Analysis II, Term I

Dr. Davoud Cheraghi

October 3, 2023

Introduction to the module

This is a continuation of Analysis I module you had in year-one. In that module, you have learned about the real numbers, completeness, convergence of sequences and series, continuity and differentiability of functions on an interval or $\mathbb{R}$, integral of a function on an interval. Analysis II is a single module in year-two, delivered during term I and term II.

The content of Analysis II in term I has two parts. In the first part we complete the study of analysis on Euclidean spaces, by introducing the concepts of converges of sequences in higher dimensional Euclidean spaces $\mathbb{R}^n$, and the continuity and differentiability of maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. In the second part of the module, we generalise these notions of analysis on Euclidean spaces into a broader setting, called metric spaces and topological spaces. That is a setting where one can define the notions of converge of sequences, completeness of spaces, continuity of maps, etc. Many theorems you have learned in the previous analysis module extends into this setting, and indeed, one can give unified proofs to all those statements at once. Many theorems find a natural form in the setting of metric spaces, and you will see that the proof you already know for a statement can be adapted to the more general setting.

Any section/subsection marked with * is not examinable, but will be valuable in future courses, especially if you take pure analysis courses in your third year and beyond. You should certainly at least read through the notes on these sections, even if you choose not to attempt the questions. I will try to indicate in lectures when I'm covering those material.

Throughout this lecture notes, the definitions are numbered successively within each chapter, that is, in Chapter 1, you will see Definition 1.1, Definition 1.2, Definition 1.3, and so on. The same numbering mechanism applies to Examples, Exercises, and Remarks in each chapter. On the other hand, the results such as lemmas, propositions, corollaries, and theorems are collectively numbered in a successive fashion. That is, in Chapter 1, you will see Proposition 1.1, Theorem 1.2, Theorem 1.3, etc.

Contents

Introduction to the module i

1 Differentiation in higher dimensions 1

1.1 Euclidean spaces . . . 1

1.1.1 Preliminaries from analysis I . . . 1

1.1.2 Euclidean space of dimension n . . . 2

1.1.3 Convergence of sequences in Euclidean spaces . . . 4

1.1.4 Open sets in Euclidean spaces . . . 7

1.2 Continuity . . . 9

1.2.1 Continuity at a point, and continuity on an open set . . . 9

1.3 Derivative of a map of Euclidean spaces . . . 14

1.3.1 Derivative as a linear map . . . 14

1.3.2 Chain rule . . . 21

1.4 Directional derivatives . . . 24

1.4.1 Rates of change and partial derivatives . . . 24

1.4.2 Relation between partial derivatives and differentiability . . . 28

1.5 Higher derivatives . . . 35

1.5.1 Higher derivatives as linear maps . . . 35

1.5.2 Symmetry of mixed partial derivatives . . . 36

1.5.3 Taylor's theorem . . . 38

1.6 Inverse and Implicit function theorems . . . 40

1.6.1 Inverse function theorem . . . 40

1.6.2 Implicit Function Theorem . . . 44

1.6.3 * Sketch of the proof of the Implicit Function Theorem . . . 46

1.6.4 The general form of the Implicit Function Theorem . . . 47

1.6.5 * Equivalence of the two theorems . . . 48

2 Metric and topological spaces 50

2.1 Metric spaces . . . 50

2.1.1 Motivation and definition . . . 50

2.1.2 Examples of metric spaces . . . 52

2.1.3 Normed vector spaces 60
2.1.4 Open sets in metric spaces 62
2.1.5 Convergence in metric spaces 66
2.1.6 Closed sets in metric spaces 68
2.1.7 Interior, isolated, limit, and boundary points in metric spaces 70
2.1.8 Continuous maps of metric spaces 73
2.2 Topological spaces 78
2.2.1 Motivation 78
2.2.2 Topology on a set 78
2.2.3 Convergence, and Hausdorff property 82
2.2.4 Closed sets in topological spaces 83
2.2.5 Continuous maps on topological spaces 85
2.3 Connectedness 87
2.3.1 Connected sets 87
2.3.2 Continuous maps and connected sets 91
2.3.3 Path connected sets 92
2.4 Compactness 94
2.4.1 Compactness by covers 94
2.4.2 Sequential compactness 100
2.4.3 Continuous maps and compact sets 102
2.5 Completeness 105
2.5.1 Complete metric spaces and Banach space 105
2.5.2 Arzelà-Ascoli 110
2.5.3 Fixed point Theorem 112

Chapter 1

Differentiation in higher dimensions

1.1 Euclidean spaces

1.1.1 Preliminaries from analysis I

In this chapter we are going to extend some of the ideas that you saw last year (such as limits and continuity) to higher dimensions. The definitions are almost identical, so this should mostly feel like a review chapter to begin with, although some of the ideas we are going to approach from a different point of view.

Throughout these notes we frequently use the standard notations for the set of natural numbers

$$\mathbb{N} = \{1, 2, 3, \ldots\},$$

the set of integers

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\},$$

the set of rational numbers

$$\mathbb{Q} = \{p/q \mid p \in \mathbb{Z},\ q \in \mathbb{Z} \setminus \{0\}\},$$

and the set of real numbers $\mathbb{R}$. The set of real numbers is obtained as the *completion* of $\mathbb{Q}$. We may add, multiply and subtract elements of $\mathbb{R}$, and we can divide by elements of $\mathbb{R} \setminus \{0\}$. Note that some authors use the notation $\mathbb{N}$ to denote the set $\{0, 1, 2, \ldots\}$, but we will omit 0 from this set.

On $\mathbb{R}$ we have a notion of ordering $\leq$, so that we may say whether a real number is greater than, less than or equal to another. Moreover, $\mathbb{R}$ satisfies the **completeness axiom**, that is, if $A \subset \mathbb{R}$ is non-empty and bounded above, then A has a least upper bound. The standard notation for the least upper bound of A is $\sup(A)$.

An important function defined on all real numbers is the **modulus function**, defined as

$$|x| := \begin{cases} x & x \geq 0, \\ -x & x < 0. \end{cases}$$

This function has the following properties:

(i) for all $x \in \mathbb{R}$, we have $|x| \geq 0$, with $|x| = 0$ if and only if $x = 0$,

(ii) for all x and y in $\mathbb{R}$, $|xy| = |x|\,|y|$,

(iii) for all x and y in $\mathbb{R}$,

$$|x + y| \leq |x| + |y| .$$

The third property in the above list is called the **triangle inequality** for the modulus function.

1.1.2 Euclidean space of dimension n

For $n \geq 1$, the **n-dimensional Euclidean space**, denoted by $\mathbb{R}^n$, is defined as the set of ordered n-tuples $(x^1, x^2, \ldots, x^n)$, where each $x^i \in \mathbb{R}$, for $i = 1, 2, \ldots, n$. Each such n-tuple is denoted by a single letter $x = (x^1, x^2, \ldots, x^n)$ and will be referred to as a point in $\mathbb{R}^n$. The entries x^i are called the **coordinates** of x.

One may see each element of $\mathbb{R}^n$ as a row vector with n real components, or as a column vector with n real components. We do not make this distinction (unless when a matrix is acting on the point x. When a matrix M acts on a vector with the same components as x we use Mx^t to make it clear that x is viewed as a column vector. Here ^{t} denotes the transpose operation.)

We shall try to stick to the convention of using superscripts to label components of vectors, and subscripts to label different vectors, so that $x_1, x_2 \in \mathbb{R}^n$ are two different vectors, while $x^1, x^2 \in \mathbb{R}$ are the components of one vector.

If x and y are elements of $\mathbb{R}^n$ with

$$x = \left(x^1, \ldots, x^n\right), \qquad y = \left(y^1, \ldots, y^n\right),$$

we can add these two elements according to

$$x + y = \left(x^1 + y^1, \ldots, x^n + y^n\right).$$

Moreover, for every $\lambda \in \mathbb{R}$, we define

$$\lambda x = \left(\lambda x^1, \ldots, \lambda x^n\right).$$

With these definitions, $\mathbb{R}^n$ is a **vector space** over $\mathbb{R}$.

The **inner product**,

$$\langle \cdot\,, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},$$

is defined as

$$\left\langle (x^1, \dots, x^n), (y^1, \dots, y^n) \right\rangle = \sum_{i=1}^{n} x^i y^i.$$

Using the inner product, we may define the **length**, or **norm**, function

$$\|\cdot\| : \mathbb{R}^n \to [0, \infty)$$

as

$$\|x\| = \sqrt{\langle x, x \rangle} = \langle x, x \rangle^{1/2}.$$

Note that the inner product of two vectors is a real number, not a vector.

The norm function on $\mathbb{R}^n$ has the following properties:

(i) for all $x \in \mathbb{R}^n$, we have $\|x\| \geq 0$, with $\|x\| = 0$ if and only if $x = 0$,

(ii) for all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, $\|\lambda x\| = |\lambda| \, \|x\|$,

(iii) for all x and y in $\mathbb{R}^n$,

$$\|x + y\| \leq \|x\| + \|y\|. \tag{1.1}$$

The third property in the above list is called the **triangle inequality** for the norm on $\mathbb{R}^n$.

Remark 1.1. As we shall see later, these properties can be used in an abstract fashion to define more general "normed vector spaces". The norm gives us a useful notion of "distance" between two points, that is, the distance from x to y is given by $\|x - y\|$. Notice that if $n = 1$ we have $|\cdot| = \|\cdot\|$, and we will use either interchangeably in this case.

Exercise 1.1. (a) Show that the inner product satisfies the following properties: for all x, y, and z in $\mathbb{R}^n$ and all $a \in \mathbb{R}$,

$$\langle x, y \rangle = \langle y, x \rangle, \qquad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle, \qquad \langle ax, y \rangle = a \langle x, y \rangle.$$

(b) For $t \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$, show that:

$$\|x + ty\|^2 = \|x\|^2 + 2t \langle x, y \rangle + t^2 \|y\|^2 \geq 0. \tag{1.2}$$

(c) By thinking of (1.2) as a quadratic in t, and considering its possible roots, deduce the **Cauchy-Schwartz** inequality:

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|. \tag{1.3}$$

When does equality hold?

(d) Deduce the triangle inequality (1.1).

(e) Show the reverse triangle inequality:

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

Exercise 1.2. Suppose $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$.

