Chapter 8

Several variables and partial derivatives

8.1 Vector spaces, linear mappings, and convexity

Note: 2–3 lectures

8.1.1 Vector spaces

The euclidean space $\mathbb{R}^n$ has already made an appearance in the metric space chapter. In this chapter, we will extend the differential calculus we created for one variable to several variables. The key idea in differential calculus is to approximate functions by lines and linear functions. In several variables we must introduce a little bit of linear algebra before we can move on. So let us start with vector spaces and linear functions on vector spaces.

While it is common to use $\vec{x}$ or the bold $\mathbf{x}$ for elements of $\mathbb{R}^n$, especially in the applied sciences, we use just plain x, which is common in mathematics. That is, $v \in \mathbb{R}^n$ is a *vector*, which means $v = (v_1, v_2, \ldots, v_n)$ is an n-tuple of real numbers.*

It is common to write vectors as *column vectors*, that is, $n \times 1$ matrices:

$$v = (v_1, v_2, \ldots, v_n) = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

We will do so when convenient. We call real numbers *scalars* to distinguish them from vectors.

The set $\mathbb{R}^n$ has a so-called *vector space* structure defined on it. However, even though we will be looking at functions defined on $\mathbb{R}^n$, not all spaces we wish to deal with are equal to $\mathbb{R}^n$. Therefore, let us define the abstract notion of the vector space.

*Subscripts are used for many purposes, so sometimes we may have several vectors which may also be identified by subscript such as a finite or infinite sequence of vectors $y_1, y_2, \ldots$.

Definition 8.1.1. Let X be a set together with operations of addition, $+\colon X \times X \to X$, and multiplication, $\cdot\colon \mathbb{R} \times X \to X$, (we usually write ax instead of $a \cdot x$). X is called a *vector space* (or a *real vector space*) if the following conditions are satisfied:

(i) (Addition is associative) If $u, v, w \in X$, then $u + (v + w) = (u + v) + w$.

(ii) (Addition is commutative) If $u, v \in X$, then $u + v = v + u$.

(iii) (Additive identity) There is a $0 \in X$ such that $v + 0 = v$ for all $v \in X$.

(iv) (Additive inverse) For every $v \in X$, there is a $-v \in X$, such that $v + (-v) = 0$.

(v) (Distributive law) If $a \in \mathbb{R}$, $u, v \in X$, then $a(u + v) = au + av$.

(vi) (Distributive law) If $a, b \in \mathbb{R}$, $v \in X$, then $(a + b)v = av + bv$.

(vii) (Multiplication is associative) If $a, b \in \mathbb{R}$, $v \in X$, then $(ab)v = a(bv)$.

(viii) (Multiplicative identity) $1v = v$ for all $v \in X$.

Elements of a vector space are usually called *vectors*, even if they are not elements of $\mathbb{R}^n$ (vectors in the "traditional" sense).

If $Y \subset X$ is a subset that is a vector space itself with the same operations, then Y is called a *subspace* or *vector subspace* of X.

Example 8.1.2: An example vector space is $\mathbb{R}^n$, where addition and multiplication by a constant is done componentwise: if $a \in \mathbb{R}$, $v = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$, and $w = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$, then

$$\begin{aligned} v + w &:= (v_1, v_2, \ldots, v_n) + (w_1, w_2, \ldots, w_n) = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n), \\ av &:= a(v_1, v_2, \ldots, v_n) = (av_1, av_2, \ldots, av_n). \end{aligned}$$

In this book we mostly deal with vector spaces that can be often regarded as subsets of $\mathbb{R}^n$, but there are other vector spaces useful in analysis. Let us give a couple of examples.

Example 8.1.3: A trivial example of a vector space (the smallest one in fact) is just $X = \{0\}$. The operations are defined in the obvious way. You always need a zero vector to exist, so all vector spaces are nonempty sets.

Example 8.1.4: The space $C([0,1], \mathbb{R})$ of continuous functions on the interval $[0,1]$ is a vector space. For two functions f and g in $C([0,1], \mathbb{R})$ and $a \in \mathbb{R}$ we make the obvious definitions of $f + g$ and af:

$$(f + g)(x) := f(x) + g(x), \qquad (af)(x) := a\big(f(x)\big).$$

The 0 is the function that is identically zero. We leave it as an exercise to check that all the vector space conditions are satisfied.

Example 8.1.5: The space of polynomials $c_0 + c_1 t + c_2 t^2 + \cdots + c_m t^m$ is a vector space, let us denote it by $\mathbb{R}[t]$ (coefficients are real and the variable is t). The operations are defined in the same

way as for functions above. Suppose we have two polynomials, one of degree m and one of degree n. Assume $n \geq m$ for simplicity. Then

$$(c_0 + c_1 t + c_2 t^2 + \cdots + c_m t^m) + (d_0 + d_1 t + d_2 t^2 + \cdots + d_n t^n) = \\ (c_0 + d_0) + (c_1 + d_1)t + (c_2 + d_2)t^2 + \cdots + (c_m + d_m)t^m + d_{m+1}t^{m+1} + \cdots + d_n t^n$$

and

$$a(c_0 + c_1 t + c_2 t^2 + \cdots + c_m t^m) = (ac_0) + (ac_1)t + (ac_2)t^2 + \cdots + (ac_m)t^m.$$

Despite what it looks like, $\mathbb{R}[t]$ is not equivalent to $\mathbb{R}^n$ for any n. In particular it is not "finite dimensional", we will make this notion precise in just a little bit. One can make a finite dimensional vector subspace by restricting the degree. For example, if we say $\mathscr{P}_n$ is the space of polynomials of degree n or less, then we have a finite dimensional vector space.

The space $\mathbb{R}[t]$ can be thought of as a subspace of $C(\mathbb{R},\mathbb{R})$. If we restrict the range of t to $[0,1]$, $\mathbb{R}[t]$ can be identified with a subspace of $C([0,1],\mathbb{R})$.

It is often better to think of even simpler "finite dimensional" vector spaces using the abstract notion rather than always $\mathbb{R}^n$. It is possible to use other fields than $\mathbb{R}$ in the definition (for example it is common to use the complex numbers $\mathbb{C}$), but let us stick with the real numbers*.

8.1.2 Linear combinations and dimension

Definition 8.1.6. Suppose X is a vector space, $x_1, x_2, \ldots, x_k \in X$ are vectors, and $a_1, a_2, \ldots, a_k \in \mathbb{R}$ are scalars. Then

$$a_1 x_1 + a_2 x_2 + \cdots + a_k x_k$$

is called a *linear combination* of the vectors $x_1, x_2, \ldots, x_k$.

If $Y \subset X$ is a set then the *span* of Y, or in notation $\operatorname{span}(Y)$, is the set of all linear combinations of some finite number of elements of Y. We also say Y *spans* $\operatorname{span}(Y)$.

Example 8.1.7: Let $Y := \{(1,1)\} \subset \mathbb{R}^2$. Then

$$\operatorname{span}(Y) = \{(x,x) \in \mathbb{R}^2 : x \in \mathbb{R}\}.$$

That is, $\operatorname{span}(Y)$ is the line through the origin and the point $(1,1)$.

Example 8.1.8: Let $Y := \{(1,1),(0,1)\} \subset \mathbb{R}^2$. Then

$$\operatorname{span}(Y) = \mathbb{R}^2,$$

as any point $(x,y) \in \mathbb{R}^2$ can be written as a linear combination

$$(x,y) = x(1,1) + (y-x)(0,1).$$

*If you want a very funky vector space over a different field, $\mathbb{R}$ itself is a vector space over the rational numbers.

A sum of two linear combinations is again a linear combination, and a scalar multiple of a linear combination is a linear combination, which proves the following proposition.

Proposition 8.1.9. *Let X be a vector space. For any $Y \subset X$, the set* $\operatorname{span}(Y)$ *is a vector space itself. That is,* $\operatorname{span}(Y)$ *is a subspace of X.*

If Y is already a vector space then $\operatorname{span}(Y) = Y$.

Definition 8.1.10. A set of vectors $\{x_1, x_2, \ldots, x_k\} \subset X$ is *linearly independent*, if the only solution to

$$a_1 x_1 + a_2 x_2 + \cdots + a_k x_k = 0 \tag{8.1}$$

is the trivial solution $a_1 = a_2 = \cdots = a_k = 0$. A set that is not linearly independent, is *linearly dependent*.

A linearly independent set B of vectors such that $\operatorname{span}(B) = X$ is called a *basis* of X. For example the set Y of the two vectors in Example 8.1.8 is a basis of $\mathbb{R}^2$.

If a vector space X contains a linearly independent set of d vectors, but no linearly independent set of $d+1$ vectors then we say the *dimension* or $\dim X := d$. If for all $d \in \mathbb{N}$ the vector space X contains a set of d linearly independent vectors, we say X is infinite dimensional and write $\dim X := \infty$.

Clearly for the trivial vector space, $\dim \{0\} = 0$. We will see in a moment that any vector space that is a subset of $\mathbb{R}^n$ has a finite dimension, and that dimension is less than or equal to n.

If a set is linearly dependent, then one of the vectors is a linear combination of the others. In other words, in (8.1) if $a_j \neq 0$, then we solve for x_j

$$x_j = \frac{a_1}{a_j} x_1 + \cdots + \frac{a_{j-1}}{a_j} x_{j-1} + \frac{a_{j+1}}{a_j} x_{j+1} + \cdots + \frac{a_k}{a_k} x_k.$$

Clearly then the vector x_j has at least two different representations as linear combinations of $\{x_1, x_2, \ldots, x_k\}$.

Proposition 8.1.11. *If $B = \{x_1, x_2, \ldots, x_k\}$ is a basis of a vector space X, then every point $y \in X$ has a unique representation of the form*

$$y = \sum_{j=1}^{k} a_j x_j$$

for some scalars $a_1, a_2, \ldots, a_k$.

Proof. Every $y \in X$ is a linear combination of elements of B since X is the span of B. For uniqueness suppose

$$y = \sum_{j=1}^{k} a_j x_j = \sum_{j=1}^{k} b_j x_j,$$

then

$$\sum_{j=1}^{k}(a_j - b_j)x_j = 0.$$

By linear independence of the basis $a_j = b_j$ for all j. □

For $\mathbb{R}^n$ we define

$$e_1 := (1,0,0,\ldots,0), \quad e_2 := (0,1,0,\ldots,0), \quad \ldots, \quad e_n := (0,0,0,\ldots,1),$$

and call this the *standard basis* of $\mathbb{R}^n$. We use the same letters e_j for any $\mathbb{R}^n$, and which space $\mathbb{R}^n$ we are working in is understood from context. A direct computation shows that $\{e_1, e_2, \ldots, e_n\}$ is really a basis of $\mathbb{R}^n$; it spans $\mathbb{R}^n$ and is linearly independent. In fact,

$$x = (x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} x_j e_j.$$

Proposition 8.1.12. *Let X be a vector space and d a nonnegative integer.*

(i) If X is spanned by d vectors, then $\dim X \leq d$.

