^2 - y^2$ and $f^2(x, y, z) = 2xz + 1$.

Reduction to Component Functions For many purposes we can reduce the study of functions $\mathbf{f}$, as above, to the study of the corresponding *real*-valued functions $f^1, \ldots, f^m$. However, this is not always a good idea, since studying the f^i involves a choice of coordinates in $\mathbb{R}^m$, and this can obscure the geometry involved.

In Definitions 18.2.1, 18.3.1 and 18.4.1 we define the notion of partial derivative, directional derivative, and differential of $\mathbf{f}$ without reference to the component functions. In Propositions 18.2.2, 18.3.2 and 18.4.2 we show these definitions are equivalent to definitions in terms of the component functions.

18.2. Paths in $\mathbb{R}^m$. In this section we consider the case corresponding to $n = 1$ in the notation of the previous section. This is an important case in its own right and also helps motivate the case $n > 1$.

Definition 18.2.1. Let I be an interval in $\mathbb{R}$. If $\mathbf{f}: I \to \mathbb{R}^n$ then the *derivative* or *tangent vector* at t is the vector

$$\mathbf{f}'(t) = \lim_{s \to 0} \frac{\mathbf{f}(t+s) - \mathbf{f}(t)}{s},$$

provided the limit exists[58]. In this case we say $\mathbf{f}$ is *differentiable* at t. If, moreover, $\mathbf{f}'(t) \neq 0$ then $\mathbf{f}'(t)/|\mathbf{f}'(t)|$ is called the *unit tangent* at t.

Remark 18.2.1. Although we say $\mathbf{f}'(t)$ is the tangent vector at t, we should really think of $\mathbf{f}'(t)$ as a vector with its "base" at $\mathbf{f}(t)$. See the next diagram.

Proposition 18.2.2. *Let* $\mathbf{f}(t) = \Big(f^1(t), \ldots, f^m(t)\Big)$. *Then* $\mathbf{f}$ *is differentiable at* t *iff* $f^1, \ldots, f^m$ *are differentiable at* t. *In this case*

$$\mathbf{f}'(t) = \Big(f^{1\prime}(t), \ldots, f^{m\prime}(t)\Big).$$

Proof. Since

$$\frac{\mathbf{f}(t+s) - \mathbf{f}(t)}{s} = \left(\frac{f^1(t+s) - f^1(t)}{s}, \ldots, \frac{f^m(t+s) - f^m(t)}{s}\right),$$

The theorem follows since a function into $\mathbb{R}^m$ converges iff its component functions converge. □

[58] If t is an endpoint of I then one takes the corresponding one-sided limits.

DEFINITION 18.2.3. If $\mathbf{f}(t) = \Big(f^1(t), \dots, f^m(t)\Big)$ then $\mathbf{f}$ is C^1 if each f^i is C^1.

We have the usual rules for differentiating the sum of two functions from I to $\mathbb{R}^m$, and the product of such a function with a real valued function (*exercise*: formulate and prove such a result). The following rule for differentiating the inner product of two functions is useful.

PROPOSITION 18.2.4. *If $\mathbf{f}_1, \mathbf{f}_2 : I \to \mathbb{R}^m$ are differentable at t then*

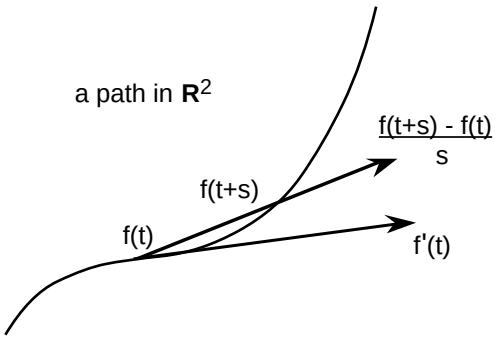
$$\frac{d}{dt}\Big(\mathbf{f}_1(t), \mathbf{f}_2(t)\Big) = \Big(\mathbf{f}_1'(t), \mathbf{f}_2(t)\Big) + \Big(\mathbf{f}_1(t), \mathbf{f}_2'(t)\Big).$$

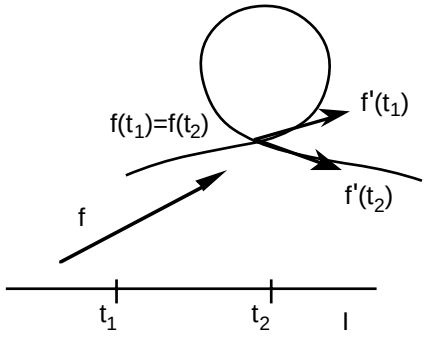
PROOF. Since

$$\Big(\mathbf{f}_1(t), \mathbf{f}_2(t)\Big) = \sum_{i=1}^{m} f_1^i(t) f_2^i(t),$$

the result follows from the usual rule for differentiation sums and products. □

If $\mathbf{f} : I \to \mathbb{R}^m$, we can think of $\mathbf{f}$ as tracing out a "curve" in $\mathbb{R}^m$ (we will make this precise later). The terminology *tangent vector* is reasonable, as we see from the following diagram. Sometimes we speak of the tangent vector *at* $\mathbf{f}(t)$ rather than *at* t, but we need to be careful if $\mathbf{f}$ is not one-one, as in the second figure.