(a) Show that:

$$\max_{k=1,\ldots,n} \left|x^k\right| \leq \|x\| \,. \tag{1.4}$$

(b) Show that:

$$\|x\| \leq \sqrt{n} \max_{k=1,\ldots,n} \left|x^k\right|. \tag{1.5}$$

1.1.3 Convergence of sequences in Euclidean spaces

Now that we have a few definitions relating to $\mathbb{R}^n$, we're ready to revisit some concepts from first year analysis and see how they can be extended to higher dimensions.

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

$$x_0, x_1, x_2, \ldots,$$

with each $x_i \in \mathbb{R}^n$, for $i = 0, 1, 2, \ldots$. This is often written $(x_i)_{i=0}^{\infty}$, or $(x_i)_{i\in\mathbb{N}}$. A very important concept relating to sequences is convergence.

Definition 1.1. A sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ **converges** to (the vector) $x \in \mathbb{R}^n$ if the following holds: For every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i \geq N$ we have

$$\|x_i - x\| < \epsilon.$$

We then write:

$$x_i \to x, \quad \text{as } i \to \infty,$$

or

$$\lim_{i\to\infty} x_i = x.$$

One may compare the above definition to the one for convergence of a sequence of real numbers. Indeed, this notion is intimately related to convergence of real numbers, as stated in the next lemma.

Proposition 1.1. *The sequence of vectors $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ converges to the vector $x \in \mathbb{R}^n$ if and only if each component of x_i converges to the corresponding component of x. That is, if we write:*

$$x_i = (x_i^1, \ldots, x_i^n), \quad \textit{and} \quad x = (x^1, \ldots x^n),$$

then, $x_i \to x$ as $i \to \infty$ if and only if for all $k = 1, \ldots n$, $x_i^k \to x^k$ as $i \to \infty$.

Proof. Let us first assume that for all $k = 1, 2, \dots, n$,

$$x_i^k \to x^k, \quad \text{as } i \to \infty.$$

Fix an arbitrary $\epsilon > 0$. Then, for each $k = 1, \dots, n$, we apply the definition of convergence of $x_i^k \to x^k$ to $\epsilon/\sqrt{n}$ to obtain $N_k \in \mathbb{N}$ such that for all $i \geq N_k$ we have

$$\left|x_i^k - x^k\right| < \frac{\epsilon}{\sqrt{n}}.$$

Let $N = \max\{N_1 \dots, N_n\}$. Then, for every $i \geq N$, we have

$$\max_{k=1,\dots,n} \left|x_i^k - x^k\right| < \frac{\epsilon}{\sqrt{n}}.$$

Now, recall from the inequality in (1.4) that for every $y = (y^1, y^2, \dots, y^n) \in \mathbb{R}^n$,

$$\|y\| \leq \sqrt{n} \max_{k=1,\dots,n} \left|y^k\right|,$$

so we deduce

$$\|x_i - x\| \leq \sqrt{n} \max_{k=1,\dots,n} \left|x_i^k - x^k\right| < \epsilon.$$

This establishes the result in one direction.

Now assume that

$$\lim_{i\to\infty} x_i = x.$$

Fix an arbitrary integer k with $1 \leq k \leq n$, and an arbitrary $\epsilon > 0$. We aim to show that $x_i^k \to x^k$, as $i \to \infty$. The definition of convergence of $x_i \to x$, as $i \to \infty$, with ϵ, gives us $N \in \mathbb{N}$ such that for all $i \geq N$ we have

$$\|x_i - x\| < \epsilon.$$

Recall from Exercise 1.1, Equation (1.5) that for every $y = (y^1, y^2, \dots, y^n) \in \mathbb{R}^n$,

$$\max_{k=1,\dots,n} \left|y^k\right| \leq \|y\|.$$

In particular, for all $i \geq N$, we have

$$\left|x_i^k - x^k\right| \leq \max_{k=1,\dots,n} \left|x_i^k - x^k\right| \leq \|x_i - x\| < \epsilon.$$

As $\epsilon > 0$ was arbitrary, this shows that x_i^k converges to x^k, as $i \to \infty$. □

Exercise 1.3. Suppose that $(x_i)_{i=0}^\infty$ and $(y_i)_{i=0}^\infty$ are two sequences in $\mathbb{R}^n$ with

$$\lim_{i\to\infty} x_i = x, \quad \lim_{i\to\infty} y_i = y.$$

(a) Show that

$$x_i + y_i \to x + y \quad \text{as } i \to \infty.$$

(b) Show that

$$\langle x_i, y_i \rangle \to \langle x, y \rangle \qquad \text{as } i \to \infty,$$

deduce that

$$\|x_i\| \to \|x\| \qquad \text{as } i \to \infty.$$

(c) Suppose that $(a_i)_{i=0}^{\infty}$ is a sequence in $\mathbb{R}$ with $a_i \to a$ as $i \to \infty$. Show that:

$$a_i x_i \to ax, \qquad \text{as } i \to \infty.$$

1.1.4 Open sets in Euclidean spaces

In dimension one, you are familiar with sets of the form (a, b) and $[a, b]$, i.e. the open interval and the closed interval respectively. These form natural domains for functions in dimension one, and it is fairly general to present theorems about maps in dimension one on such intervals. In higher dimensions, one may generalise these sets to sets of the from

$$\begin{aligned}&(a^1, b^1) \times (a^2, b^2) \times \cdots \times (a^n, b^n)\\ &\qquad = \left\{(x^1, x^2, \ldots, x^n) \in \mathbb{R}^n \mid \text{ for } 1 \leq i \leq n, a^i < x^i < b^i\right\},\end{aligned}$$

or

$$\begin{aligned}&[a^1, b^1] \times [a^2, b^2] \times \cdots \times [a^n, b^n]\\ &\qquad = \left\{(x^1, x^2, \ldots, x^n) \in \mathbb{R}^n \mid \text{ for } 1 \leq i \leq n, a^i \leq x^i \leq b^i\right\}.\end{aligned}$$

But this is very restrictive and does not capture the same level of generality of intervals in dimension one. The domains of maps in higher dimensions may appear in many forms. Due to this, we present a class of subsets of $\mathbb{R}^n$, called open sets.

For $x \in \mathbb{R}^n$ and the real number $r > 0$, the **open ball** of radius r about x is defined as the set

$$B_r(x) = \{y \in \mathbb{R}^n : \|x - y\| < r\}.$$

That is, $B_r(x)$ consists of all points in $\mathbb{R}^n$ which are at distance less than r from x. We sometimes denote the open ball $B_r(x)$ by $B(x, r)$. Both notations are widely used in mathematics.

Definition 1.2. A set $U \subseteq \mathbb{R}^n$ is called **open in** $\mathbb{R}^n$, if for every $x \in U$ there exists $r > 0$ such that $B_r(x) \subseteq U$.

In other words, about any point in an open set we can find a small ball which is entirely contained in the set. Note that in this definition, the radius of the ball is allowed to depend on x. See Figure 1.1.4.

We may compare the above definition with the definition of open sets in $\mathbb{R}$ you saw in Analysis I. Recall that a set $I \subseteq \mathbb{R}$ is open in $\mathbb{R}$, if for every $x \in I$, there is $\delta > 0$ such that $(x - \delta, x + \delta) \subseteq I$. This definition is consistent with the one we have given in $\mathbb{R}^n$, since in $\mathbb{R}^1$, $B_\delta(x) = (x - \delta, x + \delta)$.

Example 1.1. The ball $B_1(0)$ is open in $\mathbb{R}^n$. To see this, suppose $x \in B_1(0)$, so that $\|x\| < 1$. Let $r = (1 - \|x\|)/2$. We need to show that $B_r(x) \subseteq B_1(0)$. To that end, let $y \in B_r(x)$ be an arbitrary point. Using the triangle inequality for the norm in $\mathbb{R}^n$, we have

$$\|y\| = \|y - x + x\| \leq \|y - x\| + \|x\| < r + \|x\| = \frac{1 - \|x\|}{2} + \|x\| < \frac{1 + \|x\|}{2} < 1.$$

This means that $y \in B_1(0)$.

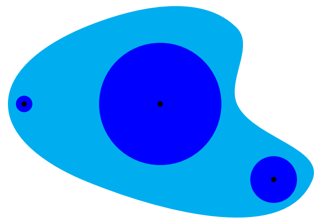

Figure 1.1: An open set in $\mathbb{R}^2$ in cyan, and some balls inside it. The radius of the ball depends on the location of the point.

Observe that in the above example, one can replace 1 with any other positive real number, and the result is still valid. That is, for every $\delta > 0$, the set $B_\delta(0)$ is open in $\mathbb{R}^n$. Similarly, one can also replace 0 with any $y \in \mathbb{R}^n$. Thus, in general, for any $y \in \mathbb{R}^n$ and any $\delta > 0$, $B_\delta(y)$ is open in $\mathbb{R}^n$.

Example 1.2. The set $A = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is not open. Clearly $y := (1, 0, \ldots, 0)$ belongs to A. On the other hand, if $r > 0$ then $z = (1 + r/2, 0, \ldots, 0)$ belongs to $B_r(y)$ but not to A, so there is no $r > 0$ such that $B_r(y) \subset A$.

Exercise 1.4. Which of the following subsets of $\mathbb{R}^n$ is open:

(a) $\mathbb{R}^n$?

(b) $\emptyset$?

(c) $\{x = (x^1, \ldots, x^n) \in \mathbb{R}^n \mid x^1 > 0\}$?

(d) $\{x = (x^1, \ldots, x^n) \in \mathbb{R}^n \mid \forall i,\ x^i \in [0, 1)\}$?

(e) $\{x = (x^1, \ldots, x^n) \in \mathbb{R}^n \mid \forall i,\ x^i \in \mathbb{Q}\}$?

Exercise 1.5. Let $(x_i)_{i=0}^{\infty}$ be a sequence in $\mathbb{R}^n$ with $\lim_{i\to\infty} x_i = x \in \mathbb{R}^n$. Assume that there is $r > 0$ such that for all $i \geq 0$, we have $\|x_i\| < r$. Show that

$$\|x\| \leq r.$$

Exercise 1.6. (a) Show that if U_1 and U_2 are open sets in $\mathbb{R}^n$, then $U_1 \cup U_2$ and $U_1 \cap U_2$ are open in $\mathbb{R}^n$.

(b) Suppose that U_α, for α in an index set I, are open sets in $\mathbb{R}^n$.

(i) Show that the set $\bigcup_{\alpha \in I} U_\alpha$ is open in $\mathbb{R}^n$.