(ii) $\dim X = d$ *if and only if X has a basis of d vectors (and so every basis has d vectors).*

(iii) In particular, $\dim \mathbb{R}^n = n$.

(iv) If $Y \subset X$ is a vector subspace and $\dim X = d$, *then* $\dim Y \leq d$.

(v) If $\dim X = d$ *and a set T of d vectors spans X, then T is linearly independent.*

(vi) If $\dim X = d$ *and a set T of m vectors is linearly independent, then there is a set S of $d-m$ vectors such that $T \cup S$ is a basis of X.*

Proof. Let us start with (i). Suppose $S = \{x_1, x_2, \ldots, x_d\}$ spans X, and $T = \{y_1, y_2, \ldots, y_m\}$ is a set of linearly independent vectors of X. We wish to show that $m \leq d$. Write

$$y_1 = \sum_{k=1}^{d} a_{k,1} x_k,$$

for some numbers $a_{1,1}, a_{2,1}, \ldots, a_{d,1}$, which we can do as S spans X. One of the $a_{k,1}$ is nonzero (otherwise y_1 would be zero), so suppose without loss of generality that this is $a_{1,1}$. Then we solve

$$x_1 = \frac{1}{a_{1,1}} y_1 - \sum_{k=2}^{d} \frac{a_{k,1}}{a_{1,1}} x_k.$$

In particular $\{y_1, x_2, \ldots, x_d\}$ span X, since x_1 can be obtained from $\{y_1, x_2, \ldots, x_d\}$. Therefore, there are some numbers for some numbers $a_{1,2}, a_{2,2}, \ldots, a_{d,2}$, such that

$$y_2 = a_{1,2} y_1 + \sum_{k=2}^{d} a_{k,2} x_k.$$

As T is linearly independent, we must have that one of the $a_{k,2}$ for $k \geq 2$ must be nonzero. Without loss of generality suppose $a_{2,2} \neq 0$. Proceed to solve for

$$x_2 = \frac{1}{a_{2,2}} y_2 - \frac{a_{1,2}}{a_{2,2}} y_1 - \sum_{k=3}^{d} \frac{a_{k,2}}{a_{2,2}} x_k.$$

In particular $\{y_1, y_2, x_3, \ldots, x_d\}$ spans X.

We continue this procedure. If $m < d$, then we are done. So suppose $m \geq d$. After d steps we obtain that $\{y_1, y_2, \ldots, y_d\}$ spans X. Any other vector v in X is a linear combination of $\{y_1, y_2, \ldots, y_d\}$, and hence cannot be in T as T is linearly independent. So $m = d$.

Let us look at (ii). First, if we have a set T of k linearly independent vectors that do not span X, then we can always choose a vector $v \in X \setminus \operatorname{span}(T)$. The set $T \cup \{v\}$ is linearly independent (exercise). If $\dim X = d$, then there must exist some linearly independent set of d vectors T, and it must span X, otherwise we could choose a larger set of linearly independent vectors. So we have a basis of d vectors. On the other hand if we have a basis of d vectors, it is linearly independent and spans X. By (i) we know there is no set of $d+1$ linearly independent vectors, so dimension must be d.

For (iii) notice that $\{e_1, e_2, \ldots, e_n\}$ is a basis of $\mathbb{R}^n$.

To see (iv), suppose Y is a vector space and $Y \subset X$, where $\dim X = d$. As X cannot contain $d+1$ linearly independent vectors, neither can Y.

For (v) suppose T is a set of m vectors that is linearly dependent and spans X. Then one of the vectors is a linear combination of the others. Therefore if we remove it from T we obtain a set of $m-1$ vectors that still span X and hence $\dim X \leq m-1$ by (i).

For (vi) suppose $T = \{x_1, x_2, \ldots, x_m\}$ is a linearly independent set. We follow the procedure above in the proof of (ii) to keep adding vectors while keeping the set linearly independent. As the dimension is d we can add a vector exactly $d-m$ times. □

8.1.3 Linear mappings

A function $f \colon X \to Y$, when Y is not $\mathbb{R}$, is often called a *mapping* or a *map* rather than a *function*.

Definition 8.1.13. A mapping $A \colon X \to Y$ of vector spaces X and Y is *linear* (or a *linear transformation*) if for every $a \in \mathbb{R}$ and $x, y \in X$ we have

$$A(ax) = aA(x), \qquad \text{and} \qquad A(x+y) = A(x) + A(y).$$

We usually write Ax instead of $A(x)$ if A is linear.

If A is one-to-one an onto then we say A is *invertible* and we denote the inverse by A^{-1}.

If $A \colon X \to X$ is linear then we say A is a *linear operator* on X.

We write $L(X,Y)$ for the set of all linear transformations from X to Y, and just $L(X)$ for the set of linear operators on X. If $a \in \mathbb{R}$ and $A, B \in L(X,Y)$, define the transformations aA and $A+B$ by

$$(aA)(x) := aAx, \qquad (A+B)(x) := Ax + Bx.$$

If $A \in L(Y,Z)$ and $B \in L(X,Y)$, define the transformation AB as the composition $A \circ B$, that is,

$$ABx := A(Bx).$$

Finally denote by $I \in L(X)$ the *identity*: the linear operator such that $Ix = x$ for all x.

It is not hard to see that $aA \in L(X,Y)$ and $A+B \in L(X,Y)$, and that $AB \in L(X,Z)$. In particular, $L(X,Y)$ is a vector space. As the set $L(X)$ is not only a vector space, but also admits a product, it is often called an *algebra*.

An immediate consequence of the definition of a linear mapping is that if A is linear then $A0 = 0$.

Proposition 8.1.14. *If $A \in L(X,Y)$ is invertible, then A^{-1} is linear.*

Proof. Let $a \in \mathbb{R}$ and $y \in Y$. As A is onto, then there is an x such that $y = Ax$, and further as it is also one-to-one $A^{-1}(Az) = z$ for all $z \in X$. So

$$A^{-1}(ay) = A^{-1}(aAx) = A^{-1}\big(A(ax)\big) = ax = aA^{-1}(y).$$

Similarly let $y_1, y_2 \in Y$, and $x_1, x_2 \in X$ such that $Ax_1 = y_1$ and $Ax_2 = y_2$, then

$$A^{-1}(y_1+y_2) = A^{-1}(Ax_1+Ax_2) = A^{-1}\big(A(x_1+x_2)\big) = x_1+x_2 = A^{-1}(y_1)+A^{-1}(y_2). \qquad \square$$

Proposition 8.1.15. *If $A \in L(X,Y)$ is linear then it is completely determined by its values on a basis of X. Furthermore, if B is a basis of X, then any function $\widetilde{A}\colon B \to Y$ extends to a linear function on X.*

We will only prove this proposition for finite dimensional spaces, as we do not need infinite dimensional spaces. For infinite dimensional spaces, the proof is essentially the same, but a little trickier to write, so let us stick with finitely many dimensions.

Proof. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis and suppose $Ax_j = y_j$. Every $x \in X$ has a unique representation

$$x = \sum_{j=1}^{n} b_j x_j$$

for some numbers $b_1, b_2, \ldots, b_n$. By linearity

$$Ax = A\sum_{j=1}^{n} b_j x_j = \sum_{j=1}^{n} b_j Ax_j = \sum_{j=1}^{n} b_j y_j.$$

The "furthermore" follows by setting $y_j := \widetilde{A}(x_j)$, and defining the extension as $Ax := \sum_{j=1}^{n} b_j y_j$. The function is well defined by uniqueness of the representation of x. We leave it to the reader to check that A is linear. $\square$

The next proposition only works for finite dimensional vector spaces. It is a special case of the so-called rank-nullity theorem from linear algebra.

Proposition 8.1.16. *If X is a finite dimensional vector space and $A \in L(X)$, then A is one-to-one if and only if it is onto.*

Proof. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis for X. Suppose A is one-to-one. Now suppose

$$\sum_{j=1}^{n} c_j A x_j = A \sum_{j=1}^{n} c_j x_j = 0.$$

As A is one-to-one, the only vector that is taken to 0 is 0 itself. Hence,

$$0 = \sum_{j=1}^{n} c_j x_j$$

and $c_j = 0$ for all j. So $\{Ax_1, Ax_2, \ldots, Ax_n\}$ is a linearly independent set. By Proposition 8.1.12 and the fact that the dimension is n, we have that $\{Ax_1, Ax_2, \ldots, Ax_n\}$ span X. Any point $x \in X$ can be written as

$$x = \sum_{j=1}^{n} a_j A x_j = A \sum_{j=1}^{n} a_j x_j,$$

so A is onto.

Now suppose A is onto. As A is determined by the action on the basis we see that every element of X has to be in the span of $\{Ax_1, Ax_2, \ldots, Ax_n\}$. Suppose

$$A \sum_{j=1}^{n} c_j x_j = \sum_{j=1}^{n} c_j A x_j = 0.$$

By Proposition 8.1.12 as $\{Ax_1, Ax_2, \ldots, Ax_n\}$ span X, the set is independent, and hence $c_j = 0$ for all j. In other words if $Ax = 0$, then $x = 0$. This means that A is one-to-one: If $Ax = Ay$, then $A(x - y) = 0$ and so $x = y$. □

We leave the proof of the next proposition as an exercise.

Proposition 8.1.17. *If X and Y are finite dimensional vector spaces, then $L(X, Y)$ is also finite dimensional.*

Finally let us note that we often identify a finite dimensional vector space X of dimension n with $\mathbb{R}^n$, provided we fix a basis $\{x_1, x_2, \ldots, x_n\}$ in X. That is, we define a bijective linear map $A \in L(X, \mathbb{R}^n)$ by $Ax_j = e_j$, where $\{e_1, e_2, \ldots, e_n\}$. Then we have the correspondence

$$\sum_{j=1}^{n} c_j x_j \in X \quad \overset{A}{\mapsto} \quad (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n.$$

8.1.4 Convexity

A subset U of a vector space is *convex* if whenever $x, y \in U$, the line segment from x to y lies in U. That is, if the *convex combination* $(1-t)x+ty$ is in U for all $t \in [0,1]$. See Figure 8.1.

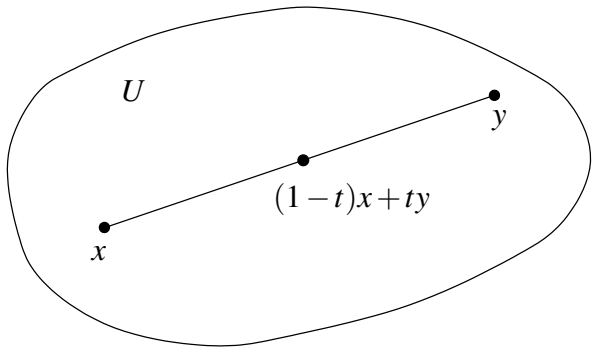

Figure 8.1: Convexity.