EXAMPLE 18.2.5.

1. Let

$$\mathbf{f}(t) = (\cos t, \sin t) \quad t \in [0, 2\pi).$$

This traces out a circle in $\mathbb{R}^2$ and

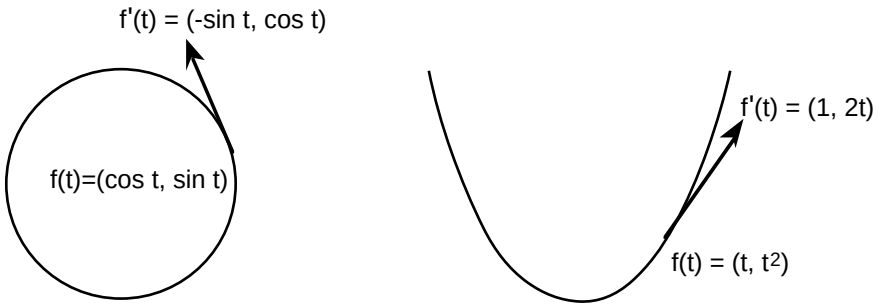
$$\mathbf{f}'(t) = (-\sin t, \cos t).$$

2. Let

$$\mathbf{f}(t) = (t, t^2).$$

This traces out a parabola in $\mathbb{R}^2$ and

$$\mathbf{f}'(t) = (1, 2t).$$



EXAMPLE 18.2.6. Consider the functions

1. $\mathbf{f}_1(t) = (t, t^3) \quad t \in \mathbb{R}$,
2. $\mathbf{f}_2(t) = (t^3, t^9) \quad t \in \mathbb{R}$,
3. $\mathbf{f}_3(t) = (\sqrt[3]{t}, t) \quad t \in \mathbb{R}$.

Then each function $\mathbf{f}_i$ traces out the same "cubic" curve in $\mathbb{R}^2$, (i.e., the image is the same set of points), and

$$\mathbf{f}_1(0) = \mathbf{f}_2(0) = \mathbf{f}_3(0) = (0, 0).$$

However,

$$\mathbf{f}_1'(0) = (1, 0),\ \mathbf{f}_2'(0) = (0, 0),\ \mathbf{f}_3'(0) \text{ is undefined.}$$

Intuitively, we will think of a *path* in $\mathbb{R}^m$ as a function $\mathbf{f}$ which neither stops nor reverses direction. It is often convenient to consider the variable t as representing "time". We will think of the corresponding *curve* as the set of points traced out by $\mathbf{f}$. Many different paths (i.e. functions) will give the same curve; they correspond to tracing out the curve at different times and velocities. We make this precise as follows:

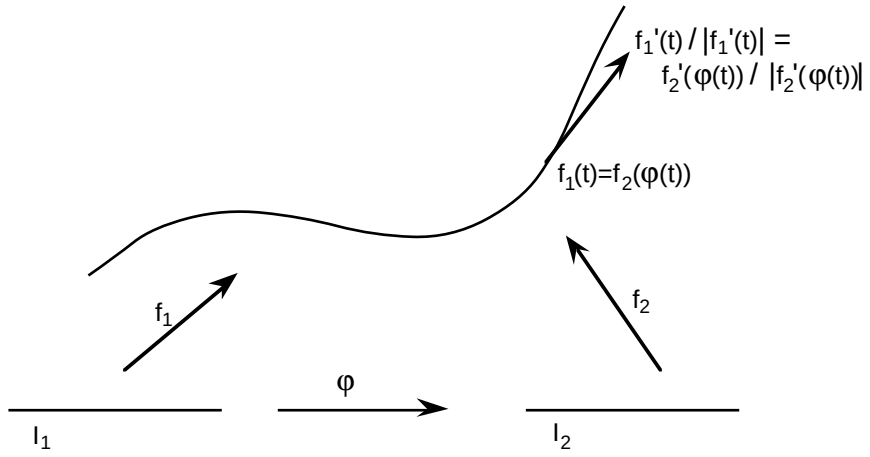
DEFINITION 18.2.7. We say $\mathbf{f}: I \to \mathbb{R}^m$ is a *path*[59] in $\mathbb{R}^m$ if $\mathbf{f}$ is C^1 and $\mathbf{f}'(t) \neq 0$ for $t \in I$. We say the two paths $\mathbf{f}_1 : I_1 \to \mathbb{R}^m$ and $\mathbf{f}_2 : I_2 \to \mathbb{R}^m$ are *equivalent* if there exists a function $\phi: I_1 \to I_2$ such that $\mathbf{f}_1 = \mathbf{f}_2 \circ \phi$, where ϕ is C^1 and $\phi'(t) > 0$ for $t \in I_1$.

A *curve* is an equivalence class of paths. Any path in the equivalence class is called a *parametrisation* of the curve.