(ii) Give an example showing that $\bigcap_{\alpha \in I} U_\alpha$ need not be open.

Remark 1.2. It is worth noting that the notion of open sets in $\mathbb{R}^n$ relies on the length function $\|\cdot\|$ we have on $\mathbb{R}^n$. As we shall see in the next chapter, one can consider functions (called metric) with similar properties on a wide range of other sets (such as the set of all continuous functions from $[0,1]$ to $\mathbb{R}$ or the set of all sequences in $[0,1]$, etc). These lead to notions of open sets on such sets. We will look into this in the next chapter.

1.2 Continuity

Last year, you learned about the notion of continuity for functions from $\mathbb{R}$ (or subsets of $\mathbb{R}$) to $\mathbb{R}$. In this section we revisit those definitions and upgrade them to higher dimensions. In fact, the definitions we shall give are almost identical: the only thing that changes is that we use the appropriate "norm" for the domain and range.

1.2.1 Continuity at a point, and continuity on an open set

We start with the simple definition

Definition 1.3. Let $A \subset \mathbb{R}^n$ be an open set, and suppose $f : A \to \mathbb{R}^m$. We say that f is **continuous at** $p \in A$ if the following holds: for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $\|x - p\| < \delta$ we have

$$\|f(x) - f(p)\| < \epsilon.$$

If f is continuous at every p in A, we say f is **continuous on** A.

We can think of this as saying "f maps points in A close to p to points in $\mathbb{R}^m$ close to $f(p)$". Notice that in the definition above, the symbol $\|\cdot\|$ is playing two slightly different roles: as the norm on $\mathbb{R}^n$ and the norm on $\mathbb{R}^m$.

Remark 1.3. The words "function" and "map" are not identical. For $f : X \to Y$, we use the word "function" when the target space Y is the real numbers or the complex numbers (or in general a field). Otherwise, we use the word "map". Of course it is correct to refer to $f : X \to \mathbb{R}$ as a map, but it is uncommon to refer to $f : X \to Y$ as a function, when Y is not a set of numbers where one can not add and multiply elements. On the other hand, it is common in analysis and geometry to see expressions like, "let f be a function on X", which means that $f : X \to \mathbb{R}$ or $f : X \to \mathbb{C}$. In those cases, the target space is understood from the context.

Example 1.3. The map $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) = \|x\|$ is continuous on $\mathbb{R}^n$.

To show this, fix an arbitrary $p \in \mathbb{R}^n$. Suppose $\|x - p\| < \delta$, then by the reverse triangle inequality (see Exercise 1.1) we have:

$$|f(x) - f(p)| = \big| \|x\| - \|p\| \big| \leq \|x - p\| < \delta.$$

Thus we can take $\delta = \epsilon$ and we have satisfied the criteria for continuity of f at p.

Example 1.4. Every linear map $\Lambda : \mathbb{R}^n \to \mathbb{R}^m$ is continuous.

Let $\{e_j\}_{j=1}^n$ be the canonical basis for $\mathbb{R}^n$, that is, all entries of e_j are 0 except the j-th entry which is 1. We may define the real number

$$M = \max_{j=1,\ldots,n} \|\Lambda(e_j)\| .$$

We note that,

$$\begin{aligned} \|\Lambda(x) - \Lambda(p)\| = \|\Lambda(x-p)\| &= \left\| \Lambda\Big(\sum_{j=1}^n e_j (x-p)^j \Big) \right\| \\ &= \left\| \sum_{j=1}^n (x-p)^j \Lambda(e_j) \right\| \\ &\leq \sum_{j=1}^n \big\| (x-p)^j \Lambda(e_j) \big\| \\ &\leq \sum_{j=1}^n \big| (x-p)^j \big| \, \|\Lambda(e_j)\| \\ &\leq M \sum_{j=1}^n \big| (x-p)^j \big| \end{aligned}$$

Thus, using the inequality in Equation (1.4),

$$\|\Lambda(x) - \Lambda(p)\| \leq M \sum_{j=1}^n \|x - p\| = Mn \|x - p\| .$$

Thus, if we take $\delta = \epsilon/(2Mn)$, then for any x with $0 < \|x - p\| < \delta$, we have

$$\|\Lambda(x) - \Lambda(p)\| < \frac{\epsilon}{2Mn} Mn < \epsilon,$$

so Λ is continuous.

Example 1.5. The map $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x^1, \ldots, x^n) = x^1$ is continuous on $\mathbb{R}^n$.

To see this, fix an arbitrary $p \in \mathbb{R}^n$. Suppose $\|x - p\| < \delta$, then by the inequality in (1.5) we have:

$$|f(x) - f(p)| = \big| x^1 - p^1 \big| \leq \max_{k=1,\ldots,n} \left| x^k - p^k \right| \leq \|x - p\| < \delta,$$

so we may take $\delta = \epsilon$ and we have satisfied the condition for continuity. Obviously the same argument shows that all of the coordinate maps (i.e. the map taking x to x^k) are continuous.

Theorem 1.2. *Let A be an open subset of $\mathbb{R}^n$ and B be an open subset of $\mathbb{R}^m$. Suppose $f : A \to B$ is continuous at p and $g : B \to \mathbb{R}^l$ is continuous at $f(p)$. Then $g \circ f : A \to \mathbb{R}^l$ is continuous at p.*

Proof. Fix an arbitrary $\epsilon > 0$. Since g is continuous at $f(p)$, we know that there exists $\delta_1 > 0$ such that for any $y \in B$ with $\|y - f(p)\| < \delta_1$, we have $\|g(y) - g(f(p))\| < \epsilon$. Similarly, since f is continuous at p, we know that there exists $\delta > 0$ such that for any $x \in A$ with $\|x - p\| < \delta$, we have $\|f(x) - f(p)\| < \delta_1$. Combining these two statements and taking $y = f(x)$, we deduce that if $x \in A$ with $\|x - p\| < \delta$, we have $\|g(f(x)) - g(f(p))\| < \epsilon$. $\square$

It is sometimes useful to express the continuity of a map in a slightly different way, for which we need the following definition:

Definition 1.4. Let A be an open subset of $\mathbb{R}^n$ and suppose $f : A \to \mathbb{R}^m$. For $p \in A$, we say that the **limit** of f as x tends to p is equal to $q \in \mathbb{R}^m$, if the following holds: for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in A$ with $0 < \|x - p\| < \delta$ we have

$$\|f(x) - q\| < \epsilon.$$

In this case, we write

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

Note that in the above definition, we do not allow $x = p$. With this notion of a limit in hand, we can give the definition of continuity more compactly as:

"f is continuous at p, if $\lim_{x \to p} f(x) = f(p)$."

Theorem 1.3. *Suppose A is an open subset of $\mathbb{R}^n$, $p \in A$, and $f, g : A \to \mathbb{R}$ with*

$$\lim_{x \to p} f(x) = F, \qquad \lim_{x \to p} g(x) = G.$$

Then

(i) $\lim_{x \to p} (f(x) + g(x)) = F + G$,

(ii) $\lim_{x \to p} (f(x)g(x)) = FG$,

(iii) If, furthermore $G \neq 0$, then:

$$\lim_{x \to p} \frac{f(x)}{g(x)} = \frac{F}{G}.$$

Proof. (i) Fix an arbitrary $\epsilon > 0$. Since $\lim_{x\to p} f(x) = F$, we know that there exists $\delta_1 > 0$ such that for every $x \in A$ with $0 < \|x - p\| < \delta_1$,

$$|f(x) - F| < \frac{\epsilon}{2}.$$

Similarly, there exists $\delta_2 > 0$ such that for every $x \in A$ with $0 < \|x - p\| < \delta_2$,

$$|g(x) - G| < \frac{\epsilon}{2}.$$

Define $\delta = \min\{\delta_1, \delta_2\}$. Evidently $\delta > 0$. For every $x \in A$ with $0 < \|x - p\| < \delta$, by the triangle inequality, we have

$$|f(x) + g(x) - (F + G)| \leq |f(x) - F| + |g(x) - G| < \epsilon.$$

(ii) Fix an arbitrary $\epsilon > 0$, and assume without loss of generality that $\epsilon < 3$ (Why can see assume this?). Since $\lim_{x\to p} f(x) = F$, we know that there exists $\delta_1 > 0$ such that for every $x \in A$ with $0 < \|x - p\| < \delta_1$,

$$|f(x) - F| < \frac{\epsilon}{3(1 + |G|)}.$$

Similarly, there exists $\delta_2 > 0$ such that for every $x \in A$ with $0 < \|x - p\| < \delta_2$,

$$|g(x) - G| < \frac{\epsilon}{3(1 + |F|)}.$$

To control $f(x)g(x) - FG$, we add and subtract the same terms, so that we obtain terms of the form $f(x) - F$ and $g(x) - G$. That is,

$$\begin{aligned} f(x)g(x) - FG &= f(x)g(x) - f(x) \cdot G + f(x) \cdot G - F \cdot G \\ &= f(x)(g(x) - G) + (f(x) - F) \cdot G \\ &= (f(x) - F + F)(g(x) - G) + (f(x) - F) \cdot G \\ &= (f(x) - F)(g(x) - G) + F \cdot (g(x) - G) + (f(x) - F) \cdot G \end{aligned}$$

Now, take $\delta = \min\{\delta_1, \delta_2\}$. For every $x \in A$ with $0 < \|x - p\| < \delta$, by the triangle inequality, we have

$$\begin{aligned} |f(x)g(x) - FG| &\leq |f(x) - F|\,|g(x) - G| + |F|\,|g(x) - G| + |G|\,|f(x) - F| \\ &< \frac{\epsilon^2}{9(1 + |F|)(1 + |G|)} + \frac{\epsilon\,|F|}{3(1 + |F|)} + \frac{\epsilon\,|G|}{3(1 + |G|)} \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

(iii) Given the previous part, it suffices to show that if $\lim_{x\to p} g(x) = G$ with $G \neq 0$, then

$$\lim_{x\to p} \frac{1}{g(x)} = \frac{1}{G}.$$

Fix an arbitrary $\epsilon > 0$. Since $\lim_{x\to p} g(x) = G$, we know that there exist $\delta_1 > 0$ such that for every $x \in A$ with $0 < \|x - p\| < \delta_1$,

$$|g(x) - G| < \frac{\epsilon |G|^2}{2}.$$

Also, since $G \neq 0$, $G/2 > 0$, and hence, there is $\delta_2 > 0$ such that for every $x \in A$ with $0 < \|x - p\| < \delta_2$,

$$|g(x) - G| < \frac{|G|}{2}.$$

By the triangle inequality, this implies that

$$|g(x)| = |g(x) - G + G| \geq |G| - |g(x) - G| > |G| - \frac{|G|}{2} = \frac{|G|}{2}.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. For every $x \in A$ with $0 < \|x - p\| < \delta$, we have

$$\left|\frac{1}{g(x)} - \frac{1}{G}\right| = |G - g(x)| \cdot \frac{1}{|G|} \cdot \frac{1}{|g(x)|} < \frac{\epsilon |G|^2}{2} \cdot \frac{1}{|G|} \cdot \frac{2}{|G|} = \epsilon.$$

This completes the proof.