Note that in $\mathbb{R}$, every connected interval is convex. In $\mathbb{R}^2$ (or higher dimensions) there are lots of nonconvex connected sets. For example the set $\mathbb{R}^2 \setminus \{0\}$ is not convex but it is connected. To see this simply take any $x \in \mathbb{R}^2 \setminus \{0\}$ and let $y := -x$. Then $(1/2)x + (1/2)y = 0$, which is not in the set. On the other hand, the ball $B(x,r) \subset \mathbb{R}^n$ (using the standard metric on $\mathbb{R}^n$) is always convex by the triangle inequality.

Exercise* 8.1.1: Show that in $\mathbb{R}^n$ any ball $B(x,r)$ for $x \in \mathbb{R}^n$ and $r > 0$ is convex.*

Example 8.1.18: Any subspace V of a vector space X is convex.

Example 8.1.19: A somewhat more complicated example is given by the following. Let $C([0,1],\mathbb{R})$ be the vector space of continuous real valued functions on $\mathbb{R}$. Let $X \subset C([0,1],\mathbb{R})$ be the set of those f such

$$\int_0^1 f(x)\,dx \leq 1 \qquad \text{and} \qquad f(x) \geq 0 \text{ for all } x \in [0,1].$$

Then X is convex. Take $t \in [0,1]$ and note that if $f, g \in X$ then $tf(x) + (1-t)g(x) \geq 0$ for all x. Furthermore

$$\int_0^1 \big(tf(x) + (1-t)g(x)\big)\,dx = t\int_0^1 f(x)\,dx + (1-t)\int_0^1 g(x)\,dx \leq 1.$$

Note that X is not a subspace of $C([0,1],\mathbb{R})$.

Proposition 8.1.20. *The intersection two convex sets is convex. In fact, if $\{C_\lambda\}_{\lambda \in I}$ is an arbitrary collection of convex sets, then*

$$C := \bigcap_{\lambda \in I} C_\lambda$$

is convex.

Proof. If $x, y \in C$, then $x, y \in C_\lambda$ for all $\lambda \in I$, and hence if $t \in [0,1]$, then $tx + (1-t)y \in C_\lambda$ for all $\lambda \in I$. Therefore $tx + (1-t)y \in C$ and C is convex. □

Proposition 8.1.21. *Let $T : V \to W$ be a linear mapping between two vector spaces and let $C \subset V$ be a convex set. Then $T(C)$ is convex.*

Proof. Take any two points $p, q \in T(C)$. Pick $x, y \in C$ such that $Tx = p$ and $Ty = q$. As C is convex then for all $t \in [0,1]$ we have $tx + (1-t)y \in C$, so

$$tp + (1-t)q = tTx + (1-t)Ty = T\big(tx + (1-t)y\big) \in T(C). \qquad \square$$

For completeness, a very useful construction is the *convex hull*. Given any set $S \subset V$ of a vector space, define the convex hull of S, by

$$\operatorname{co}(S) := \bigcap \{C \subset V : S \subset C, \text{ and } C \text{ is convex}\}.$$

That is, the convex hull is the smallest convex set containing S. By a proposition above, the intersection of convex sets is convex and hence, the convex hull is convex.

Example 8.1.22: The convex hull of 0 and 1 in $\mathbb{R}$ is $[0,1]$. Proof: Any convex set containing 0 and 1 must contain $[0,1]$. The set $[0,1]$ is convex, therefore it must be the convex hull.

8.1.5 Exercises

Exercise* 8.1.2: Verify that $\mathbb{R}^n$ is a vector space.*

Exercise* 8.1.3: Let X be a vector space. Prove that a finite set of vectors $\{x_1, \ldots, x_n\} \subset X$ is linearly independent if and only if for every $j = 1, 2, \ldots, n$*

$$\operatorname{span}(\{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n\}) \subsetneq \operatorname{span}(\{x_1, \ldots, x_n\}).$$

That is, the span of the set with one vector removed is strictly smaller.

Exercise* 8.1.4** (Challenging): Prove that $C([0,1], \mathbb{R})$ is an infinite dimensional vector space where the operations are defined in the obvious way: $s = f + g$ and $m = fg$ are defined as $s(x) := f(x) + g(x)$ and $m(x) := f(x)g(x)$. Hint: for the dimension, think of functions that are only nonzero on the interval $(1/n+1, 1/n)$.*

Exercise* 8.1.5: Let $k \colon [0,1]^2 \to \mathbb{R}$ be continuous. Show that $L \colon C([0,1], \mathbb{R}) \to C([0,1], \mathbb{R})$ defined by*

$$Lf(y) := \int_0^1 k(x,y) f(x)\, dx$$

is a linear operator. That is, show that L is well defined (that Lf is continuous), and that L is linear.

Exercise* 8.1.6: Let $\mathscr{P}_n$ be the vector space of polynomials in one variable of degree n or less. Show that $\mathscr{P}_n$ is a vector space of dimension $n+1$.*

Exercise* 8.1.7: Let $\mathbb{R}[t]$ be the vector space of polynomials in one variable t. Let $D\colon \mathbb{R}[t] \to \mathbb{R}[t]$ be the derivative operator (derivative in t). Show that D is a linear operator.*

Exercise* 8.1.8: Let us show that Proposition 8.1.16 only works in finite dimensions. Take $\mathbb{R}[t]$ and define the operator $A\colon \mathbb{R}[t] \to \mathbb{R}[t]$ by $A\big(P(t)\big) = tP(t)$. Show that A is linear and one-to-one, but show that it is not onto.*

Exercise* 8.1.9: Finish the proof of Proposition 8.1.15 in finite dimensional case. That is, suppose, $\{x_1, x_2, \ldots x_n\}$ is a basis of X, $\{y_1, y_2, \ldots y_n\} \subset Y$ and we define a function*

$$Ax := \sum_{j=1}^{n} b_j y_j, \qquad \text{if} \quad x = \sum_{j=1}^{n} b_j x_j.$$

Then prove that $A\colon X \to Y$ is linear.

Exercise* 8.1.10: Prove Proposition 8.1.17. Hint: A linear operator is determined by its action on a basis. So given two bases $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ for X and Y respectively, consider the linear operators A_{jk} that send $A_{jk}x_j = y_k$, and $A_{jk}x_\ell = 0$ if $\ell \neq j$.*

Exercise* 8.1.11** (Easy): Suppose X and Y are vector spaces and $A \in L(X,Y)$ is a linear operator.*
a) Show that the nullspace $N := \{x \in X : Ax = 0\}$ is a vectorspace.
b) Show that the range $R := \{y \in Y : Ax = y \text{ for some } x \in X\}$ is a vectorspace.

Exercise* 8.1.12** (Easy): Show by example that a union of convex sets need not be convex.***

Exercise* 8.1.13: Compute the convex hull of the set of 3 points $\{(0,0),(0,1),(1,1)\}$ in $\mathbb{R}^2$.*

Exercise* 8.1.14: Show that the set $\{(x,y) \in \mathbb{R}^2 : y > x^2\}$ is a convex set.*

Exercise* 8.1.15: Show that every convex set in $\mathbb{R}^n$ is connected using the standard topology on $\mathbb{R}^n$.*

Exercise* 8.1.16: Suppose $K \subset \mathbb{R}^2$ is a convex set such that the only point of the form $(x,0)$ in K is the point $(0,0)$. Further suppose that there $(0,1) \in K$ and $(1,1) \in K$. Then show that if $(x,y) \in K$ then $y > 0$ unless $x = 0$.*

8.2 Analysis with vector spaces

Note: 2-3 lectures

8.2.1 Norms

Let us start measuring distance.

Definition 8.2.1. If X is a vector space, then we say a function $\|\cdot\| \colon X \to \mathbb{R}$ is a *norm* if:

(i) $\|x\| \geq 0$, with $\|x\| = 0$ if and only if $x = 0$.

(ii) $\|cx\| = |c|\,\|x\|$ for all $c \in \mathbb{R}$ and $x \in X$.

(iii) $\|x+y\| \leq \|x\| + \|y\|$ for all $x, y \in X$ (Triangle inequality).

Before defining the standard norm on $\mathbb{R}^n$, let us define the standard scalar *dot product* on $\mathbb{R}^n$. For two vectors if $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, define

$$x \cdot y := \sum_{j=1}^{n} x_j y_j.$$

It is easy to see that the dot product is linear in each variable separately, that is, it is a linear mapping when you keep one of the variables constant. The *Euclidean norm* is then defined as

$$\|x\| := \|x\|_{\mathbb{R}^n} := \sqrt{x \cdot x} = \sqrt{(x_1)^2 + (x_2)^2 + \cdots + (x_n)^2}.$$

We normally just use $\|x\|$, but sometimes it will be necessary to emphasize that we are talking about the euclidean norm and use $\|x\|_{\mathbb{R}^n}$. It is easy to see that the Euclidean norm satisfies (i) and (ii). To prove that (iii) holds, the key inequality is the so-called Cauchy-Schwarz inequality that we have seen before. As this inequality is so important let us restate and reprove it using the notation of this chapter.

Theorem 8.2.2 (Cauchy-Schwarz inequality)**.** *Let $x, y \in \mathbb{R}^n$, then*

$$|x \cdot y| \leq \|x\|\|y\| = \sqrt{x \cdot x}\,\sqrt{y \cdot y},$$

with equality if and only if the vectors are scalar multiples of each other.

Proof. If $x = 0$ or $y = 0$, then the theorem holds trivially. So assume $x \neq 0$ and $y \neq 0$.

If x is a scalar multiple of y, that is $x = \lambda y$ for some $\lambda \in \mathbb{R}$, then the theorem holds with equality:

$$|\lambda y \cdot y| = |\lambda|\,|y \cdot y| = |\lambda|\,\|y\|^2 = \|\lambda y\|\|y\|.$$

Next take $x+ty$,

$$\|x+ty\|^2 = (x+ty)\cdot(x+ty) = x\cdot x + x\cdot ty + ty\cdot x + ty\cdot ty = \|x\|^2 + 2t(x\cdot y) + t^2\|y\|^2.$$

If x is not a scalar multiple of y, then $\|x+ty\|^2 > 0$ for all t. So the above polynomial in t is never zero. From elementary algebra it follows that the discriminant must be negative:

$$4(x\cdot y)^2 - 4\|x\|^2\|y\|^2 < 0,$$

or in other words $(x\cdot y)^2 < \|x\|^2\|y\|^2$. □

Item (iii), the triangle inequality, follows via a simple computation:

$$\|x+y\|^2 = x\cdot x + y\cdot y + 2(x\cdot y) \leq \|x\|^2 + \|y\|^2 + 2(\|x\|\|y\|) = (\|x\| + \|y\|)^2.$$

The distance $d(x,y) := \|x-y\|$ is the standard distance function on $\mathbb{R}^n$ that we used when we talked about metric spaces.

In fact, on any vector space X, once we have a norm (any norm), we define a distance $d(x,y) := \|x-y\|$ that makes X into a metric space (an easy exercise).

Definition 8.2.3. Let $A \in L(X,Y)$. Define

$$\|A\| := \sup\{\|Ax\| : x \in X \text{ with } \|x\| = 1\}.$$

The number $\|A\|$ is called the *operator norm*. We will see below that indeed it is a norm (at least for finite dimensional spaces). Again, when necessary to emphasize which norm we are talking about, we may write it as $\|A\|_{L(X,Y)}$.