We can think of ϕ as giving another way of measuring "time".

We expect that the *unit* tangent vector to a curve should depend only on the curve itself, and not on the particular parametrisation. This is indeed the case, as is shown by the following Proposition.

[59] Other texts may have different terminology.



PROPOSITION 18.2.8. *Suppose* $\mathbf{f}_1 : I_1 \to \mathbb{R}^m$ *and* $\mathbf{f}_2 : I_2 \to \mathbb{R}^m$ *are equivalent parametrisations; and in particular* $\mathbf{f}_1 = \mathbf{f}_2 \circ \phi$ *where* $\phi : I_1 \to I_2$, ϕ *is* C^1 *and* $\phi'(t) > 0$ *for* $t \in I_1$. *Then* $\mathbf{f}_1$ *and* $\mathbf{f}_2$ *have the same unit tangent vector at* t *and* $\phi(t)$ *respectively.*

PROOF. From the chain rule for a function of one variable, we have

$$\begin{aligned}\mathbf{f}_1'(t) &= \Big(f_1^{1\prime}(t), \ldots, f_1^{m\prime}(t)\Big) \\ &= \Big(f_2^{1\prime}(\phi(t))\,\phi'(t), \ldots, f_2^{m\prime}(\phi(t))\,\phi'(t)\Big) \\ &= \mathbf{f}_2'(\phi(t))\,\phi'(t).\end{aligned}$$

Hence, since $\phi'(t) > 0$,

$$\frac{\mathbf{f}_1'(t)}{\left|\mathbf{f}_1'(t)\right|} = \frac{\mathbf{f}_2'(t)}{\left|\mathbf{f}_2'(t)\right|}.$$

□

DEFINITION 18.2.9. If $\mathbf{f}$ is a path in $\mathbb{R}^m$, then the *acceleration* at t is $\mathbf{f}''(t)$.

EXAMPLE 18.2.10. If $|\mathbf{f}'(t)|$ is constant (i.e. the "speed" is constant) then the velocity and the acceleration are orthogonal.

PROOF. Since $|\mathbf{f}(t)|^2 = \Big(\mathbf{f}'(t), \mathbf{f}'(y)\Big)$ is constant, we have from Proposition 18.2.4 that

$$0 = \frac{d}{dt}\Big(\mathbf{f}'(t), \mathbf{f}'(y)\Big) = 2\Big(\mathbf{f}''(t), \mathbf{f}'(y)\Big).$$

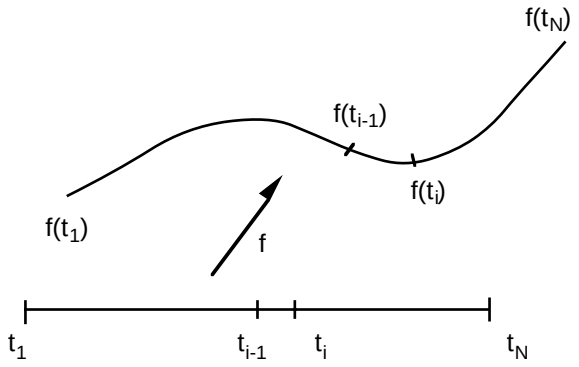
This gives the result. □

Arc length Suppose $\mathbf{f} : [a, b] \to \mathbb{R}^m$ is a path in $\mathbb{R}^m$. Let $a = t_1 < t_2 < \ldots < t_n = b$ be a partition of $[a, b]$, where $t_i - t_{i-1} = \delta t$ for all i.

We think of the length of the curve corresponding to $\mathbf{f}$ as being

$$\approx \sum_{i=2}^{N} |f(t_i) - f(t_{i-1})| = \sum_{i=2}^{N} \frac{|f(t_i) - f(t_{i-1})|}{\delta t}\delta t \approx \int_a^b \left|\mathbf{f}'(t)\right| dt.$$

See the next diagram.



Motivated by this we make the following definition.

DEFINITION 18.2.11. Let $\mathbf{f}:[a,b]\to\mathbb{R}^m$ be a path in $\mathbb{R}^m$. Then the *length* of the curve corresponding to $\mathbf{f}$ is given by

$$\int_a^b \left|\mathbf{f}'(t)\right| dt.$$

The next result shows that this definition is independent of the particular parametrisation chosen for the curve.

PROPOSITION 18.2.12. *Suppose* $\mathbf{f}_1:[a_1,b_1]\to\mathbb{R}^m$ *and* $\mathbf{f}_2:[a_2,b_2]\to\mathbb{R}^m$ *are equivalent parametrisations; and in particular* $\mathbf{f}_1=\mathbf{f}_2\circ\phi$ *where* $\phi:[a_1,b_1]\to[a_2,b_2]$, ϕ *is* C^1 *and* $\phi'(t)>0$ *for* $t\in I_1$. *Then*

$$\int_{a_1}^{b_1}\left|\mathbf{f}_1'(t)\right| dt=\int_{a_2}^{b_2}\left|\mathbf{f}_2'(s)\right| ds.$$