□

Corollary 1.4. *Suppose A is an open set in $\mathbb{R}^n$ and $f, g : A \to \mathbb{R}$ are continuous at $p \in A$. Then,*

(i) $f + g$ is continuous at p.

(ii) fg is continuous at p.

(iii) If, furthermore $g(p) \neq 0$, then $\frac{f}{g}$ is continuous at p.

Exercise 1.7. Assume that A is an open set in $\mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Show that $\lim_{x\to p} f(x) = F$, if and only if, for any sequence $(x_i)_{i=0}^{\infty}$ in $A \setminus \{p\}$ with $\lim_{i\to\infty} x_i = p$,

$$\lim_{i\to\infty} f(x_i) = F.$$

Exercise 1.8. (a) Show that the map $f : \mathbb{R} \to \mathbb{R}^n$ defined as $f(x) = (x, 0, \ldots, 0)$ is continuous on $\mathbb{R}$.

(b) Let A be an open set in $\mathbb{R}^n$ and $f^1, f^2, \ldots, f^m$ are functions from A to $\mathbb{R}$. Consider the map $f : A \to \mathbb{R}^m$ defined as

$$f(x^1, \ldots, x^n) \mapsto \left(f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)\right).$$

Show that f is continuous at $p \in A$, if and only if, for every $k = 1, \ldots, m$ the map $f^k : A \to \mathbb{R}$ is continuous at p.

(c) Show that the map $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x^1, x^2, \dots, x^n) = 3x^1(x^2)^5 + 4x^2(x^n)^7$ is continuous on $\mathbb{R}^n$. Here, $(x^j)^m$ denotes the coordinate x^j raised to power m.

With the above results, one can build many continuous maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. For example,

$$(x^1, x^2) \mapsto \left(\sin(x^1x^2), \cos(x^2)\right),$$
$$(x^1, x^2, x^3) \mapsto \left(\frac{x^1 - x^2}{1 + (x^2)^2}, e^{x^3}\right).$$

Exercise 1.9 (*). (a) Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous on $\mathbb{R}^n$, and suppose $U \subset \mathbb{R}^m$ is open. Show that:

$$f^{-1}(U) := \{x \in \mathbb{R}^n : f(x) \in U\}$$

is open.

(b) Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ has the property that $f^{-1}(U) \subset \mathbb{R}^n$ is open for every open set $U \subset \mathbb{R}^m$. Show that f is continuous on $\mathbb{R}^n$.

1.3 Derivative of a map of Euclidean spaces

So far, when differentiating functions, we've restricted ourselves to the situation where the function depends only on one variable. This covers lots of situations that we're interested in, but of course we often wish to consider maps of more than one variable. In this chapter we will see how the idea of differentiation can be extended to maps which send (subsets of) $\mathbb{R}^n$ to $\mathbb{R}^m$. The basic idea will be that the derivative of a map at a point p should be the "best linear approximation" to the map at p.

1.3.1 Derivative as a linear map

Before we think about how to define a derivative of a map in higher dimensions, let's first note some of the potential challenges. In one dimension, we say that f is differentiable at p if the limit

$$\lim_{x \to p} \frac{f(x) - f(p)}{x - p}$$

exists. If $x, p \in \mathbb{R}^n$ and $f(x), f(p) \in \mathbb{R}^m$ then we obviously have a problem: we don't even know how to make sense of 'dividing by $x - p$', and it's not clear what sort of object we should end up with.

To try and find a way through this impasse, let's just remind ourselves how the derivative is introduced in one dimension. By approximating with successive chords, we consider the tangent to the graph of f at p (see Figure 1.2). Let us think a little

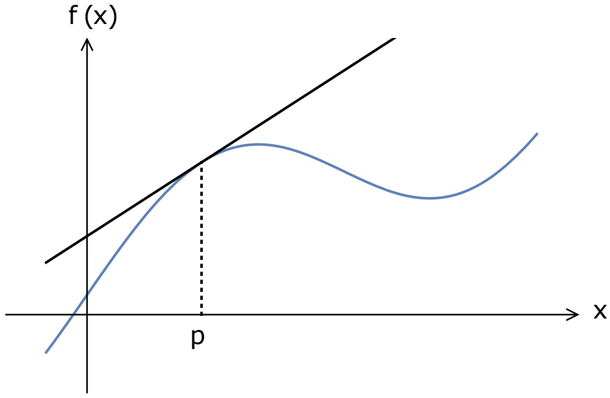

Figure 1.2: The tangent to f at p.

about how the tangent is characterised. Any (non-vertical) straight line passing through $(p, f(p))$ is the graph of the affine map

$$A_\lambda : x \mapsto \lambda(x-p) + f(p)$$

for some $\lambda \in \mathbb{R}$. Let's consider the difference between f and such an affine map

$$f(x) - A_\lambda(x) = f(x) - f(p) - \lambda(x-p).$$

In general, from the continuity of f we know that for any $\lambda \in \mathbb{R}$,

$$\lim_{x \to p} [f(x) - A_\lambda(x)] = 0. \tag{1.6}$$

However, if f is differentiable, there is a unique choice of λ that allows us to make a stronger statement. If f is differentiable, there exists a unique $\lambda \in \mathbb{R}$ such that

$$\lim_{x \to p} \frac{|f(x) - A_\lambda(x)|}{|x-p|} = 0.$$

This is a stronger statement than (1.6) because it tells us that $f(x) - A_\lambda(x)$ is going to zero faster than $|x-p|$, as $x \to p$. We make this informal discussion more precise in the following lemma.

Lemma 1.5. *The map $f : (a,b) \to \mathbb{R}$ is differentiable at $p \in (a,b)$ if and only if there exists a map of the from $A_\lambda(x) = \lambda(x-p) + f(p)$, for some $\lambda \in \mathbb{R}$, such that*

$$\lim_{x \to p} \frac{|f(x) - A_\lambda(x)|}{|x-p|} = 0.$$

Proof. We can re-write

$$\frac{|f(x) - f(p) - \lambda(x-p)|}{|x-p|} = \left| \frac{f(x) - f(p)}{x-p} - \lambda \right|,$$

so that
$$\lim_{x\to p}\frac{|f(x)-A_\lambda(x)|}{|x-p|}=0 \qquad\Longleftrightarrow\qquad \lim_{x\to p}\frac{f(x)-f(p)}{x-p}=\lambda.$$
The expression on the right-hand side of the above equation is the definition of differentiability of f at p. □

We may rewrite
$$A_\lambda(x)=\lambda(x-p)+f(p)=\lambda x+(f(p)-\lambda p).$$
Thus, $A_\lambda:\mathbb{R}\to\mathbb{R}$ is the composition of the linear map $x\mapsto\lambda x$ and the translation $x\mapsto x+(f(p)-\lambda p)$. Such maps are called affine maps of $\mathbb{R}$. By the above lemma, the map f is differentiable at p, if it is "well approximated" by an affine map at p. We may generalise this to higher dimensions.

Since we are going to frequently apply linear an nonlinear maps to variables, to distinguish between these two cases, we shall use the notation $h[v]$ when h is a linear map and v is seen as a vector, and use $h(v)$ when h is a map and v is seen as a point in the domain of h.

Let $L(\mathbb{R}^n;\mathbb{R}^m)$ denote the set of all linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. Recall that $\Lambda:\mathbb{R}^n\to\mathbb{R}^m$ is a linear map if
$$\Lambda[x+y]=\Lambda[x]+\Lambda[y],\qquad \forall x,y\in\mathbb{R}^n,$$
$$\Lambda[ax]=a\Lambda[x],\qquad \forall a\in\mathbb{R}\text{ and }x\in\mathbb{R}^n.$$

In analogy to the statement in Lemma 1.5 we propose the following definition.

Definition 1.5. Suppose $\Omega\subset\mathbb{R}^n$ is open. The map $f:\Omega\to\mathbb{R}^m$ is **differentiable** at $p\in\Omega$, if there exists a linear map $\Lambda\in L(\mathbb{R}^n;\mathbb{R}^m)$ such that
$$\lim_{x\to p}\frac{\|f(x)-(\Lambda[x-p]+f(p))\|}{\|x-p\|}=0.$$
In this case, we write
$$Df(p):=\Lambda,$$
and call $Df(p)$ the derivative of the map f at the point p.

Note that some authors refer to the derivative of a map as **total derivative**, or **differential**. We shall refer to that as derivative.

It is often useful to have the following equivalent characterisation of differentiability in higher dimensions: $f:\Omega\to\mathbb{R}^m$ is differentiable at $p\in\Omega$ if and only if there exists $\Lambda\in L(\mathbb{R}^n;\mathbb{R}^m)$ such that
$$\lim_{h\to 0}\frac{\|f(p+h)-f(p)-\Lambda[h]\|}{\|h\|}=0.$$

Note that in the above equation, $h \to 0$ in $\mathbb{R}^n$.

Recall that using a canonical basis for $\mathbb{R}^n$ and $\mathbb{R}^m$ any linear map $\Lambda \in L(\mathbb{R}^n; \mathbb{R}^m)$ can be expressed as an $m \times n$ matrix which is called the **Jacobian** of f at p. The convention is that an $m \times n$ matrix has m rows and n columns. For the purposes of this course, we won't make a big deal of the difference between a linear map and its matrix representation with respect to the canonical basis, so will use the words derivative and Jacobian essentially indistinguishably.