By linearity for any nonzero vector $x \in X$, $\left\|A\frac{x}{\|x\|}\right\| = \frac{\|Ax\|}{\|x\|}$. The vector $\frac{x}{\|x\|}$ is of norm 1. Therefore,

$$\|A\| = \sup\{\|Ax\| : x \in X \text{ with } \|x\| = 1\} = \sup_{\substack{x\in X\\ x\neq 0}} \frac{\|Ax\|}{\|x\|}.$$

This implies that

$$\|Ax\| \leq \|A\|\|x\|.$$

It is not hard to see from the definition that $\|A\| = 0$ if and only if $A = 0$, that is, if A takes every vector to the zero vector.

It is also not difficult to see the norm of the identity operator:

$$\|I\| = \sup_{\substack{x\in X\\ x\neq 0}} \frac{\|Ix\|}{\|x\|} = \sup_{\substack{x\in X\\ x\neq 0}} \frac{\|x\|}{\|x\|} = 1.$$

For finite dimensional spaces $\|A\|$ is always finite as we prove below. This also implies that A is continuous. For infinite dimensional spaces neither statement needs to be true. For a simple example, take the vector space of continuously differentiable functions on $[0,1]$ and as the norm use the uniform norm. The functions $\sin(nx)$ have norm 1, but the derivatives have norm n. So differentiation (which is a linear operator) has unbounded norm on this space. But let us stick to finite dimensional spaces now.

When we talk about finite dimensional vector space, one often thinks of $\mathbb{R}^n$, although if we have a norm, the norm might perhaps not be the standard euclidean norm. In the exercises, you can prove that every norm is in some sense "equivalent" to the euclidean norm in that the topology it generates is the same. For simplicity we only prove the following proposition for the euclidean space, and the proof for a general finite dimensional space is left as an exercise.

Proposition 8.2.4. *Let X and Y be finite dimensional vector spaces with a norm. If $A \in L(X,Y)$, then $\|A\| < \infty$, and A is uniformly continuous (Lipschitz with constant $\|A\|$).*

Proof. As we said we only prove the proposition for euclidean space so suppose that $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ and the norm is the standard euclidean norm. The general case is left as an exercise.

Let $\{e_1, e_2, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Write $x \in \mathbb{R}^n$, with $\|x\| = 1$, as

$$x = \sum_{j=1}^{n} c_j e_j.$$

Since $e_j \cdot e_\ell = 0$ whenever $j \neq \ell$ and $e_j \cdot e_j = 1$ then $c_j = x \cdot e_j$ and

$$\left|c_j\right| = \left|x \cdot e_j\right| \leq \|x\| \|e_j\| = 1.$$

Then

$$\|Ax\| = \left\| \sum_{j=1}^{n} c_j A e_j \right\| \leq \sum_{j=1}^{n} \left|c_j\right| \|Ae_j\| \leq \sum_{j=1}^{n} \|Ae_j\|.$$

The right hand side does not depend on x. We found a finite upper bound independent of x, so $\|A\| < \infty$.

Now for any vector spaces X and Y, and $A \in L(X,Y)$, suppose that $\|A\| < \infty$. For $v, w \in X$,

$$\|A(v-w)\| \leq \|A\| \|v-w\|.$$

As $\|A\| < \infty$, then this says A is Lipschitz with constant $\|A\|$. □

Proposition 8.2.5. *Let X, Y, and Z be finite dimensional vector spaces with a norm.*

(i) If $A, B \in L(X,Y)$ and $c \in \mathbb{R}$, then

$$\|A+B\| \leq \|A\| + \|B\|, \qquad \|cA\| = |c| \, \|A\|.$$

In particular, the operator norm is a norm on the vector space $L(X,Y)$.

(ii) If $A \in L(X,Y)$ and $B \in L(Y,Z)$, then

$$\|BA\| \leq \|B\|\|A\|.$$

Proof. For (i), let us note that

$$\|(A+B)x\| = \|Ax+Bx\| \leq \|Ax\| + \|Bx\| \leq \|A\|\|x\| + \|B\|\|x\| = (\|A\| + \|B\|)\|x\|.$$

So $\|A+B\| \leq \|A\| + \|B\|$.

Similarly

$$\|(cA)x\| = |c|\,\|Ax\| \leq (|c|\,\|A\|)\|x\|.$$

Thus $\|cA\| \leq |c|\,\|A\|$. Next note

$$|c|\,\|Ax\| = \|cAx\| \leq \|cA\|\|x\|.$$

Hence $|c|\,\|A\| \leq \|cA\|$.

For (ii) write

$$\|BAx\| \leq \|B\|\|Ax\| \leq \|B\|\|A\|\|x\|. \qquad \square$$

As a norm defines a metric, we have a metric space topology on $L(X,Y)$, so we can talk about open/closed sets, continuity, and convergence.

Proposition 8.2.6. *Let X be a finite dimensional vector space with a norm. Let $U \subset L(X)$ be the set of invertible linear operators.*

(i) If $A \in U$ and $B \in L(X)$, and

$$\|A-B\| < \frac{1}{\|A^{-1}\|}, \tag{8.2}$$

then B is invertible.

(ii) U is open and $A \mapsto A^{-1}$ is a continuous function on U.

Let us make sense of this on a simple example. Think back to $\mathbb{R}^1$, where linear operators are just numbers a and the operator norm of a is simply $|a|$. The operator a is invertible ($a^{-1} = 1/a$) whenever $a \neq 0$. The condition $|a-b| < \frac{1}{|a^{-1}|}$ does indeed imply that b is not zero. And $a \mapsto 1/a$ is a continuous map. When $n > 1$, then there are other noninvertible operators than just zero, and in general things are a bit more difficult.

Proof. Let us prove (i). We know something about A^{-1} and something about $A-B$. These are linear operators so let us apply them to a vector.

$$A^{-1}(A-B)x = x - A^{-1}Bx.$$

Therefore,

$$\begin{aligned}\|x\| &= \|A^{-1}(A-B)x + A^{-1}Bx\| \\ &\leq \|A^{-1}\|\|A-B\|\|x\| + \|A^{-1}\|\|Bx\|.\end{aligned}$$

Now assume $x \neq 0$ and so $\|x\| \neq 0$. Using (8.2) we obtain

$$\|x\| < \|x\| + \|A^{-1}\|\|Bx\|,$$

or in other words $\|Bx\| \neq 0$ for all nonzero x, and hence $Bx \neq 0$ for all nonzero x. This is enough to see that B is one-to-one (if $Bx = By$, then $B(x-y) = 0$, so $x = y$). As B is one-to-one operator from X to X which is finite dimensional and hence is invertible.

Let us look at (ii). Fix some $A \in U$. Let B be invertible and near A, that is $\|A-B\|\|A^{-1}\| < 1/2$. Then (8.2) is satisfied. We have shown above (using $B^{-1}y$ instead of x)

$$\|B^{-1}y\| \leq \|A^{-1}\|\|A-B\|\|B^{-1}y\| + \|A^{-1}\|\|y\| \leq 1/2\|B^{-1}y\| + \|A^{-1}\|\|y\|,$$

or

$$\|B^{-1}y\| \leq 2\|A^{-1}\|\|y\|.$$

So $\|B^{-1}\| \leq 2\|A^{-1}\|$.

Now note that

$$A^{-1}(A-B)B^{-1} = A^{-1}(AB^{-1} - I) = B^{-1} - A^{-1},$$

and

$$\|B^{-1} - A^{-1}\| = \|A^{-1}(A-B)B^{-1}\| \leq \|A^{-1}\|\|A-B\|\|B^{-1}\| \leq 2\|A^{-1}\|^2\|A-B\|.$$

Therefore, if as B tends to A, $\|B^{-1} - A^{-1}\|$ tends to 0, and therefore the inverse operation is a continuous function at A. □

8.2.2 Matrices

As we previously noted, once we fix a basis in a finite dimensional vector space X, we can represent a vector of X as an n-tuple of numbers, that is a vector in $\mathbb{R}^n$. The same thing can be done with $L(X,Y)$, which brings us to matrices, which are a convenient way to represent finite-dimensional linear transformations. Suppose we have bases $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$ for vector spaces X and Y respectively. A linear operator is determined by its values on the basis. Given $A \in L(X,Y)$, Ax_j is an element of Y. Therefore, define the numbers $\{a_{i,j}\}$ as follows

$$Ax_j = \sum_{i=1}^{m} a_{i,j} y_i,$$

and write them as a *matrix*

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

And we say A is an m-by-n matrix. The *columns* of the matrix are precisely the coefficients that represent Ax_j. Let us derive the familiar rule for matrix multiplication.

When

$$z = \sum_{j=1}^{n} c_j x_j,$$

then

$$Az = \sum_{j=1}^{n} c_j A x_j = \sum_{j=1}^{n} c_j \left(\sum_{i=1}^{m} a_{i,j} y_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{i,j} c_j \right) y_i,$$

which gives rise to the familiar rule for matrix multiplication.

There is a one-to-one correspondence between matrices and linear operators in $L(X,Y)$. That is, once we fix a basis in X and in Y. If we would choose a different basis, we would get different matrices. This is important, the operator A acts on elements of X, the matrix is something that works with n-tuples of numbers, that is, vectors of $\mathbb{R}^n$.

If B is an n-by-r matrix with entries $b_{j,k}$, then the matrix for $C = AB$ is an m-by-r matrix whose i,kth entry $c_{i,k}$ is

$$c_{i,k} = \sum_{j=1}^{n} a_{i,j} b_{j,k}.$$

A way to remember it is if you order the indices as we do, that is *row,column*, and put the elements in the same order as the matrices, then it is the "middle index" that is "summed-out."

A linear mapping changing one basis to another is a square matrix in which the columns represent basis elements of the second basis in terms of the first basis. We call such a linear mapping an *change of basis*.

Now suppose all the bases are just the standard bases and $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. If we recall the Cauchy-Schwarz inequality we note that

$$\|Az\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{i,j} c_j \right)^2 \leq \sum_{i=1}^{m} \left(\sum_{j=1}^{n} (c_j)^2 \right) \left(\sum_{j=1}^{n} (a_{i,j})^2 \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} (a_{i,j})^2 \right) \|z\|^2.$$

In other words, we have a bound on the operator norm (note that equality rarely happens)

$$\|A\| \leq \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{i,j})^2}.$$

If the entries go to zero, then $\|A\|$ goes to zero. In particular, if A if fixed and B is changing such that the entries of $A-B$ go to zero then B goes to A in operator norm. That is B goes to A in the metric space topology induced by the operator norm. We have proved the first part of:

Proposition 8.2.7. *If $f\colon S \to \mathbb{R}^{nm}$ is a continuous function for a metric space S, then taking the components of f as the entries of a matrix, f is a continuous mapping from S to $L(\mathbb{R}^n, \mathbb{R}^m)$. Conversely if $f\colon S \to L(\mathbb{R}^n, \mathbb{R}^m)$ is a continuous function then the entries of the matrix are continuous functions.*

The proof of the second part is rather easy. Take $f(x)e_j$ and note that is a continuous function to $\mathbb{R}^m$ with standard Euclidean norm: notice that $\|f(x)e_j - f(y)e_j\| = \|\big(f(x)-f(y)\big)e_j\| \leq \|f(x)-f(y)\|$. So as $x \to y$, then $\|f(x)-f(y)\| \to 0$ and so $\|f(x)e_j - f(y)e_j\| \to 0$. Such a function is continuous if and only if its components are continuous and these are the components of the jth column of the matrix $f(x)$.