PROOF. From the chain rule and then the rule for change of variable of integration,

$$\begin{aligned}\int_{a_1}^{b_1}\left|\mathbf{f}_1'(t)\right| dt &= \int_{a_1}^{b_1}\left|\mathbf{f}_2'(\phi(t))\right|\phi'(t)dt\\ &= \int_{a_2}^{b_2}\left|\mathbf{f}_2'(s)\right| ds.\end{aligned}$$

□

18.3. Partial and Directional Derivatives. Analogous to Definitions 17.3.1 and 17.4.1 we have:

DEFINITION 18.3.1. The ith *partial derivative* of $\mathbf{f}$ at $\mathbf{x}$ is defined by

$$\frac{\partial\mathbf{f}}{\partial x^i}(\mathbf{x})\ \Big(\text{or } D_i\mathbf{f}(\mathbf{x})\Big)=\lim_{t\to0}\frac{\mathbf{f}(\mathbf{x}+t\mathbf{e}_i)-\mathbf{f}(\mathbf{x})}{t},$$

provided the limit exists. More generally, the *directional derivative* of $\mathbf{f}$ at $\mathbf{x}$ in the direction $\mathbf{v}$ is defined by

$$D_{\mathbf{v}}\mathbf{f}(\mathbf{x})=\lim_{t\to0}\frac{\mathbf{f}(\mathbf{x}+t\mathbf{v})-\mathbf{f}(\mathbf{x})}{t},$$

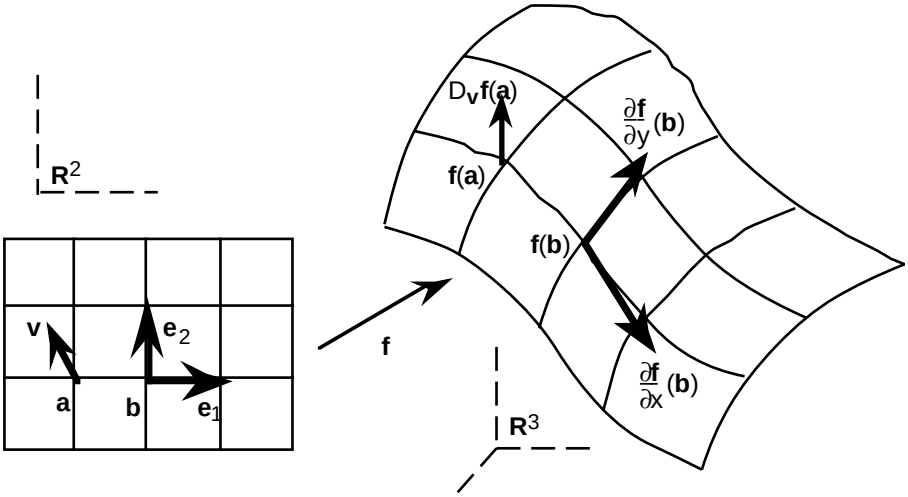
provided the limit exists.

REMARK 18.3.1.

1. It follows immediately from the Definitions that

$$\frac{\partial\mathbf{f}}{\partial x^i}(\mathbf{x})=D_{\mathbf{e}_i}\mathbf{f}(\mathbf{x}).$$

2. The partial and directional derivatives are *vectors in* $\mathbb{R}^m$. In the terminology of the previous section, $\dfrac{\partial \mathbf{f}}{\partial x^i}(\mathbf{x})$ is tangent to the path $t \mapsto \mathbf{f}(\mathbf{x} + t\mathbf{e}_i)$ and $D_{\mathbf{v}}\mathbf{f}(\mathbf{x})$ is tangent to the path $t \mapsto \mathbf{f}(\mathbf{x} + t\mathbf{v})$. Note that the curves corresponding to these paths are subsets of the image of $\mathbf{f}$.
3. As we will discuss later, we may regard the partial derivatives at $\mathbf{x}$ as a basis for the tangent space to the image of $\mathbf{f}$ at $\mathbf{f}(\mathbf{x})$[60].



PROPOSITION 18.3.2. *If $f^1, \ldots, f^m$ are the component functions of $\mathbf{f}$ then*

$$\begin{aligned} \frac{\partial \mathbf{f}}{\partial x^i}(\mathbf{a}) &= \left(\frac{\partial f^1}{\partial x^i}(\mathbf{a}), \ldots, \frac{\partial f^m}{\partial x^i}(\mathbf{a})\right) \quad \text{for } i = 1, \ldots, n \\ D_{\mathbf{v}}\mathbf{f}(\mathbf{a}) &= \Big(D_{\mathbf{v}} f^1(\mathbf{a}), \ldots, D_{\mathbf{v}} f^m(\mathbf{a})\Big) \end{aligned}$$

in the sense that if one side of either equality exists, then so does the other, and both sides are then equal.