Lemma 1.6. *Let $\Omega \subset \mathbb{R}^n$ be an open set. If $f : \Omega \to \mathbb{R}^m$ is differentiable at $p \in \Omega$, then it is continuous at p.*

Proof. Since

$$\lim_{h \to 0} \frac{\|f(p+h) - f(p) - \Lambda[h]\|}{\|h\|} = 0,$$

we must have

$$\lim_{h \to 0} \|f(p+h) - f(p) - \Lambda[h]\| = 0.$$

On the other hand, since linear maps are continuous, see Example 1.4, we obtain

$$0 = \lim_{h \to 0} (f(p+h) - f(p) - \Lambda[h]) = \lim_{h \to 0} (f(p+h) - f(p)).$$

□

Example 1.6. By Lemma 1.5 any function $f : (a, b) \to \mathbb{R}$ which is differentiable at p satisfies the conditions of 1.5 with $Df(p) = f'(p)$. Notice that a 1×1 matrix is simply a real number.

Example 1.7. Let $B \in L(\mathbb{R}^n; \mathbb{R}^m)$ and $V \in \mathbb{R}^m$. Then, the map $f : \mathbb{R}^n \to \mathbb{R}^m$ defined as

$$f(x) = B(x) + V$$

is differentiable at each $p \in \mathbb{R}^n$, and $Df(p) = B$. To see this, note that

$$\begin{aligned} f(p+h) - f(p) - B(h) &= (B(p+h) + V) - (B(p) + V) - B(h) \\ &= B(p) + B(h) + V - B(p) - V - B(h) = 0. \end{aligned}$$

Thus,

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

Example 1.8. The map $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(x) = \|x\|^2$$

is differentiable at each $p \in \mathbb{R}^n$, and $Df(p)$ is the linear map

$$Df(p)[h] = 2 \langle p, h \rangle, \quad \forall h \in \mathbb{R}^n.$$

From the properties of the inner product in Exercise 1.1-(a), we can see that the map $h \mapsto 2\langle p, h\rangle$ is a linear map.

We note that

$$f(p+h) = \|p+h\|^2 = \langle p+h, p+h\rangle = \|p\|^2 + 2\langle p, h\rangle + \|h\|^2,$$

so that

$$\lim_{h\to 0} \frac{\|f(p+h) - f(p) - 2\langle p, h\rangle\|}{\|h\|} = \lim_{h\to 0} \|h\| = 0.$$

As a matrix, we have that $Df(p) = 2p$, where p is viewed as a row vector with n components (this is in line with our convention that a $1 \times n$ matrix maps $\mathbb{R}^n$ to $\mathbb{R}^1$). So the Jacobian is a row vector for this map.

Example 1.9. Let $m \geq 1$ be an integer, and assume that for $i = 1, 2, \ldots, m$, the map $f^i : (a, b) \to \mathbb{R}$ is differentiable at $p \in (a, b)$. Then the map $f : (a, b) \to \mathbb{R}^m$ defined as

$$f(x) = \left(f^1(x), f^2(x), \ldots, f^m(x)\right),$$

is differentiable at p, and the derivative $Df(p) : \mathbb{R} \to \mathbb{R}^m$ has the matrix representation

$$Df(p) = \begin{pmatrix} (f^1)'(p) \\ \vdots \\ (f^m)'(p) \end{pmatrix}.$$

To see this, we note that

$$f(p+h) - f(p) - \begin{pmatrix} (f^1)'(p) \\ \vdots \\ (f^m)'(p) \end{pmatrix} h = \begin{pmatrix} f^1(p+h) - f^1(p) - (f^1)'(p)h \\ \vdots \\ f^m(p+h) - f^m(p) - (f^m)'(p)h \end{pmatrix}$$

so that, using the inequality in (1.5),

$$\frac{\|f(p+h) - f(p) - Df(p)[h]\|}{\|h\|} \leq \sqrt{m} \max_{j=1,\ldots,m} \frac{\left|f^j(p+h) - f^j(p) - (f^j)'(p)h\right|}{|h|}.$$

Since each f^j is differentiable at p, the left hand side of the above equation tends to 0, as $h \to 0$. And since the left hand side of the equation is non-negative, it must tend to 0, as $h \to 0$. Notice here that the expression $Df(p)[h]$ means applying the linear map $Df(p)$ to the one dimensional vector h, which gives us an element of $\mathbb{R}^m$.

Implicitly in the discussion above, we've assumed that $Df(p)$, if it exists, must be unique. Of course, this is something that we need to prove.

Theorem 1.7. *The derivative, if it exists, is unique.*

Proof. Suppose $\Omega \subset \mathbb{R}^n$ is open, $f : \Omega \to \mathbb{R}^m$, $p \in \Omega$ and that Λ and Λ' satisfy:

$$\lim_{h \to 0} \frac{\|f(p+h) - f(p) - \Lambda[h]\|}{\|h\|} = \lim_{h \to 0} \frac{\|f(p+h) - f(p) - \Lambda'[h]\|}{\|h\|} = 0.$$

Let e be an arbitrary vector in $\mathbb{R}^n$ with $\|e\| = 1$. Then for any real number $\alpha \neq 0$ we have

$$\frac{\Lambda[\alpha e]}{\alpha} = \Lambda[e].$$

Now, let $(\alpha_j)_{j=0}^{\infty}$ be a sequence of non-zero real numbers tending to 0 as $j \to \infty$. By adding and subtracting identical terms, we see that

$$\begin{aligned}
&\|\Lambda[e] - \Lambda'[e]\| \\
&\quad = \left\| \frac{\Lambda[\alpha_j e]}{\alpha_j} - \frac{\Lambda'[\alpha_j e]}{\alpha_j} \right\| \\
&\quad = \lim_{j \to \infty} \frac{\|\Lambda[\alpha_j e] - \Lambda'[\alpha_j e]\|}{\|\alpha_j e\|} \\
&\quad = \lim_{j \to \infty} \frac{\|-f(p + \alpha_j e) + f(p) + \Lambda[\alpha_j e] + f(p + \alpha_j e) - f(p) - \Lambda'[\alpha_j e]\|}{\|\alpha_j e\|} \\
&\quad \leq \lim_{j \to \infty} \frac{\|f(p + \alpha_j e) - f(p) - \Lambda[\alpha_j e]\|}{\|\alpha_j e\|} + \lim_{j \to \infty} \frac{\|f(p + \alpha_j e) - f(p) - \Lambda'[\alpha_j e]\|}{\|\alpha_j e\|} \\
&\quad = 0.
\end{aligned}$$

For the last equality in the above equation we have used that $\alpha_j e \to 0$ as $j \to \infty$. By the above equation, for any unit vector e we have $\Lambda[e] = \Lambda'[e]$, which implies that (as linear maps) $\Lambda = \Lambda'$. □

Exercise 1.10. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is given by $f(x) = x$. Show that f is differentiable at each $p \in \mathbb{R}^n$ and

$$Df(p) = \mathrm{id},$$

where $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is the identity map.

Exercise 1.11. Show that the map $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f : (x, y) \mapsto x^2 + y^2,$$

is differentiable at all points $p = (\xi, \eta) \in \mathbb{R}^2$ with Jacobian

$$Df(p) = \begin{pmatrix} 2\xi & 2\eta \end{pmatrix}.$$

Exercise 1.12. One might hope that the derivative can be calculated by finding

$$\lim_{x \to p} \frac{f(x) - f(p)}{\|x - p\|}.$$

By considering the example of Exercise 1.10 or otherwise, show that this limit may not always exist, even if f is differentiable at p.

Exercise 1.13. Suppose that $\Omega \subset \mathbb{R}^n$ is open, and $f, g : \Omega \to \mathbb{R}^m$ are differentiable at $p \in \Omega$. Show that $h = f + g$ is differentiable at p and

$$Dh(p) = Df(p) + Dg(p)$$

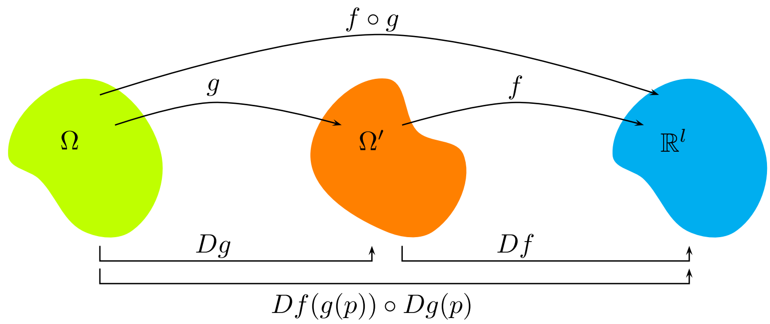

Figure 1.3: Illustration of Theorem 1.8.

1.3.2 Chain rule

In dimension one there is a simple "algorithm" which allows us to calculate the derivative of more complicated maps using the derivative of simpler ones. That algorithm is the chain rule. If $f, g : \mathbb{R} \to \mathbb{R}$, with g differentiable at p and f differentiable at $g(p)$, then $f \circ g$ is differentiable at p with

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

Now, suppose that $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^l$, with g differentiable at p and f differentiable at $g(p)$. Let $h = f \circ g$. We know that $Dg(p) : \mathbb{R}^n \to \mathbb{R}^m$ and $Df(g(p)) : \mathbb{R}^m \to \mathbb{R}^l$ are linear maps, so it certainly makes sense to consider $Df(g(p)) \circ Dg(p)$, where "$\circ$" denotes the composition of linear maps (corresponding to matrix multiplication). This will be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^l$, which is the right kind of object to be $Dh(p)$. In fact, it is the case that $h = f \circ g$ is differentiable at p with

$$Dh(p) = Df(g(p)) \circ Dg(p)$$

Theorem 1.8. *Assume $\Omega \subseteq \mathbb{R}^n$ and $\Omega' \subseteq \mathbb{R}^m$ are open sets, with $g : \Omega \to \Omega'$ differentiable at $p \in \Omega$ and $f : \Omega' \to \mathbb{R}^l$ differentiable at $g(p) \in \Omega'$. Then $h = f \circ g : \Omega \to \mathbb{R}^l$ is differentiable at p with derivative*

$$Dh(p) = Df(g(p)) \circ Dg(p).$$

(*) *Proof.* Let $g(p) = q$, $A = Dg(p)$, $B = Df(q)$. We define the map

$$\begin{aligned}
\phi(x) &= g(x) - g(p) - A(x-p), \quad \forall x \in \Omega \\
\psi(y) &= f(y) - f(q) - B(y-q), \quad \forall y \in \Omega' \\
\tau(x) &= f(g(x)) - f(g(p)) - B\left(A(x-p)\right), \quad \forall x \in \Omega.
\end{aligned}$$