8.2.3 Determinants

A certain number can be assigned to square matrices that measures how the corresponding linear mapping stretches space. In particular this number, called the determinant, can be used to test for invertibility of a matrix.

First define the symbol $\operatorname{sgn}(x)$ for a number is defined by

$$\operatorname{sgn}(x) := \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Suppose $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a *permutation* of the integers $(1, 2, \ldots, n)$, that is, a reordering of $(1, 2, \ldots, n)$. Any permutation can be obtained by a sequence of transpositions (switchings of two elements). Call a permutation *even* (resp. *odd*) if it takes an even (resp. odd) number of transpositions to get from σ to $(1, 2, \ldots, n)$. It can be shown that this is well defined (exercise). In fact, define

$$\operatorname{sgn}(\sigma) := \operatorname{sgn}(\sigma_1, \ldots, \sigma_n) = \prod_{p<q} \operatorname{sgn}(\sigma_q - \sigma_p). \tag{8.3}$$

Then it can be shown that $\operatorname{sgn}(\sigma)$ is is 1 if σ is even and -1 if σ is odd. This fact can be proved by noting that applying a transposition changes the sign. Then note that the sign of $(1, 2, \ldots, n)$ is 1.

Let S_n be the set of all permutations on n elements (the *symmetric group*). Let $A = [a_{i,j}]$ be a square $n \times n$ matrix. Define the *determinant* of A

$$\det(A) := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i}.$$

Proposition 8.2.8.

(i) $\det(I) = 1$.

(ii) $\det([x_1 \;\; x_2 \;\; \cdots \;\; x_n])$ *as a function of column vectors* x_j *is linear in each variable* x_j *separately.*

(iii) If two columns of a matrix are interchanged, then the determinant changes sign.

(iv) If two columns of A *are equal, then* $\det(A) = 0$.

(v) If a column is zero, then $\det(A) = 0$.

(vi) $A \mapsto \det(A)$ *is a continuous function.*

(vii) $\det \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] = ad - bc$, *and* $\det[a] = a$.

In fact, the determinant is the unique function that satisfies (i), (ii), and (iii). But we digress. By (ii), we mean that if we fix all the vectors $x_1, \ldots, x_n$ except for x_j and think of the determinant as function of x_j, it is a linear function, that is, if $v, w \in \mathbb{R}^n$ are two vectors, and $a, b \in \mathbb{R}$ are scalars then

$$\det([x_1 \;\; \cdots \;\; x_{j-1} \;\; (av+bw) \;\; x_{j+1} \;\; \cdots \;\; x_n]) = \\ a\det([x_1 \;\; \cdots \;\; x_{j-1} \;\; v \;\; x_{j+1} \;\; \cdots \;\; x_n]) + b\det([x_1 \;\; \cdots \;\; x_{j-1} \;\; w \;\; x_{j+1} \;\; \cdots \;\; x_n]).$$

Proof. We go through the proof quickly, as you have likely seen this before.

(i) is trivial. For (ii), notice that each term in the definition of the determinant contains exactly one factor from each column.

Part (iii) follows by noting that switching two columns is like switching the two corresponding numbers in every element in S_n. Hence all the signs are changed. Part (iv) follows because if two columns are equal and we switch them we get the same matrix back and so part (iii) says the determinant must have been 0.

Part (v) follows because the product in each term in the definition includes one element from the zero column. Part (vi) follows as det is a polynomial in the entries of the matrix and hence continuous. We have seen that a function defined on matrices is continuous in the operator norm if it is continuous in the entries. Finally, part (vii) is a direct computation. □

The determinant tells us about areas and volumes, and how they change. For example, in the 1×1 case, a matrix is just a number, and the determinant is exactly this number. It says how the linear mapping "stretches" the space. Similarly for $\mathbb{R}^2$ (and in fact for $\mathbb{R}^n$). Suppose $A \in L(\mathbb{R}^2)$ is a linear transformation. It can be checked directly that the area of the image of the unit square $A([0,1]^2)$ is precisely $|\det(A)|$. The sign of the determinant tells us if the image is flipped or not. This works with arbitrary figures, not just the unit square. The determinant tells us the stretch in the area. In $\mathbb{R}^3$ it will tell us about the 3 dimensional volume, and in n-dimensions about the n-dimensional volume. We claim this without proof.

Proposition 8.2.9. *If* A *and* B *are* $n \times n$ *matrices, then* $\det(AB) = \det(A)\det(B)$. *In particular,* A *is invertible if and only if* $\det(A) \neq 0$ *and in this case,* $\det(A^{-1}) = \frac{1}{\det(A)}$.

Proof. Let $b_1, b_2, \ldots, b_n$ be the columns of B. Then

$$AB = [Ab_1 \quad Ab_2 \quad \cdots \quad Ab_n].$$

That is, the columns of AB are $Ab_1, Ab_2, \ldots, Ab_n$.

Let $b_{j,k}$ denote the elements of B and a_j the columns of A. Note that $Ae_j = a_j$. By linearity of the determinant as proved above we have

$$\begin{aligned}
\det(AB) = \det([Ab_1 \quad Ab_2 \quad \cdots \quad Ab_n]) &= \det\left(\left[\sum_{j=1}^{n} b_{j,1}a_j \quad Ab_2 \quad \cdots \quad Ab_n\right]\right) \\
&= \sum_{j=1}^{n} b_{j,1} \det([a_j \quad Ab_2 \quad \cdots \quad Ab_n]) \\
&= \sum_{1 \leq j_1, j_2, \ldots, j_n \leq n} b_{j_1,1} b_{j_2,2} \cdots b_{j_n,n} \det([a_{j_1} \quad a_{j_2} \quad \cdots \quad a_{j_n}]) \\
&= \left(\sum_{(j_1, j_2, \ldots, j_n) \in S_n} b_{j_1,1} b_{j_2,2} \cdots b_{j_n,n} \operatorname{sgn}(j_1, j_2, \ldots, j_n)\right) \det([a_1 \quad a_2 \quad \cdots \quad a_n]).
\end{aligned}$$

In the above, go from all integers between 1 and n, to just elements of S_n by noting that when two columns in the determinant are the same then the determinant is zero. We then reorder the columns to the original ordering and obtain the sgn.

The conclusion that $\det(AB) = \det(A)\det(B)$ follows by recognizing the determinant of B. We obtain this by plugging in $A = I$. The expression we got for the determinant of B has rows and columns swapped, so as a sidenote, we have also just proved that the determinant of a matrix and its transpose are equal.

To prove the second part of the theorem, suppose A is invertible. Then $A^{-1}A = I$ and consequently $\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I) = 1$. If A is not invertible, then the columns are linearly dependent. That is, suppose

$$\sum_{j=1}^{n} \gamma_j a_j = 0,$$

where not all γ_j are equal to 0. Without loss of generality suppose $\gamma_1 \neq 1$. Take

$$B := \begin{bmatrix} \gamma_1 & 0 & 0 & \cdots & 0 \\ \gamma_2 & 1 & 0 & \cdots & 0 \\ \gamma_3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_n & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Applying the definition of the determinant we see $\det(B) = \gamma_1 \neq 0$. Then $\det(AB) = \det(A)\det(B) = \gamma_1 \det(A)$. The first column of AB is zero, and hence $\det(AB) = 0$. Thus $\det(A) = 0$. $\square$

Proposition 8.2.10. *Determinant is independent of the basis. In other words, if B is invertible then,*

$$\det(A) = \det(B^{-1}AB).$$

Proof follows by noting $\det(B^{-1}AB) = \frac{1}{\det(B)}\det(A)\det(B) = \det(A)$. If in one basis A is the matrix representing a linear operator, then for another basis we can find a matrix B such that the matrix $B^{-1}AB$ takes us to the first basis, applies A in the first basis, and takes us back to the basis we started with. We choose a basis on X, and we represent a linear mapping using a matrix with respect to this basis. We obtain the same determinant as if we had used any other basis. It follows that

$$\det\colon L(X) \to \mathbb{R}$$

is a well-defined function (not just on matrices).

There are three types of so-called *elementary matrices*. Recall again that e_j are the standard basis of $\mathbb{R}^n$. First for some $j = 1, 2, \ldots, n$ and some $\lambda \in \mathbb{R}$, $\lambda \neq 0$, an $n \times n$ matrix E defined by

$$Ee_i = \begin{cases} e_i & \text{if } i \neq j, \\ \lambda e_i & \text{if } i = j. \end{cases}$$

Given any $n \times m$ matrix M the matrix EM is the same matrix as M except with the kth row multiplied by λ. It is an easy computation (exercise) that $\det(E) = \lambda$.

Second, for some j and k with $j \neq k$, and $\lambda \in \mathbb{R}$ an $n \times n$ matrix E defined by

$$Ee_i = \begin{cases} e_i & \text{if } i \neq j, \\ e_i + \lambda e_k & \text{if } i = j. \end{cases}$$

Given any $n \times m$ matrix M the matrix EM is the same matrix as M except with λ times the kth row added to the jth row. It is an easy computation (exercise) that $\det(E) = 1$.

Finally, for some j and k with $j \neq k$ an $n \times n$ matrix E defined by

$$Ee_i = \begin{cases} e_i & \text{if } i \neq j \text{ and } i \neq k, \\ e_k & \text{if } i = j, \\ e_j & \text{if } i = k. \end{cases}$$

Given any $n \times m$ matrix M the matrix EM is the same matrix with jth and kth rows swapped. It is an easy computation (exercise) that $\det(E) = -1$.

Elementary matrices are useful for computing the determinant. The proof of the following proposition is left as an exercise.

Proposition 8.2.11. *Let T be an $n \times n$ invertible matrix. Then there exists a finite sequence of elementary matrices $E_1, E_2, \ldots, E_k$ such that*

$$T = E_1 E_2 \cdots E_k,$$

and

$$\det(T) = \det(E_1)\det(E_2)\cdots\det(E_k).$$

8.2.4 Exercises

Exercise* 8.2.1: If X is a vector space with a norm $\|\cdot\|$, then show that $d(x,y) := \|x-y\|$ makes X a metric space.*

Exercise* 8.2.2** (Easy): Show that for square matrices A and B,* $\det(AB) = \det(BA)$.