PROOF. Essentially the same as for the proof of Proposition 18.2.2. □

EXAMPLE 18.3.3. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$\mathbf{f}(x, y) = (x^2 - 2xy,\ x^2 + y^3,\ \sin x).$$

Then

$$\begin{aligned} \frac{\partial \mathbf{f}}{\partial x}(x, y) &= \left(\frac{\partial f^1}{\partial x}, \frac{\partial f^2}{\partial x}, \frac{\partial f^3}{\partial x}\right) = (2x - 2y,\ 2x,\ \cos x), \\ \frac{\partial \mathbf{f}}{\partial y}(x, y) &= \left(\frac{\partial f^1}{\partial y}, \frac{\partial f^2}{\partial y}, \frac{\partial f^3}{\partial y}\right) = (-2x,\ 3y^2,\ 0), \end{aligned}$$

are vectors in $\mathbb{R}^3$.

18.4. The Differential. Analogous to Definition 17.5.1 we have:

DEFINITION 18.4.1. Suppose $\mathbf{f} : D\,(\subset \mathbb{R}^n) \to \mathbb{R}^m$. Then $\mathbf{f}$ is *differentiable* at $\mathbf{a} \in D$ if there is a linear transformation $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\frac{\Big|\mathbf{f}(\mathbf{x}) - \Big(\mathbf{f}(\mathbf{a}) + \mathbf{L}(\mathbf{x} - \mathbf{a})\Big)\Big|}{|\mathbf{x} - \mathbf{a}|} \to 0 \quad \text{as } \mathbf{x} \to \mathbf{a}. \tag{68}$$

[60]More precisely, if $n \leq m$ and the differential $d\mathbf{f}(\mathbf{x})$ has rank n. See later.

The linear transformation $\mathbf{L}$ is denoted by $\mathbf{f}'(\mathbf{a})$ or $d\mathbf{f}(\mathbf{a})$ and is called the *derivative* or *differential* of $\mathbf{f}$ at $\mathbf{a}$[61].

A vector-valued function is differentiable iff the corresponding component functions are differentiable. More precisely:

PROPOSITION 18.4.2. *$\mathbf{f}$ is differentiable at $\mathbf{a}$ iff $f^1, \dots, f^m$ are differentiable at $\mathbf{a}$. In this case the differential is given by*

$$\langle d\mathbf{f}(\mathbf{a}), \mathbf{v}\rangle = \Big(\langle df^1(\mathbf{a}), \mathbf{v}\rangle, \dots, \langle df^m(\mathbf{a}), \mathbf{v}\rangle\Big). \tag{69}$$

In particular, the differential is unique.

PROOF. For any linear map $\mathbf{L}: \mathbb{R}^n \to \mathbb{R}^m$, and for each $i = 1, \dots, m$, let $L^i: \mathbb{R}^n \to \mathbb{R}$ be the linear map defined by $L^i(\mathbf{v}) = \Big(\mathbf{L}(\mathbf{v})\Big)^i$.

Since a function into $\mathbb{R}^m$ converges iff its component functions converge, it follows

$$\frac{\Big|\mathbf{f}(\mathbf{x}) - \Big(\mathbf{f}(\mathbf{a}) + \mathbf{L}(\mathbf{x}-\mathbf{a})\Big)\Big|}{|\mathbf{x}-\mathbf{a}|} \to 0 \quad \text{as } \mathbf{x} \to \mathbf{a}$$

iff

$$\frac{\Big|f^i(\mathbf{x}) - \Big(f^i(\mathbf{a}) + L^i(\mathbf{x}-\mathbf{a})\Big)\Big|}{|\mathbf{x}-\mathbf{a}|} \to 0 \quad \text{as } \mathbf{x} \to \mathbf{a} \quad \text{for } i = 1, \dots, m.$$

Thus $\mathbf{f}$ is differentiable at $\mathbf{a}$ iff $f^1, \dots, f^m$ are differentiable at $\mathbf{a}$.

In this case we must have

$$L^i = df^i(\mathbf{a}) \quad i = 1, \dots, m$$

(by uniqueness of the differential for *real*-valued functions), and so

$$\mathbf{L}(\mathbf{v}) = \Big(\langle df^1(\mathbf{a}), \mathbf{v}\rangle, \dots, \langle df^m(\mathbf{a}), \mathbf{v}\rangle\Big).$$

But this says that the differential $d\mathbf{f}(\mathbf{a})$ is unique and is given by (69). □

COROLLARY 18.4.3. *If $\mathbf{f}$ is differentiable at $\mathbf{a}$ then the linear transformation $d\mathbf{f}(\mathbf{a})$ is represented by the matrix*

$$\begin{bmatrix} \dfrac{\partial f^1}{\partial x^1}(\mathbf{a}) & \cdots & \dfrac{\partial f^1}{\partial x^n}(\mathbf{a}) \\ \vdots & \cdots & \vdots \\ \dfrac{\partial f^m}{\partial x^1}(\mathbf{a}) & \cdots & \dfrac{\partial f^m}{\partial x^n}(\mathbf{a}) \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^m \tag{70}$$

PROOF. The ith column of the matrix corresponding to $d\mathbf{f}(\mathbf{a})$ is the vector $\langle d\mathbf{f}(\mathbf{a}), \mathbf{e}_i\rangle$[62]. From Proposition 18.4.2 this is the *column* vector corresponding to

$$\Big(\langle df^1(\mathbf{a}), \mathbf{e}_i\rangle, \dots, \langle df^m(\mathbf{a}), \mathbf{e}_i\rangle\Big),$$

i.e. to

$$\Big(\frac{\partial f^1}{\partial x^i}(\mathbf{a}), \dots, \frac{\partial f^m}{\partial x^i}(\mathbf{a})\Big).$$

This proves the result. □

[61] It follows from Proposition 18.4.2 that if $\mathbf{L}$ exists then it is unique and is given by the right side of (69).