By the assumptions in the theorem we know that

$$0 = \lim_{x\to p} \frac{\phi(x)}{\|x-p\|}, \tag{1.7}$$

$$0 = \lim_{y\to q} \frac{\psi(y)}{\|y-q\|}, \tag{1.8}$$

and we need to show that

$$\lim_{x\to p} \frac{\tau(x)}{\|x-p\|} = 0.$$

We may rewrite the map τ as

$$\begin{aligned}
\tau(x) &= f(g(x)) - f(g(p)) - B\left(A(x-p)\right) \\
&= f(g(x)) - f(g(p)) - B\left(g(x) - g(p) - \phi(x)\right) \\
&= f(g(x)) - f(g(p)) - B\left(g(x) - g(p)\right) + B(\phi(x)) \\
&= \psi(g(x)) + B(\phi(x)).
\end{aligned}$$

On the other hand, we recall from Example 1.4 that there is a real number M such that

$$\|A(x)\| \le M\|x\|, \quad \forall x \in \mathbb{R}^n.$$

Since B is linear, and hence continuous by Example 1.4, we have that

$$\lim_{x\to p} \frac{B(\phi(x))}{\|x-p\|} = \lim_{x\to p} B\left(\frac{\phi(x)}{\|x-p\|}\right) = B\left(\lim_{x\to p} \frac{\phi(x)}{\|x-p\|}\right) = 0.$$

Fix an arbitrary $\epsilon > 0$. It follows from (1.8) that there exists $\delta > 0$ such that for $y \in \Omega'$ with $\|y-q\| < \delta$ we have

$$\frac{\|\psi(y)\|}{\|y-q\|} < \epsilon$$

which implies

$$\|\psi(y)\| < \epsilon\|y-q\|.$$

On the other hand, since g is continuous, there exists δ_1 such that if $x \in \Omega$ with $\|x-p\| < \delta_1$ then

$$\|g(x) - g(p)\| = \|g(x) - q\| < \delta.$$

Thus, for every $x \in \Omega$ with $\|x-p\| < \delta_1$, we have

$$\begin{aligned}
\|\psi(g(x))\| &< \epsilon\|g(x) - q\| \\
&= \epsilon\|\phi(x) + A(x-p)\| \\
&\le \epsilon\|\phi(x)\| + \epsilon M\|x-p\|.
\end{aligned}$$

Dividing through by $\|x-p\|$ and taking the limit, we deduce that

$$\lim_{x\to p} \frac{\|\psi(g(x))\|}{\|x-p\|} \le \epsilon M.$$

Since $\epsilon > 0$ was arbitrary, we conclude

$$\lim_{x\to p} \frac{\|\psi(g(x))\|}{\|x-p\|} = 0,$$

and we are done. $\square$

Example 1.10. Let $m \geq 1$ be an integer, and assume that for $i = 1, 2, \dots, m$, the functions $g^i : (a, b) \to \mathbb{R}$ are differentiable at some $p \in (a, b)$. Then, the map $k : (a, b) \to \mathbb{R}$, defined as

$$k(x) = \left\| \left(g^1(x), g^2(x), \dots, g^m(x)\right) \right\|^2$$

is differentiable at p, and its Jacobian matrix has one real entry

$$2g^1(p)(g^1)'(p) + 2g^2(p)(g^2)'(p) \cdots + 2g^m(p)(g^m)'(p).$$

We note that by Example 1.9, the map $g : (a, b) \to \mathbb{R}^m$ defined as

$$g(x) = (g^1(p), g^2(p), \dots, g^m(p))$$

is differentiable at p with derivative

$$Dg(p) = \begin{pmatrix} (g^1)'(p) \\ \vdots \\ (g^m)'(p) \end{pmatrix}.$$

On the other hand, in Example 1.8, we saw that the map $f(x) = \|x\|^2$ is differentiable at every point in $\mathbb{R}^m$ with derivative $Df(q)[h] = 2\langle q, h\rangle$. We have $k = f \circ g$ on (a, b). Thus, by the chain rule, the map h is differentiable at p, with derivative

$$\begin{aligned}
Dk(p)[h] &= Df(g(p)) \circ Dg(p)[h] \\
&= D(f(g(P)) \left[\left((g^1)'(p)h, \dots, (g^m)'(p)h\right) \right] \\
&= 2 \left\langle g(p), ((g^1)'(p)h, \dots, (g^m)'(p)h) \right\rangle \\
&= 2 \left\langle g(p), h((g^1)'(p), \dots, (g^m)'(p)) \right\rangle \\
&= 2 \langle g(p), Dg(p) \rangle h.
\end{aligned}$$

Thus, the Jacobian of k at p is the one by one matrix with real entry

$$2\langle g(p), Dg(p)\rangle = 2g^1(p)(g^1)'(p) + 2g^2(p)(g^2)'(p) \cdots + 2g^m(p)(g^m)'(p).$$

Exercise 1.14. Assume Ω and Ω' are open sets in $\mathbb{R}^n$, $g : \Omega \to \Omega'$ differentiable at $p \in \Omega$ and $f : \Omega' \to \Omega$ differentiable at $g(p) \in \Omega'$. Moreover,

$$\begin{aligned}
f \circ g(x) &= x, \qquad \forall\, x \in \Omega. \\
g \circ f(x) &= x, \qquad \forall\, x \in \Omega'.
\end{aligned}$$

Show that

$$Df(g(p)) = (Dg(p))^{-1}.$$

Exercise 1.15 (*)**.** (a) Show that the map $P : \mathbb{R}^2 \to \mathbb{R}$ given by

$$P(x, y) = xy$$

is differentiable at each point $p = (\xi, \eta) \in \mathbb{R}^2$, with Jacobian

$$DP(p) = (\eta \quad \xi)\,.$$

(b) Suppose that $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $q \in \mathbb{R}^n$. Show that the map $Q : \mathbb{R}^n \to \mathbb{R}^2$ defined as

$$Q(z) = (f(z), g(z))$$

is differentiable at q, with derivative

$$DQ(q) = \begin{pmatrix} Df(q) \\ Dg(q) \end{pmatrix}$$

(c) Show that the map $F : \mathbb{R}^n \to \mathbb{R}$ defined as $F(z) = f(z)g(z)$, for all $z \in \mathbb{R}^n$, is differentiable at q, with derivative

$$DF(q) = g(q)Df(q) + f(q)Dg(q)$$

1.4 Directional derivatives

1.4.1 Rates of change and partial derivatives

Although the definitions of differentiability in dimension one and in higher dimensions appear similar, there is a major difference which makes the latter a more difficult concept. In dimension one, to see if a map : $f(a, b) \to \mathbb{R}$ is differentiable at some $x \in (a, b)$, we only need to verify that the limit of $(f(x) - f(p)/(x - p))$ exists as $x \to p$. To verify this, we do not need to know the value of the limit beforehand, that is, the value of the limit does not appear in this ratio. However, in higher dimensions, to verify if a map $f : \Omega \to \mathbb{R}^n$ is differentiable at some $p \in \Omega$, we need to know the derivative at that point. In other words, the derivative of the map at p appears in the criteria for differentiability. For basic maps, it is possible to guess the derivative, but in general, it may not be obvious what the derivative is. See for instance the map in Example 1.8. The purpose of this section is to present a simple approach to identify a candidate for the derivative in higher dimensions.

For a function $f : (a, b) \to \mathbb{R}$, we are familiar with the idea of $f'(p)$ telling us something about the rate of change of $f(x)$ as we vary x near $p \in (a, b)$. We can connect the derivative to this sort of concept with the directional derivative. Let us suppose that we are given a function $f : \mathbb{R}^3 \to \mathbb{R}$, which is supposed to represent the temperature of some three dimensional body which is not changing in time. Suppose we start at the origin $0 \in \mathbb{R}^3$ and travel along the curve $t \mapsto vt$, for some fixed $v \in \mathbb{R}^3$, that is we move along a straight line with velocity v passing through the origin at time 0. We can record the temperature of our surroundings as a function of time, $\theta(t)$ and we will find $\theta(t) = f(vt)$. Suppose we ask what the rate of change of temperature is at $t = 0$. This will of course be $\theta'(0)$. Now, we notice that we can write:

$$\theta = f \circ V$$

where V is the linear map $V : \mathbb{R} \to \mathbb{R}^3$ given by $V(t) = vt$. Now, we can use the chain rule to calculate $\theta'(0) = D\theta(0)$ and we find:

$$\theta'(0) = D\theta(0) = Df(0) \circ DV(0).$$

Now, since V is a linear map, we have $DV(0) = v$ and we conclude:

$$\theta'(0) = Df(0)[v].$$

This gives us a nice interpretation of the derivative $Df(0)$. When we apply $Df(0)$ to a vector v, we find the rate of change of f at 0 as we travel along a line with velocity v. More generally, we can consider travelling along the line given by $V(t) = p + tv$ for some $p \in \mathbb{R}^3$. Then at $t = 0$, we are passing through the point $p \in \mathbb{R}^3$. Setting $\theta(t) = f(p + tv)$, We call the quantity:

$$\theta'(0) = D\theta(p) = Df(p)[v]$$

the **directional derivative** of f at p in the direction v. Sometimes the notation

$$\frac{\partial f}{\partial v}(p) := \lim_{t \to 0} \frac{1}{t} \left[f(p + vt) - f(p) \right] = Df(p)[v]$$

is used for the directional derivative.