Exercise* 8.2.3: For $\mathbb{R}^n$ define*

$$\|x\|_\infty := \max\{|x_1|, |x_2|, \ldots, |x_n|\},$$

sometimes called the sup or the max norm.
a) Show that $\|\cdot\|_\infty$ is a norm on $\mathbb{R}^n$ (defining a different distance).
b) What is the unit ball $B(0,1)$ in this norm?

Exercise* 8.2.4: For $\mathbb{R}^n$ define*

$$\|x\|_1 := \sum_{j=1}^{n} |x_j|,$$

sometimes called the 1-norm (or L^1 norm).
a) Show that $\|\cdot\|_1$ is a norm on $\mathbb{R}^n$ (defining a different distance, sometimes called the taxicab distance).
b) What is the unit ball $B(0,1)$ in this norm?

Exercise* 8.2.5: Using the euclidean norm on $\mathbb{R}^2$. Compute the operator norm of the operators in $L(\mathbb{R}^2)$ given by the matrices:*
a) $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ *b)* $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ *c)* $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ *d)* $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$

Exercise* 8.2.6: Using the standard euclidean norm $\mathbb{R}^n$. Show*
a) Suppose $A \in L(\mathbb{R}, \mathbb{R}^n)$ is defined for $x \in \mathbb{R}$ by $Ax = xa$ for a vector $a \in \mathbb{R}^n$. Then the operator norm $\|A\|_{L(\mathbb{R},\mathbb{R}^n)} = \|a\|_{\mathbb{R}^n}$. (that is the operator norm of A is the euclidean norm of a).
b) Suppose $B \in L(\mathbb{R}^n, \mathbb{R})$ is defined for $x \in \mathbb{R}^n$ by $Bx = b \cdot x$ for a vector $b \in \mathbb{R}^n$. Then the operator norm $\|B\|_{L(\mathbb{R}^n,\mathbb{R})} = \|b\|_{\mathbb{R}^n}$

Exercise* 8.2.7: Suppose $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a permutation of $(1, 2, \ldots, n)$.*
a) Show that we can make a finite number of transpositions (switching of two elements) to get to $(1, 2, \ldots, n)$.
b) Using the definition (8.3) *show that σ is even if* $\operatorname{sgn}(\sigma) = 1$ *and σ is odd if* $\operatorname{sgn}(\sigma) = -1$. *In particular showing that being odd or even is well defined.*

Exercise* 8.2.8: Verify the computation of the determinant for the three types of elementary matrices.*

Exercise* 8.2.9: Prove Proposition 8.2.11.*

Exercise* 8.2.10: a) Suppose $D = [d_{i,j}]$ is an n-by-n* diagonal matrix*, that is, $d_{i,j} = 0$ whenever $i \neq j$. Show that* $\det(D) = d_{1,1} d_{2,2} \cdots d_{n,n}$.
b) Suppose A is a diagonalizable matrix. That is, there exists a matrix B such that $B^{-1}AB = D$ for a diagonal matrix $D = [d_{i,j}]$. Show that $\det(A) = d_{1,1} d_{2,2} \cdots d_{n,n}$.

Exercise* 8.2.11: Take the vectorspace of polynomials $\mathbb{R}[t]$ and the linear operator $D \in L(\mathbb{R}[t])$ that is the differentiation (we proved in an earlier exercise that D is a linear operator). Define the norm on $P(t) = c_0 + c_1 t + \cdots + c_n t^n$ as $\|P\| := \sup\{\left|c_j\right| : j = 0, 1, \ldots, n\}$.*
a) Show that $\|P\|$ is a norm on $\mathbb{R}[t]$.
b) Show that D does not have bounded operator norm, that is $\|D\| = \infty$. Hint: consider the polynomials t^n as n tends to infinity.

Exercise* 8.2.12: In this exercise we finish the proof of Proposition 8.2.4. Let X be any finite dimensional vector space with a norm. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis for X.*
a) Show that the function $f \colon \mathbb{R}^n \to \mathbb{R}$

$$f(c_1, c_2, \ldots, c_n) = \|c_1 x_1 + c_2 x_2 + \cdots + c_n x_n\|$$

is continuous.
b) Show that there exists numbers m and M such that if $c = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$ with $\|c\| = 1$ (standard euclidean norm), then $m \leq \|c_1 x_1 + c_2 x_2 + \cdots + c_n x_n\| \leq M$ (here the the norm is on X).
c) Show that there exists a number B such that if $\|c_1 x_1 + c_2 x_2 + \cdots + c_n x_n\| = 1$, then $\left|c_j\right| \leq B$.
d) Use part (c) to show that if X and Y are finite dimensional vector spaces and $A \in L(X, Y)$, then $\|A\| < \infty$.

Exercise* 8.2.13: Let X be any finite dimensional vector space with a norm $\|\cdot\|$ and basis $\{x_1, x_2, \ldots, x_n\}$. Let $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ and $\|c\|$ be the standard euclidean norm on $\mathbb{R}^n$.*
a) Find that there exist positive numbers $m, M > 0$ such that for

$$m\|c\| \leq \|c_1 x_1 + c_2 x_2 + \cdots + c_n x_n\| \leq M\|c\|.$$

Hint: See previous exercise.
b) Use part (a) to show that of $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on X, then there exist positive numbers $m, M > 0$ (perhaps different than above) such that for all $x \in X$ we have

$$m\|x\|_1 \leq \|x\|_2 \leq M\|x\|_1.$$

c) Now show that $U \subset X$ is open in the metric defined by $\|x - y\|_1$ if and only if it is open in the metric defined by $\|x - y\|_2$. In other words, convergence of sequences, continuity of functions is the same in either norm.

8.3 The derivative

Note: 2–3 lectures

8.3.1 The derivative

Recall that for a function $f\colon \mathbb{R} \to \mathbb{R}$, we defined the derivative at x as

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

In other words, there was a number a (the derivative of f at x) such that

$$\lim_{h\to 0} \left| \frac{f(x+h)-f(x)}{h} - a \right| = \lim_{h\to 0} \left| \frac{f(x+h)-f(x)-ah}{h} \right| = \lim_{h\to 0} \frac{|f(x+h)-f(x)-ah|}{|h|} = 0.$$

Multiplying by a is a linear map in one dimension. That is, we think of $a \in L(\mathbb{R}^1,\mathbb{R}^1)$ which is the best linear approximation of f near x. We use this definition to extend differentiation to more variables.

Definition 8.3.1. Let $U \subset \mathbb{R}^n$ be an open subset and $f\colon U \to \mathbb{R}^m$. We say f is *differentiable* at $x \in U$ if there exists an $A \in L(\mathbb{R}^n,\mathbb{R}^m)$ such that

$$\lim_{\substack{h\to 0\\ h\in\mathbb{R}^n}} \frac{\|f(x+h)-f(x)-Ah\|}{\|h\|} = 0.$$

We write $Df(x) := A$, or $f'(x) := A$, and we say A is the *derivative* of f at x. When f is *differentiable* at all $x \in U$, we say simply that f is differentiable.

For a differentiable function, the derivative of f is a function from U to $L(\mathbb{R}^n,\mathbb{R}^m)$. Compare to the one dimensional case, where the derivative is a function from U to $\mathbb{R}$, but we really want to think of $\mathbb{R}$ here as $L(\mathbb{R}^1,\mathbb{R}^1)$.

The norms above must be on the right spaces of course. The norm in the numerator is on $\mathbb{R}^m$, and the norm in the denominator is on $\mathbb{R}^n$ where h lives. Normally it is understood that $h \in \mathbb{R}^n$ from context. We will not explicitly say so from now on.

We have again cheated somewhat and said that A is *the* derivative. We have not shown yet that there is only one, let us do that now.

Proposition 8.3.2. *Let $U \subset \mathbb{R}^n$ be an open subset and $f\colon U \to \mathbb{R}^m$. Suppose $x \in U$ and there exist $A,B \in L(\mathbb{R}^n,\mathbb{R}^m)$ such that*

$$\lim_{h\to 0} \frac{\|f(x+h)-f(x)-Ah\|}{\|h\|} = 0 \qquad \textit{and} \qquad \lim_{h\to 0} \frac{\|f(x+h)-f(x)-Bh\|}{\|h\|} = 0.$$

Then $A = B$.

Proof.

$$\begin{aligned}\frac{\|(A-B)h\|}{\|h\|} &= \frac{\|f(x+h)-f(x)-Ah-(f(x+h)-f(x)-Bh)\|}{\|h\|} \\ &\leq \frac{\|f(x+h)-f(x)-Ah\|}{\|h\|} + \frac{\|f(x+h)-f(x)-Bh\|}{\|h\|}.\end{aligned}$$

So $\frac{\|(A-B)h\|}{\|h\|} \to 0$ as $h \to 0$. That is, given $\varepsilon > 0$, then for all h in some δ-ball around the origin

$$\varepsilon > \frac{\|(A-B)h\|}{\|h\|} = \left\|(A-B)\frac{h}{\|h\|}\right\|.$$

For any x with $\|x\| = 1$, let $h = (\delta/2)x$, then $\|h\| < \delta$ and $\frac{h}{\|h\|} = x$. So $\|(A-B)x\| < \varepsilon$. Taking the supremum over all x with $\|x\| = 1$ we get the operator norm $\|A-B\| \leq \varepsilon$. As $\varepsilon > 0$ was arbitrary $\|A-B\| = 0$ or in other words $A = B$. □

Example 8.3.3: If $f(x) = Ax$ for a linear mapping A, then $f'(x) = A$. This is easily seen:

$$\frac{\|f(x+h)-f(x)-Ah\|}{\|h\|} = \frac{\|A(x+h)-Ax-Ah\|}{\|h\|} = \frac{0}{\|h\|} = 0.$$

Example 8.3.4: Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x,y) = \big(f_1(x,y), f_2(x,y)\big) := (1+x+2y+x^2, 2x+3y+xy)$. Let us show that f is differentiable at the origin and let us compute compute the derivative, directly using the definition. The derivative is in $L(\mathbb{R}^2,\mathbb{R}^2)$ so it can be represented by a 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$. Suppose $h = (h_1,h_2)$. We need the following expression to go to zero.

$$\frac{\|f(h_1,h_2)-f(0,0)-(ah_1+bh_2, ch_1+dh_2)\|}{\|(h_1,h_2)\|} = \frac{\sqrt{\big((1-a)h_1+(2-b)h_2+h_1^2\big)^2 + \big((2-c)h_1+(3-d)h_2+h_1h_2\big)^2}}{\sqrt{h_1^2+h_2^2}}.$$

If we choose $a = 1$, $b = 2$, $c = 2$, $d = 3$, the expression becomes

$$\frac{\sqrt{h_1^4+h_1^2h_2^2}}{\sqrt{h_1^2+h_2^2}} = |h_1|\,\frac{\sqrt{h_1^2+h_2^2}}{\sqrt{h_1^2+h_2^2}} = |h_1|\,.$$

And this expression does indeed go to zero as $h \to 0$. Therefore the function is differentiable at the origin and the derivative can be represented by the matrix $\left[\begin{smallmatrix} 1 & 2 \\ 2 & 3 \end{smallmatrix}\right]$.