[62] For any linear transformation $\mathbf{L}: \mathbb{R}^n \to \mathbb{R}^m$, the ith column of the corresponding matrix is $\mathbf{L}(\mathbf{e}_i)$.

REMARK 18.4.1. The jth *column* is the vector in $\mathbb{R}^m$ corresponding to the partial derivative $\dfrac{\partial \mathbf{f}}{\partial x^j}(\mathbf{a})$. The ith *row* represents $df^i(\mathbf{a})$.

The following proposition is immediate.

PROPOSITION 18.4.4. *If* $\mathbf{f}$ *is differentiable at* $\mathbf{a}$ *then*

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \langle d\mathbf{f}(\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle + \psi(\mathbf{x}),$$

where $\psi(\mathbf{x}) = o(|\mathbf{x}-\mathbf{a}|)$.

Conversely, suppose

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + L(\mathbf{x}-\mathbf{a}) + \psi(\mathbf{x}),$$

where $L:\mathbb{R}^n \to \mathbb{R}^m$ *is linear and* $\psi(\mathbf{x}) = o(|\mathbf{x}-\mathbf{a}|)$. *Then* $\mathbf{f}$ *is differentiable at* $\mathbf{a}$ *and* $d\mathbf{f}(\mathbf{a}) = L$.

PROOF. As for Proposition 17.5.7. □

Thus as is the case for real-valued functions, the previous proposition implies $\mathbf{f}(\mathbf{a}) + \langle d\mathbf{f}(\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle$ gives the best first order approximation to $\mathbf{f}(\mathbf{x})$ for $\mathbf{x}$ near $\mathbf{a}$.

EXAMPLE 18.4.5. Let $\mathbf{f}:\mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$\mathbf{f}(x,y) = (x^2 - 2xy, x^2 + y^3).$$

Find the best first order approximation to $\mathbf{f}(\mathbf{x})$ for $\mathbf{x}$ near $(1,2)$.
Solution:

$$\begin{aligned}
\mathbf{f}(1,2) &= \begin{bmatrix} -3 \\ 9 \end{bmatrix}, \\
d\mathbf{f}(x,y) &= \begin{bmatrix} 2x-2y & -2x \\ 2x & 3y^2 \end{bmatrix}, \\
d\mathbf{f}(1,2) &= \begin{bmatrix} -2 & -2 \\ 2 & 12 \end{bmatrix}.
\end{aligned}$$

So the best first order approximation near $(1,2)$ is

$$\begin{aligned}
&\mathbf{f}(1,2) + \langle d\mathbf{f}(1,2), (x-1, y-2)\rangle \\
&\quad = \begin{bmatrix} -3 \\ 9 \end{bmatrix} + \begin{bmatrix} -2 & -2 \\ 2 & 12 \end{bmatrix} \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} \\
&\quad = \begin{bmatrix} -3 - 2(x-1) - 4(y-2) \\ 9 + 2(x-1) + 12(y-2) \end{bmatrix} \\
&\quad = \begin{bmatrix} 7 - 2x - 4y \\ -17 + 2x + 12y \end{bmatrix}.
\end{aligned}$$

Alternatively, working with each component separately, the best first order approximation is

$$\begin{aligned}
&\Big(f^1(1,2) + \frac{\partial f^1}{\partial x}(1,2)(x-1) + \frac{\partial f^1}{\partial y}(1,2)(y-2), \\
&\qquad f^2(1,2) + \frac{\partial f^2}{\partial x}(1,2)(x-1) + \frac{\partial f^2}{\partial y}(y-2)\Big) \\
&= \Big(-3 - 2(x-1) - 4(y-2),\ 9 + 2(x-1) + 12(y-2)\Big) \\
&= \Big(7 - 2x - 4y,\ -17 + 2x + 12y\Big).
\end{aligned}$$

REMARK 18.4.2. One similarly obtains second and higher order approximations by using Taylor's formula for each component function.

PROPOSITION 18.4.6. *If* $\mathbf{f}, \mathbf{g} : D\,(\subset \mathbb{R}^n) \to \mathbb{R}^m$ *are differentiable at* $\mathbf{a} \in D$*, then so are* $\alpha\mathbf{f}$ *and* $\mathbf{f} + \mathbf{g}$*. Moreover,*

$$\begin{aligned} d(\alpha\mathbf{f})(\mathbf{a}) &= \alpha d\mathbf{f}(\mathbf{a}), \\ d(\mathbf{f}+\mathbf{g})(\mathbf{a}) &= df(\mathbf{a}) + dg(\mathbf{a}). \end{aligned}$$

PROOF. This is straightforward (*exercise*) from Proposition 18.4.4. □

The previous proposition corresponds to the fact that the partial derivatives for $\mathbf{f} + \mathbf{g}$ are the sum of the partial derivatives corresponding to $\mathbf{f}$ and $\mathbf{g}$ respectively. Similarly for $\alpha\mathbf{f}$.