Now, if we take $\{e_1, e_2, e_3\}$ to be the canonical basis vectors for $\mathbb{R}^3$, then we can write $v = v^1 e_1 + v^2 e_2 + v^3 e_3$ for $v^i \in \mathbb{R}$. Doing this, and recalling that $Df(p)$ is a linear map, we have:

$$\begin{aligned} \frac{\partial f}{\partial v}(p) &= Df(p) \left[v^1 e_1 + v^2 e_2 + v^3 e_3 \right] && (1.9) \\ &= v^1 Df(p) [e_1] + v^2 Df(p) [e_2] + v^3 Df(p) [e_3] \\ &= v^1 D_1 f(p) + v^2 D_2 f(p) + v^3 D_3 f(p). && (1.10) \end{aligned}$$

In other words, we can find any directional derivative at p, provided we know the three numbers:

$$D_i f(p) = \frac{\partial f}{\partial e_i}(p), \qquad i = 1, 2, 3.$$

called the **partial derivatives** of f at p. Equivalently, these can be defined as

$$D_i f(p) := \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t}.$$

If $f : \mathbb{R}^3 \to \mathbb{R}$, then for x, y and z in $\mathbb{R}$,

$$D_1 f(x, y, z) = \lim_{t \to 0} \frac{f(x+t, y, z) - f(x, y, z)}{t} =: \frac{\partial f}{\partial x}(x, y, z),$$

where we've introduced yet more notation. The expression $\frac{\partial f}{\partial x}$ you should think of as meaning 'differentiate f with respect to x, while treating y, z as constants. Returning to (1.10), we see that for any $v = (v^1, v^2, v^3)$, we have

$$Df(p)[v] = \begin{pmatrix} D_1 f(p) & D_2 f(p) & D_3 f(p) \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix},$$

so that the Jacobian of f at p is given by

$$Df(p) = \begin{pmatrix} D_1 f(p) & D_2 f(p) & D_3 f(p) \end{pmatrix}.$$

To introduce even more notation, we sometimes write

$$\nabla f(p) = \begin{pmatrix} D_1 f(p) \\ D_2 f(p) \\ D_3 f(p) \end{pmatrix},$$

which is called the **gradient** of f at p, and with this notation

$$Df(p) = (\nabla f(p))^t.$$

We can extend all of these notions to more general range and domains, which leads us to the following definition.

Definition 1.6. Suppose $\Omega \subset \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}^m$ is differentiable at $p \in \Omega$. For any vector $v \in \mathbb{R}^n$ with $\|v\| = 1$, the **directional derivative** of f at p in the direction v is given by

$$\frac{\partial f}{\partial v}(p) = \lim_{t \to 0} \frac{f(p + tv) - f(p)}{t} = Df(p)[v]$$

The partial derivatives of f at p are given by

$$D_i f(p) = \frac{\partial f}{\partial e_i}(p) = \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t}, \quad i = 1, \ldots, n.$$

Notice that $f(x)$ is now a vector in $\mathbb{R}^m$, so expressions like $\lim_{t \to 0} \frac{f(p+tv)-f(p)}{t}$ have to be understood as limits in $\mathbb{R}^m$, so that $\frac{\partial f}{\partial v}(p)$ will be an $m-$dimensional column vector. That is, if

$$f(x) = (f^1(x), f^2(x), \ldots, f^m(x)),$$

then

$$D_i f(p) = \begin{pmatrix} D_i f^1(p) \\ \vdots \\ D_i f^m(p) \end{pmatrix}.$$

Theorem 1.9. *Suppose $\Omega \subset \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}^m$ is of the form*

$$f(x) = (f^1(x), f^2(x), \ldots, f^m(x)).$$

If f is differentiable at some $p \in \Omega$, then the Jacobian of f at p is

$$Df(p) = \begin{pmatrix} D_1 f^1(p) & \ldots & D_n f^1(p) \\ \vdots & \ddots & \vdots \\ D_1 f^m(p) & \ldots & D_n f^m(p) \end{pmatrix}.$$

Proof. Let $\{e_i\}$ be the canonical basis for $\mathbb{R}^n$. For any $v \in \mathbb{R}^n$, we write $v = \sum_{i=1}^n v^i e_i$. Then by the linearity of $Df(p)$ we have:

$$\begin{aligned} Df(p)[v] = Df(p)\left[\sum_{i=1}^n v^i e_i\right] &= \sum_{i=1}^n v^i Df(p)\,[e_i] = \sum_{i=1}^n v^i D_i f(p). \\ &= \begin{pmatrix} \sum_{i=1}^n v^i D_i f^1(p) \\ \vdots \\ \sum_{i=1}^n v^i D_i f^m(p) \end{pmatrix} \\ &= \begin{pmatrix} D_1 f^1(p) & \ldots & D_n f^1(p) \\ \vdots & \ddots & \vdots \\ D_1 f^m(p) & \ldots & D_n f^m(p) \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \end{aligned}$$

□

This allows us to restate the chain rule in terms of the partial derivatives of the functions.

Corollary 1.10. *Suppose $\Omega \subset \mathbb{R}^n$ and $\Omega' \subset \mathbb{R}^m$ are open sets, $g : \Omega \to \Omega'$ is differentiable at $p \in \Omega$, and $f : \Omega' \to \mathbb{R}^l$ is differentiable at $g(p)$. Then $h = f \circ g$ is differentiable at p with Jacobian*

$$Dh(p) = \begin{pmatrix} D_1 f^1(g(p)) & \ldots & D_m f^1(g(p)) \\ \vdots & \ddots & \vdots \\ D_1 f^l(g(p)) & \ldots & D_m f^l(g(p)) \end{pmatrix} \begin{pmatrix} D_1 g^1(p) & \ldots & D_n g^1(p) \\ \vdots & \ddots & \vdots \\ D_1 g^m(p) & \ldots & D_n g^m(p) \end{pmatrix}$$

In the one dimensional case, we often use the derivative to search for turning points, i.e. maxima and minima, since a differentiable function will have vanishing derivative at a local maximum or minimum. A similar result holds in the higher dimensional case.

Lemma 1.11. *Let $\Omega \subset \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$ be differentiable at each point in Ω. Suppose that f has a local maximum at $p \in \Omega$. Then:*

$$Df(p) = 0.$$

Similarly if p is a local minimum.

Proof. Pick $v \in \mathbb{R}^n$. Since Ω is open, there exists $\epsilon > 0$ such that $p + tv \in \Omega$ for $t \in (-\epsilon, \epsilon)$. Consider the function $g_v : (-\epsilon, \epsilon) \to \mathbb{R}$ defined as

$$g_v(t) = f(p + tv).$$

Since f has a local maximum at p, g_v has a local maximum at 0 and moreover, g_v is differentiable by the chain rule, so we deduce

$$0 = g_v'(0) = Df(p)[v].$$

Since v was arbitrary, we have that $Df(p) = 0$. A similar argument deals with the case where p is a minimum. □

Exercise 1.16. (i) Let the function $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$f(x, y) = (x^2 + e^{x+y}, x - \log y, 2xy + 1).$$

Assuming f is differentiable at a point (x, y), what is its derivative?

(ii) Let $g : \mathbb{R}^3 \to \mathbb{R}^1$ be given by

$$g(x, y, z) = x + y + z.$$

Compute the derivative of $g \circ f$ assuming it exists. Compute it in 2 ways, with and without the chain rule.

1.4.2 Relation between partial derivatives and differentiability

We have seen above that for a function $f : \mathbb{R}^n \to \mathbb{R}$ which is differentiable at some point p, the limits

$$D_i f(p) := \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t} \tag{1.11}$$

exist for $i = 1, \ldots n$, and moreover these limits completely determine the derivative of f at p. One might hope, based on this, that in order for f to be differentiable at p it is enough to know that the partial derivatives (i.e. the limits in (1.11)) of f at p all exist. Unfortunately, this is not the case, as we show in the following example.

Example 1.11. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$f(x, y) = \begin{cases} 0 & x = y = 0 \\ \frac{xy}{\sqrt{x^2+y^2}} & \text{otherwise} \end{cases}$$

See Figure 1.4 for the graph of the function f.

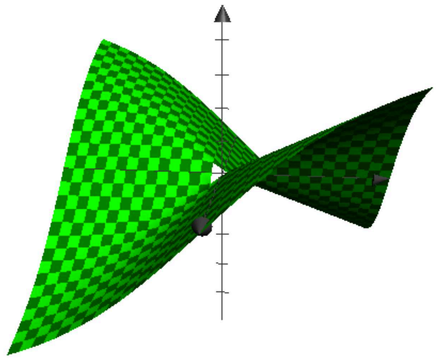

Figure 1.4: The graph of the function in Example 1.11.

First note that this function is continuous at the origin. Since $|xy| \leq \frac{1}{2}\left(x^2+y^2\right)$, we have that for $p = (x,y) \neq (0,0)$:

$$|f(p)| \leq \frac{1}{2}\sqrt{x^2+y^2},$$

so that

$$\lim_{p\to 0} f(p) = 0.$$

Now consider the partial derivatives. We have

$$D_1 f(0) = \lim_{t\to 0} \frac{1}{t}\left[f(te_1) - f(0)\right] = \lim_{t\to 0} \frac{0-0}{t} = 0$$

since $f(te_1) = 0$ for all t. Similarly, we also have

$$D_2 f(0) = \lim_{t\to 0} \frac{1}{t}\left[f(te_2) - f(0)\right] = \lim_{t\to 0} \frac{0-0}{t} = 0$$

Thus, if f is differentiable, then it must be that $Df = 0$, so all directional derivatives at 0 exist and are equal to zero. However, let $h = \frac{1}{\sqrt{2}}(1,1)$. For $t > 0$, we have

$$\frac{f(th) - f(0)}{t} = \frac{t^2/2}{t^2} = \frac{1}{2},$$

which contradicts the differentiability of f at the origin. Thus, even though the partial derivatives exist for this function, the function is not differentiable.

Away from the origin, the function is a composition of smooth functions so is differentiable. We can calculate the partial derivatives at a point $p = (x,y) \neq (0,0)$ and we find

$$D_1 f(p) = \frac{y}{\sqrt{x^2+y^2}} - \frac{x^2 y}{(x^2+y^2)^{\frac{3}{2}}} = \frac{y^3}{(x^2+y^2)^{\frac{3}{2}}},$$

and by symmetry:

$$D_2 f(p) = \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}}.$$

We claim that the function $g : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ given by

$$g(x,y) = \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}}$$

has no limit as $p = (x,y)$ converges to $(0,0)$. To see this, let $p = (r\cos\theta, r\sin\theta)$ for some $r \in (0,\infty)$, $\theta \in [0, 2\pi)$. Then

$$g(p) = \cos^3\theta,$$

so there can be no limit as $r \to 0$, since g approaches a different value depending on which angle we approach from.

As it happens, the fact that the partial derivatives are not continuous in a neighbourhood of the origin is the only barrier to differentiability there.

Theorem 1.12. *Let $\Omega \subset \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$. Suppose the partial derivatives*

$$D_i f(x) := \lim_{t\to 0} \frac{f(x+te_i) - f(x)}{t}$$

exist for all $x \in \Omega$, and moreover suppose that the maps

$$x \mapsto D_i f(x)$$

are continuous at $p \in \Omega$ for all $i = 1, \ldots, n$. Then f is differentiable at p.