Proposition 8.3.5. *Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^m$ be differentiable at $p \in U$. Then f is continuous at p.*

Proof. Another way to write the differentiability of f at p is to first write

$$r(h) := f(p+h) - f(p) - f'(p)h,$$

and $\frac{\|r(h)\|}{\|h\|}$ must go to zero as $h \to 0$. So $r(h)$ itself must go to zero. The mapping $h \mapsto f'(p)h$ is a linear mapping between finite dimensional spaces, it is therefore continuous and goes to zero as $h \to 0$. Therefore, $f(p+h)$ must go to $f(p)$ as $h \to 0$. That is, f is continuous at p. □

Theorem 8.3.6 (Chain rule)**.** *Let $U \subset \mathbb{R}^n$ be open and let $f \colon U \to \mathbb{R}^m$ be differentiable at $p \in U$. Let $V \subset \mathbb{R}^m$ be open, $f(U) \subset V$ and let $g \colon V \to \mathbb{R}^\ell$ be differentiable at $f(p)$. Then*

$$F(x) = g\big(f(x)\big)$$

is differentiable at p and

$$F'(p) = g'\big(f(p)\big)f'(p).$$

Without the points where things are evaluated, this is sometimes written as $F' = (f \circ g)' = g'f'$. The way to understand it is that the derivative of the composition $g \circ f$ is the composition of the derivatives of g and f. That is, if $f'(p) = A$ and $g'\big(f(p)\big) = B$, then $F'(p) = BA$.

Proof. Let $A := f'(p)$ and $B := g'\big(f(p)\big)$. Take $h \in \mathbb{R}^n$ and write $q = f(p)$, $k = f(p+h) - f(p)$. Let

$$r(h) := f(p+h) - f(p) - Ah.$$

Then $r(h) = k - Ah$ or $Ah = k - r(h)$. We look at the quantity we need to go to zero:

$$\begin{aligned}
\frac{\|F(p+h) - F(p) - BAh\|}{\|h\|} &= \frac{\|g\big(f(p+h)\big) - g\big(f(p)\big) - BAh\|}{\|h\|} \\
&= \frac{\|g(q+k) - g(q) - B\big(k - r(h)\big)\|}{\|h\|} \\
&\leq \frac{\|g(q+k) - g(q) - Bk\|}{\|h\|} + \|B\| \frac{\|r(h)\|}{\|h\|} \\
&= \frac{\|g(q+k) - g(q) - Bk\|}{\|k\|} \frac{\|f(p+h) - f(p)\|}{\|h\|} + \|B\| \frac{\|r(h)\|}{\|h\|}.
\end{aligned}$$

First, $\|B\|$ is constant and f is differentiable at p, so the term $\|B\| \frac{\|r(h)\|}{\|h\|}$ goes to 0. Next as f is continuous at p, we have that as h goes to 0, then k goes to 0. Therefore $\frac{\|g(q+k)-g(q)-Bk\|}{\|k\|}$ goes to 0 because g is differentiable at q. Finally

$$\frac{\|f(p+h) - f(p)\|}{\|h\|} \leq \frac{\|f(p+h) - f(p) - Ah\|}{\|h\|} + \frac{\|Ah\|}{\|h\|} \leq \frac{\|f(p+h) - f(p) - Ah\|}{\|h\|} + \|A\|.$$

As f is differentiable at p, for small enough h $\|f(p+h) - f(p) - Ah\|\|h\|$ is bounded. Therefore the term $\frac{\|f(p+h)-f(p)\|}{\|h\|}$ stays bounded as h goes to 0. Therefore, $\frac{\|F(p+h)-F(p)-BAh\|}{\|h\|}$ goes to zero, and $F'(p) = BA$, which is what was claimed. □

8.3.2 Partial derivatives

There is another way to generalize the derivative from one dimension. We hold all but one variable constant and take the regular derivative.

Definition 8.3.7. Let $f\colon U \to \mathbb{R}$ be a function on an open set $U \subset \mathbb{R}^n$. If the following limit exists we write

$$\frac{\partial f}{\partial x_j}(x) := \lim_{h\to 0} \frac{f(x_1,\ldots,x_{j-1},x_j+h,x_{j+1},\ldots,x_n)-f(x)}{h} = \lim_{h\to 0}\frac{f(x+he_j)-f(x)}{h}.$$

We call $\frac{\partial f}{\partial x_j}(x)$ the *partial derivative* of f with respect to x_j. Sometimes we write $D_j f$ instead.

For a mapping $f\colon U \to \mathbb{R}^m$ we write $f = (f_1, f_2, \ldots, f_m)$, where f_k are real-valued functions. Then we define $\frac{\partial f_k}{\partial x_j}$ (or write it as $D_j f_k$).

Partial derivatives are easier to compute with all the machinery of calculus, and they provide a way to compute the derivative of a function.

Proposition 8.3.8. *Let $U \subset \mathbb{R}^n$ be open and let $f\colon U \to \mathbb{R}^m$ be differentiable at $p \in U$. Then all the partial derivatives at p exist and in terms of the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$, $f'(p)$ is represented by the matrix*

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p) & \frac{\partial f_1}{\partial x_2}(p) & \ldots & \frac{\partial f_1}{\partial x_n}(p) \\ \frac{\partial f_2}{\partial x_1}(p) & \frac{\partial f_2}{\partial x_2}(p) & \ldots & \frac{\partial f_2}{\partial x_n}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \frac{\partial f_m}{\partial x_2}(p) & \ldots & \frac{\partial f_m}{\partial x_n}(p) \end{bmatrix}.$$

In other words

$$f'(p)\,e_j = \sum_{k=1}^{m} \frac{\partial f_k}{\partial x_j}(p)\,e_k.$$

If $v = \sum_{j=1}^n c_j e_j = (c_1, c_2, \ldots, c_n)$, then

$$f'(p)\,v = \sum_{j=1}^{n}\sum_{k=1}^{m} c_j \frac{\partial f_k}{\partial x_j}(p)\,e_k = \sum_{k=1}^{m}\left(\sum_{j=1}^{n} c_j \frac{\partial f_k}{\partial x_j}(p)\right) e_k.$$

Proof. Fix a j and note that

$$\begin{aligned} \left\| \frac{f(p+he_j)-f(p)}{h} - f'(p)e_j \right\| &= \left\| \frac{f(p+he_j)-f(p)-f'(p)he_j}{h} \right\| \\ &= \frac{\|f(p+he_j)-f(p)-f'(p)he_j\|}{\|he_j\|}. \end{aligned}$$

As h goes to 0, the right hand side goes to zero by differentiability of f, and hence

$$\lim_{h\to 0} \frac{f(p+he_j)-f(p)}{h} = f'(p)e_j.$$

Note that f is vector valued. So represent f by components $f=(f_1,f_2,\ldots,f_m)$, and note that taking a limit in $\mathbb{R}^m$ is the same as taking the limit in each component separately. Therefore for any k the partial derivative

$$\frac{\partial f_k}{\partial x_j}(p) = \lim_{h\to 0} \frac{f_k(p+he_j)-f_k(p)}{h}$$

exists and is equal to the kth component of $f'(p)e_j$, and we are done. □

The converse of the proposition is not true. Just because the partial derivatives exist, does not mean that the function is differentiable. See the exercises. However, when the partial derivatives are continuous, we will prove that the converse holds. One of the consequences of the proposition is that if f is differentiable on U, then $f'\colon U \to L(\mathbb{R}^n,\mathbb{R}^m)$ is a continuous function if and only if all the $\frac{\partial f_k}{\partial x_j}$ are continuous functions.

8.3.3 Gradient and directional derivatives

Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ is a differentiable function. We define the *gradient* as

$$\nabla f(x) := \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x)\, e_j.$$

Notice that the gradient gives us a way to represent the action of the derivative as a dot product: $f'(x)v = \nabla f(x)\cdot v$.

Suppose $\gamma\colon (a,b) \subset \mathbb{R} \to \mathbb{R}^n$ is a differentiable function and the image $\gamma\big((a,b)\big) \subset U$. Such a function and its image is sometimes called a *curve*, or a *differentiable curve*. Write $\gamma = (\gamma_1,\gamma_2,\ldots,\gamma_n)$. Let

$$g(t) := f\big(\gamma(t)\big).$$

The function g is differentiable. For purposes of computation we identify $L(\mathbb{R}^1)$ with $\mathbb{R}$, and hence $g'(t)$ can be computed as a number:

$$g'(t) = f'\big(\gamma(t)\big)\gamma'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\big(\gamma(t)\big)\frac{d\gamma_j}{dt}(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\frac{d\gamma_j}{dt}.$$

For convenience, we sometimes leave out the points where we are evaluating as on the right hand side above. Let us rewrite this with the notation of the gradient and the dot product:

$$g'(t) = (\nabla f)\big(\gamma(t)\big)\cdot\gamma'(t) = \nabla f\cdot\gamma'.$$

We use this idea to define derivatives in a specific direction. A direction is simply a vector pointing in that direction. So pick a vector $u \in \mathbb{R}^n$ such that $\|u\| = 1$. Fix $x \in U$. Then define a curve

$$\gamma(t) := x + tu.$$

It is easy to compute that $\gamma'(t) = u$ for all t. By chain rule

$$\frac{d}{dt}\Big|_{t=0}\big[f(x+tu)\big] = (\nabla f)(x) \cdot u,$$

where the notation $\frac{d}{dt}\big|_{t=0}$ represents the derivative evaluated at $t = 0$. We also compute directly

$$\frac{d}{dt}\Big|_{t=0}\big[f(x+tu)\big] = \lim_{h\to 0} \frac{f(x+hu) - f(x)}{h}.$$

We obtain the *directional derivative*, denoted by

$$D_u f(x) := \frac{d}{dt}\Big|_{t=0}\big[f(x+tu)\big],$$

which can be computed by one of the methods above.

Let us suppose $(\nabla f)(x) \neq 0$. By Cauchy-Schwarz inequality we have

$$|D_u f(x)| \leq \|(\nabla f)(x)\|.$$

Equality is achieved when u is a scalar multiple of $(\nabla f)(x)$. That is, when

$$u = \frac{(\nabla f)(x)}{\|(\nabla f)(x)\|},$$

we get $D_u f(x) = \|(\nabla f)(x)\|$. The gradient points in the direction in which the function grows fastest, in other words, in the direction in which $D_u f(x)$ is maximal.

8.3.4 The Jacobian

Definition 8.3.9. Let $U \subset \mathbb{R}^n$ and $f\colon U \to \mathbb{R}^n$ be a differentiable mapping. Then define the *Jacobian*, or *Jacobian determinant**, of f at x as

$$J_f(x) := \det\big(f'(x)\big).$$

Sometimes this is written as

$$\frac{\partial(f_1, f_2, \ldots, f_n)}{\partial(x_1, x_2, \ldots, x_n)}.$$

*The matrix from Proposition 8.3.8 representing $f'(x)$ is sometimes called the *Jacobian matrix*.

This last piece of notation may seem somewhat confusing, but it is useful when you need to specify the exact variables and function components used.