Higher Derivatives We say $\mathbf{f} \in C^k(D)$ iff $f^1, \ldots, f^m \in C^k(D)$. It follows from the corresponding results for the component functions that

1. $\mathbf{f} \in C^1(D) \Rightarrow \mathbf{f}$ is differentiable in D;
2. $C^0(D) \supset C^1(D) \supset C^2(D) \supset \ldots$.

18.5. The Chain Rule. The chain rule for the composition of functions of one variable says that

$$\frac{d}{dx} g\Big(f(x)\Big) = g'\Big(f(x)\Big)\, f'(x).$$

Or to use a more informal notation, if $g = g(f)$ and $f = f(x)$, then

$$\frac{dg}{dx} = \frac{dg}{df}\frac{df}{dx}.$$

This is generalised in the following theorem. The theorem says that the linear approximation to $\mathbf{g} \circ \mathbf{f}$ (computed at $\mathbf{x}$) is the composition of the linear approximation to $\mathbf{f}$ (computed at $\mathbf{x}$) followed by the linear approximation to $\mathbf{g}$ (computed at $\mathbf{f}(\mathbf{x})$).

A Little Linear Algebra Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear* map. Then we define the *norm* of L by

$$\|L\| = \max\{|L(x)| : |x| \le 1\}^{63}.$$

A simple result *(exercise)* is that

$$|L(x)| \le ||L||\,|x| \tag{71}$$

for *any* $x \in \mathbb{R}^n$.

It is also easy to check *(exercise)* that $||\cdot||$ *does* define a norm on the vector space of linear maps from $\mathbb{R}^n$ into $\mathbb{R}^m$.

THEOREM 18.5.1 (Chain Rule). *Suppose* $\mathbf{f} : D\,(\subset \mathbb{R}^n) \to \Omega\,(\subset \mathbb{R}^m)$ *and* $g : \Omega\,(\subset \mathbb{R}^m) \to \mathbb{R}^r$*. Suppose* $\mathbf{f}$ *is differentiable at* $\mathbf{x}$ *and* $\mathbf{g}$ *is differentiable at* $\mathbf{f}(\mathbf{x})$*. Then* $\mathbf{g} \circ \mathbf{f}$ *is differentiable at* $\mathbf{x}$ *and*

$$d(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = d\mathbf{g}(\mathbf{f}(\mathbf{x})) \circ d\mathbf{f}(\mathbf{x}). \tag{72}$$

Schematically:

$$\begin{array}{ccccc} & & \xrightarrow{\quad g\circ f \quad} & & \\ D\,(\subset \mathbb{R}^n) & \xrightarrow{f} & \Omega\,(\subset \mathbb{R}^m) & \xrightarrow{g} & \mathbb{R}^r \\ & & \xrightarrow{\quad d(g\circ f)(x) = dg(f(x)) \circ df(x) \quad} & & \\ \mathbb{R}^n & \xrightarrow{df(x)} & \mathbb{R}^m & \xrightarrow{dg(f(x))} & \mathbb{R}^r \end{array}$$

[63] Here $|x|$, $|L(x)|$ are the usual Euclidean norms on $\mathbb{R}^n$ and $\mathbb{R}^m$. Thus $\|L\|$ corresponds to the maximum value taken by L on the unit ball. The maximum value *is* achieved, as L is continuous and $\{x : |x| \le 1\}$ is compact.

EXAMPLE 18.5.2. To see how all this corresponds to other formulations of the chain rule, suppose we have the following:

$$\begin{array}{ccccc} \mathbb{R}^3 & \xrightarrow{f} & \mathbb{R}^2 & \xrightarrow{g} & \mathbb{R}^2 \\ (x,y,z) & & (u,v) & & (p,q) \end{array}$$

Thus coordinates in $\mathbb{R}^3$ are denoted by (x,y,z), coordinates in the first copy of $\mathbb{R}^2$ are denoted by (u,v) and coordinates in the second copy of $\mathbb{R}^2$ are denoted by (p,q).

The functions f and g can be written as follows:

$$\begin{aligned} f &: \quad u = u(x,y,z),\ v = v(x,y,z), \\ g &: \quad p = p(u,v),\ q = q(u,v). \end{aligned}$$

Thus we think of u and v as functions of x, y and z; and p and q as functions of u and v.