(*) *Proof.* Since Ω is open, there exists $r > 0$ such that $B_r(p) \subset \Omega$. Suppose $h \in B_r(0)$ has components h^i, so that $h = \sum_{i=1}^n h^i e_i$. We consider

$$\begin{aligned}
f(p+h) - f(p) &= f\left(p + \sum_{i=1}^n h^i e_i\right) - f(p) \\
&= f\left(p + \sum_{i=1}^n h^i e_i\right) - f\left(p + \sum_{i=1}^{n-1} h^i e_i\right) \\
&\quad + f\left(p + \sum_{i=1}^{n-1} h^i e_i\right) - f\left(p + \sum_{i=1}^{n-2} h^i e_i\right) \\
&\quad + \ldots \\
&\quad + f(p + h^1 e_1) - f(p).
\end{aligned}$$

Let's consider a typical line in the right hand side of the above equation, that is,

$$f\left(p + \sum_{i=1}^k h^i e_i\right) - f\left(p + \sum_{i=1}^{k-1} h^i e_i\right) = f(q + h^k e_k) - f(q),$$

where $k \in \{1, \dots, n\}$ and $q = p + \sum_{i=1}^{k-1} h^i e_i$. Now, applying the mean value theorem to the function $g(t) = f(q + te_k)$, which is differentiable by assumption, there exists $s \in \left[-\left|h^k\right|, \left|h^k\right|\right)$ such that:

$$f(q + h^k e_k) - f(q) = h^k D_k f(q + se_k) = h^k D_k f(p + c_k),$$

where $c_k = \sum_{i=1}^{k-1} h^i e_i + se_k$. One has to consider separately the cases $h^k > 0$, $h^k < 0$ and $h^k = 0$. Now, note that since $|s| \leq \left|h^k\right|$, we have

$$\|c_k\| \leq \|h\|.$$

Putting this together, we conclude that there exists $c_1, \dots, c_n \in \mathbb{R}^n$ with $\|c_k\| \leq \|h\|$ such that

$$f(p + h) - f(p) = \sum_{k=1}^{n} h^k D_k f(p + c_k).$$

From here we can estimate using the Cauchy-Schwartz identity

$$\begin{aligned} \left| f(p + h) - f(p) - \sum_{k=1}^{n} h^k D_k f(p) \right| &\leq \sum_{k=1}^{n} h^k \left| D_k f(p + c_k) - D_k f(p) \right| \\ &\leq \|h\| \left(\sum_{k=1}^{n} \left| D_k f(p + c_k) - D_k f(p) \right|^2 \right)^{\frac{1}{2}}, \end{aligned}$$

so that

$$\frac{\left| f(p + h) - f(p) - \sum_{k=1}^{n} h^k D_k f(p) \right|}{\|h\|} \leq \left(\sum_{k=1}^{n} \left| D_k f(p + c_k) - D_k f(p) \right|^2 \right)^{\frac{1}{2}}.$$

Now, fix $\epsilon > 0$. Since $x \mapsto D_k f(x)$ is continuous at p, for each $k = 1, \dots, n$, there exists δ_k such that if $\|c\| < \delta_k$ we have:

$$|D_k f(p + c) - D_k f(p)| < \frac{\epsilon}{\sqrt{n}}.$$

Suppose $\|h\| < \min\{\delta_1, \dots, \delta_n\} =: \delta$. Then as $\|c_k\| \leq \|h\|$, we deduce

$$\frac{\left| f(p + h) - f(p) - \sum_{k=1}^{n} h^k D_k f(p) \right|}{\|h\|} < \left(\sum_{k=1}^{n} \frac{\epsilon^2}{n} \right)^{\frac{1}{2}} = \epsilon.$$

As ϵ was arbitrary, we conclude that f is differentiable at p, with derivative

$$Df(p)[h] = \sum_{k=1}^{n} D_k f(p) h^k. \qquad \square$$

Exercise 1.17. Show that each of the following maps $f : \mathbb{R}^2 \to \mathbb{R}$ is everywhere differentiable

(a) $f(x,y) = x^2 + y^2 - x - xy$,

(b) $f(x,y) = \frac{1}{\sqrt{1+x^2+y^2}}$,

(c) $f(x,y) = x^5y^2$.

For maps $f : (a,b) \to \mathbb{R}$ we have learned that when f is differentiable at some $p \in (a,b)$, then there is a tangent line to the graph of f that passes through $(p, f(p))$ and approximates the graph of f near p. This is an intuitive picture that is only valid when we consider the graph of a function from one dimension to one dimension. By an example below, we show that this intuition should not be employed for maps of higher dimensions.

Example 1.12. Let $f : (-1,+1) \to \mathbb{R}^2$ be define as

$$f(x) = \begin{cases} (x^2, 0) & \text{if } x \geq 0 \\ (0, x^2) & \text{if } x < 0. \end{cases}$$

See Figure 1.12 for the image of the map f.

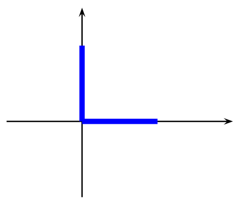

Figure 1.5: The image of the map f in Example 1.12

Clearly, f is continuous at 0 with

$$\lim_{x\to 0} f(x) = (0,0) = f(0).$$

The map f is differentiable at 0 with derivative equal to the constant linear map $\Lambda = 0$. To see this, note that

$$\lim_{h\to 0} \frac{\|f(0+h) - f(0) - \Lambda[h]\|}{\|h\|} = \lim_{h\to 0} \frac{\|f(h)\|}{\|h\|} = \lim_{h\to 0} \frac{h^2}{|h|} = \lim_{h\to 0} |h| = 0.$$

In fact, it is not possible to understand just by looking at the image of a map whether it is differentiable or not. As the example below shows, maps with the same image may or may not be differentiable.

Example 1.13. Define the maps k and g from $(-1,+1)$ to $\mathbb{R}^2$ as

$$k(x) = (x, x^3), \quad g(x) = (x^{1/3}, x).$$

See Figure 1.6 for the images of the maps f and g.

The maps k and g are continuous at 0 with

$$\lim_{x\to 0} k(x) = (0,0) = k(0),$$

and

$$\lim_{x\to 0} g(x) = (0,0) = g(0).$$

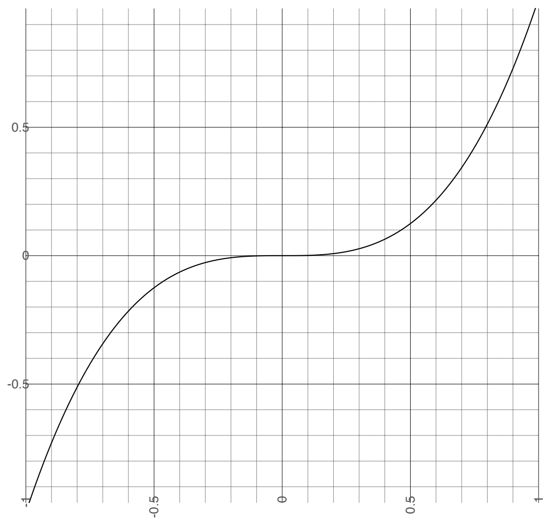

Figure 1.6: The image of the maps f and g in Example 1.13. The differentiability at $(0,0)$ depends on "how fast" we pass through the point $(0,0)$.

The maps k and g have the same image, that is, they map the interval $(-1,+1)$ to the same curve, which is the graph of the function $t \mapsto t^3$ on the interval $(-1,+1)$. However, k is differentiable at 0, but g is not differentiable at 0, as we show below.

We claim that the derivative of the map k at 0 is equal to the linear map $\Lambda(h) = (h,0)$. To see this, note that

$$\begin{aligned}\lim_{h\to 0} \frac{\|k(0+h) - k(0) - \Lambda[h]\|}{\|h\|} &= \lim_{h\to 0} \frac{\big\|(h,h^3) - (h,0)\big\|}{\|h\|}\\ &= \lim_{h\to 0} \frac{\big\|(0,h^3)\big\|}{\|h\|}\\ &= \lim_{h\to 0} \frac{|h|^3}{|h|} = 0.\end{aligned}$$

To prove that g is not differentiable at 0, we need to show that there is no linear map $\Lambda : \mathbb{R} \to \mathbb{R}^2$ which is the derivative of the map g at 0. In contrary assume that there is a linear map $\Lambda : \mathbb{R} \to \mathbb{R}^2$ such that

$$\lim_{h\to 0} \frac{\|g(0+h) - g(0) - \Lambda[h]\|}{\|h\|} = 0.$$

Let $\Lambda(1) = (a, b) \in \mathbb{R}^2$, for some real constants a and b in $\mathbb{R}$. It follows that for every $h \in \mathbb{R}$ we have

$$\Lambda(h) = \Lambda(h \cdot 1) = h\Lambda(1) = h(a, b) = (ha, hb).$$

Therefore,

$$\begin{aligned} 0 = \lim_{h\to 0} \frac{\|g(0+h) - g(0) - \Lambda[h]\|}{\|h\|} &= \lim_{h\to 0} \frac{\left\|(h^{1/3} - ah, h - bh)\right\|}{|h|} \\ &= \lim_{h\to 0} \frac{\left\|h(h^{-2/3} - a, 1 - b)\right\|}{|h|} \\ &= \lim_{h\to 0} \left\|(h^{-2/3} - a, 1 - b)\right\| \\ &= \left\|\lim_{h\to 0}(h^{-2/3} - a, 1 - b)\right\| \end{aligned}$$

In the last line of the above equation we have used that $\|\cdot\|$ is a continuous function, so we may interchange the limit and the norm. Now recall that $\|y\| = 0$, if and only if $y = 0$. Thus we must have

$$\lim_{h\to 0}(h^{-2/3} - a, 1 - b) = (0, 0)$$

which implies that

$$\lim_{h\to 0} h^{-2/3} - a = 0, \quad \text{and} \lim_{h\to 0} 1 - b = 0.$$

This is a contradiction, since for any real number a we have

$$\lim_{h\to 0} h^{-2/3} - a = \infty.$$

This contradiction shows that there is no linear map $\Lambda : \mathbb{R} \to \mathbb{R}^2$ satisfying the definition of differentiability for g at 0.

Note that the value of the other limit does not lead to any contradiction, it only says that b must be equal to 1.