We can also represent p and q as functions of x, y and z via

$$p = p\Big(u(x,y,z),\, v(x,y,z)\Big),\ q = q\Big(u(x,y,z),\, v(x,y,z)\Big).$$

The usual version of the chain rule in terms of partial derivatives is:

$$\begin{aligned} \frac{\partial p}{\partial x} &= \frac{\partial p}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial p}{\partial v}\frac{\partial v}{\partial x} \\ \frac{\partial p}{\partial x} &= \frac{\partial p}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial p}{\partial v}\frac{\partial v}{\partial x} \\ &\vdots \\ \frac{\partial q}{\partial z} &= \frac{\partial q}{\partial u}\frac{\partial u}{\partial z} + \frac{\partial q}{\partial v}\frac{\partial v}{\partial z}. \end{aligned}$$

In the first equality, $\frac{\partial p}{\partial x}$ is evaluated at (x,y,z), $\frac{\partial p}{\partial u}$ and $\frac{\partial p}{\partial v}$ are evaluated at $\Big(u(x,y,z), v(x,y,z)\Big)$, and $\frac{\partial u}{\partial x}$ and $\frac{\partial v}{\partial x}$ are evaluated at (x,y,z). Similarly for the other equalities.

In terms of the matrices of partial derivatives:

$$\underbrace{\begin{bmatrix} \frac{\partial p}{\partial x} & \frac{\partial p}{\partial y} & \frac{\partial p}{\partial z} \\ \frac{\partial q}{\partial x} & \frac{\partial q}{\partial y} & \frac{\partial q}{\partial z} \end{bmatrix}}_{d(g\circ f)(\mathbf{x})} = \underbrace{\begin{bmatrix} \frac{\partial p}{\partial u} & \frac{\partial p}{\partial v} \\ \frac{\partial q}{\partial u} & \frac{\partial q}{\partial v} \end{bmatrix}}_{dg(f(\mathbf{x}))} \underbrace{\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \end{bmatrix}}_{df(\mathbf{x})},$$

where $\mathbf{x} = (x,y,z)$.

PROOF OF CHAIN RULE: We want to show

$$(\mathbf{f}\circ\mathbf{g})(\mathbf{a}+\mathbf{h}) = (\mathbf{f}\circ\mathbf{g})(\mathbf{a}) + L(\mathbf{h}) + o(|\mathbf{h}|), \tag{73}$$

where $L = d\mathbf{f}(\mathbf{g}(\mathbf{a}))\circ d\mathbf{g}(\mathbf{a})$.

Now

$$\begin{aligned} (\mathbf{f}\circ\mathbf{g})(\mathbf{a}+\mathbf{h}) &= \mathbf{f}\Big(\mathbf{g}(\mathbf{a}+\mathbf{h})\Big) \\ &= \mathbf{f}\Big(\mathbf{g}(\mathbf{a}) + \mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})\Big) \\ &= \mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big) + \Big\langle d\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big),\, \mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})\Big\rangle \\ &\qquad + o\Big(|\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})|\Big) \\ &\qquad\qquad \dots \text{ by the differentiability of } \mathbf{f} \\ &= \mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big) + \Big\langle d\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big),\, \langle d\mathbf{g}(\mathbf{a}), \mathbf{h}\rangle + o(|\mathbf{h}|)\Big\rangle \end{aligned}$$

$$
\begin{aligned}
&\quad + o\Big(\big|\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})\big|\Big) \\
&\qquad \dots \text{ by the differentiability of } \mathbf{g} \\
&= \mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big) + \Big\langle d\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big),\, \langle d\mathbf{g}(\mathbf{a}), \mathbf{h}\rangle \Big\rangle \\
&\quad + \Big\langle d\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big),\, o(|\mathbf{h}|)\Big\rangle + o\Big(\big|\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})\big|\Big) \\
&= A + B + C + D
\end{aligned}
$$

But $B = \Big\langle d\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big) \circ d\mathbf{g}(\mathbf{a}),\, \mathbf{h}\Big\rangle$, by definition of the "composition" of two maps. Also $C = o(|\mathbf{h}|)$ from (71) (*exercise*). Finally, for D we have

$$
\begin{aligned}
\Big|\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})\Big| &= \Big|\langle d\mathbf{g}(\mathbf{a}), \mathbf{h}\rangle + o(|\mathbf{h}|)\Big| \ \dots \text{ by differentiability of } \mathbf{g} \\
&\leq ||d\mathbf{g}(\mathbf{a})||\,|\mathbf{h}| + o(|\mathbf{h}|) \ \dots \text{ from (71)} \\
&= O(|\mathbf{h}|) \ \dots \ \textit{why?}
\end{aligned}
$$

Substituting the above expressions into $A + B + C + D$, we get

$$
(\mathbf{f} \circ \mathbf{g})(\mathbf{a}+\mathbf{h}) = \mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big) + \Big\langle d\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big) \circ d\mathbf{g}(\mathbf{a}),\, \mathbf{h}\Big\rangle + o(|\mathbf{h}|). \tag{74}
$$

If follows that $\mathbf{f} \circ \mathbf{g}$ *is* differentiable at $\mathbf{a}$, *and* moreover the differential equals $d\mathbf{f}(\mathbf{g}(\mathbf{a})) \circ d\mathbf{g}(\mathbf{a})$. This proves the theorem. ∎