The Jacobian J_f is a real valued function, and when $n = 1$ it is simply the derivative. From the chain rule and the fact that $\det(AB) = \det(A)\det(B)$, it follows that:

$$J_{f\circ g}(x) = J_f\big(g(x)\big)J_g(x).$$

As we mentioned the determinant tells us what happens to area/volume. Similarly, the Jacobian measures how much a differentiable mapping stretches things locally, and if it flips orientation. In particular if the Jacobian is non-zero than we would assume that locally the mapping is invertible (and we would be correct as we will later see).

8.3.5 Exercises

Exercise* 8.3.1: Suppose $\gamma\colon (-1,1) \to \mathbb{R}^n$ and $\alpha\colon (-1,1) \to \mathbb{R}^n$ be two differentiable curves such that $\gamma(0) = \alpha(0)$ and $\gamma'(0) = \alpha'(0)$. Suppose $F\colon \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. Show that*

$$\frac{d}{dt}\Big|_{t=0} F\big(\gamma(t)\big) = \frac{d}{dt}\Big|_{t=0} F\big(\alpha(t)\big).$$

Exercise* 8.3.2: Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x,y) = \sqrt{x^2+y^2}$. Show that f is not differentiable at the origin.*

Exercise* 8.3.3: Using only the definition of the derivative, show that the following $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ are differentiable at the origin and find their derivative.*
a) $f(x,y) := (1+x+xy, x)$,
b) $f(x,y) := (y-y^{10}, x)$,
c) $f(x,y) := \big((x+y+1)^2, (x-y+2)^2\big)$.

Exercise* 8.3.4: Suppose $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ are differentiable functions. Using only the definition of the derivative, show that $h\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $h(x,y) := \big(f(x), g(y)\big)$ is a differentiable function and find the derivative at any point (x,y).*

Exercise* 8.3.5: Define a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ by*

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

a) Show that partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist at all points (including the origin).
b) Show that f is not continuous at the origin (and hence not differentiable).

***Exercise* 8.3.6:** *Define a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ by*

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

a) Show that partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist at all points.
b) Show that for all $u \in \mathbb{R}^2$ with $\|u\| = 1$, the directional derivative $D_u f$ exists at all points.
c) Show that f is continuous at the origin.
d) Show that f is not differentiable at the origin.

***Exercise* 8.3.7:** *Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one, onto, differentiable at all points, and such that f^{-1} is also differentiable at all points.*
a) Show that $f'(p)$ is invertible at all points p and compute $\left(f^{-1}\right)'\big(f(p)\big)$. Hint: consider $p = f^{-1}\big(f(p)\big)$.
b) Let $g\colon \mathbb{R}^n \to \mathbb{R}^n$ be a function differentiable at $q \in \mathbb{R}^n$ and such that $g(q) = q$. Suppose $f(p) = q$ for some $p \in \mathbb{R}^n$. Show $J_g(q) = J_{f^{-1}\circ g\circ f}(p)$ where J_g is the Jacobian determinant.

***Exercise* 8.3.8:** *Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is differentiable and such that $f(x,y) = 0$ if and only if $y = 0$ and such that $\nabla f(0,0) = (1,1)$. Prove that $f(x,y) > 0$ whenever $y > 0$ and $f(x,y) < 0$ whenever $y < 0$.*

***Exercise* 8.3.9:** *Suppose $U \subset \mathbb{R}^n$ is open and $f\colon U \to \mathbb{R}$ is differentiable. Suppose f has a local maximum at $p \in U$. Show that $f'(p) = 0$, that is the zero mapping in $L(\mathbb{R}^n, \mathbb{R})$. That is p is a* critical point *of f.*

***Exercise* 8.3.10:** *Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is differentiable and suppose that whenever $x^2 + y^2 = 1$, then $f(x,y) = 0$. Prove that there exists at least one point (x_0, y_0) such that $\frac{\partial f}{\partial x}(x_0,y_0) = \frac{\partial f}{\partial y}(x_0,y_0) = 0$.*

***Exercise* 8.3.11** (The Peano surface)**:** *Define $f(x,y) := (x - y^2)(2y^2 - x)$. Show*
a) $(0,0)$ is a critical point, that is $f'(0,0) = 0$, that is the zero linear map in $L(\mathbb{R}^2, \mathbb{R})$.
b) For every direction, that is (x,y) such that $x^2 + y^2 = 1$ the "restriction of f to the line containing the points $(0,0)$ and (x,y)", that is a function $g(t) := f(tx,ty)$ has a local maximum at $t = 0$.
c) f does not have a local maximum at $(0,0)$.

***Exercise* 8.3.12:** *Suppose $f\colon \mathbb{R} \to \mathbb{R}^n$ is differentiable and $\|f(t)\| = 1$ for all t (that is, we have a curve in the unit sphere). Then show that for all t, treating f' as a vector we have, $f'(t) \cdot f(t) = 0$.*

***Exercise* 8.3.13:** *Define $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x,y) := \big(x, y + \varphi(x)\big)$ for some differentiable function φ of one variable. Show that f is differentiable and and find f'.*

8.4 Continuity and the derivative

Note: 1–2 lectures

8.4.1 Bounding the derivative

Let us prove a "mean value theorem" for vector valued functions.

Lemma 8.4.1. *If $\varphi\colon [a,b] \to \mathbb{R}^n$ is differentiable on (a,b) and continuous on $[a,b]$, then there exists a $t_0 \in (a,b)$ such that*

$$\|\varphi(b)-\varphi(a)\| \leq (b-a)\|\varphi'(t_0)\|.$$

Proof. By mean value theorem on the function $\big(\varphi(b)-\varphi(a)\big)\cdot\varphi(t)$ (the dot is the scalar dot product again) we obtain there is a $t_0 \in (a,b)$ such that

$$\big(\varphi(b)-\varphi(a)\big)\cdot\varphi(b)-\big(\varphi(b)-\varphi(a)\big)\cdot\varphi(a) = \|\varphi(b)-\varphi(a)\|^2 = (b-a)\big(\varphi(b)-\varphi(a)\big)\cdot\varphi'(t_0)$$

where we treat φ' as a simply a column vector of numbers by abuse of notation. Note that in this case, if we think of $\varphi'(t)$ as simply a vector, then by Exercise 8.2.6, $\|\varphi'(t)\|_{L(\mathbb{R},\mathbb{R}^n)} = \|\varphi'(t)\|_{\mathbb{R}^n}$. That is, the euclidean norm of the vector is the same as the operator norm of $\varphi'(t)$.

By Cauchy-Schwarz inequality

$$\|\varphi(b)-\varphi(a)\|^2 = (b-a)\big(\varphi(b)-\varphi(a)\big)\cdot\varphi'(t_0) \leq (b-a)\|\varphi(b)-\varphi(a)\|\|\varphi'(t_0)\|. \qquad \square$$

Recall that a set U is convex if whenever $x,y \in U$, the line segment from x to y lies in U.

Proposition 8.4.2. *Let $U \subset \mathbb{R}^n$ be a convex open set, $f\colon U \to \mathbb{R}^m$ a differentiable function, and an M such that*

$$\|f'(x)\| \leq M$$

for all $x \in U$. Then f is Lipschitz with constant M, that is

$$\|f(x)-f(y)\| \leq M\|x-y\|$$

for all $x,y \in U$.

Proof. Fix x and y in U and note that $(1-t)x+ty \in U$ for all $t \in [0,1]$ by convexity. Next

$$\frac{d}{dt}\Big[f\big((1-t)x+ty\big)\Big] = f'\big((1-t)x+ty\big)(y-x).$$

By the mean value theorem above we get for some $t_0 \in (0,1)$

$$\|f(x)-f(y)\| \leq \left\|\frac{d}{dt}\Big|_{t=t_0}\Big[f\big((1-t)x+ty\big)\Big]\right\| \leq \|f'\big((1-t_0)x+t_0y\big)\|\|y-x\| \leq M\|y-x\|. \qquad \square$$

Example 8.4.3: If U is not convex the proposition is not true. To see this fact, take the set

$$U = \{(x,y) : 0.9 < x^2 + y^2 < 1.1\} \setminus \{(x,0) : x < 0\}.$$

Let $f(x,y)$ be the angle that the line from the origin to (x,y) makes with the positive x axis. You can even write the formula for f:

$$f(x,y) = 2\arctan\left(\frac{y}{x+\sqrt{x^2+y^2}}\right).$$

Think spiral staircase with room in the middle. See Figure 8.2.

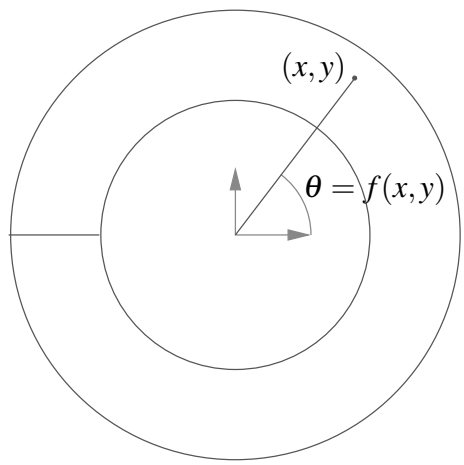

Figure 8.2: A non-Lipschitz function with uniformly bounded derivative.

The function is differentiable, and the derivative is bounded on U, which is not hard to see. Thinking of what happens near where the negative x-axis cuts the annulus in half, we see that the conclusion of the proposition cannot hold.

Let us solve the differential equation $f' = 0$.

Corollary 8.4.4. *If $U \subset \mathbb{R}^n$ is connected and $f\colon U \to \mathbb{R}^m$ is differentiable and $f'(x) = 0$, for all $x \in U$, then f is constant.*

Proof. For any $x \in U$, there is a ball $B(x,\delta) \subset U$. The ball $B(x,\delta)$ is convex. Since $\|f'(y)\| \leq 0$ for all $y \in B(x,\delta)$ then by the theorem, $\|f(x) - f(y)\| \leq 0\|x-y\| = 0$, so $f(x) = f(y)$ for all $y \in B(x,\delta)$.

This means that $f^{-1}(c)$ is open for any $c \in \mathbb{R}^m$. Suppose $f^{-1}(c)$ is nonempty. The two sets

$$U' = f^{-1}(c), \qquad U'' = f^{-1}(\mathbb{R}^m \setminus \{c\}) = \bigcup_{\substack{a \in \mathbb{R}^m \\ a \neq c}} f^{-1}(a)$$

are open disjoint, and further $U = U' \cup U''$. So as U' is nonempty, and U is connected, we have that $U'' = \emptyset$. So $f(x) = c$ for all $x \in U$. $\square$

8.4.2 Continuously differentiable functions

Definition 8.4.5. We say $f\colon U \subset \mathbb{R}^n \to \mathbb{R}^m$ is *continuously differentiable*, or $C^1(U)$ if f is differentiable and $f'\colon U \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.

Proposition 8.4.6. *Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^m$. The function f is continuously differentiable if and only if all the partial derivatives exist and are continuous.*

Without continuity the theorem does not hold. Just because partial derivatives exist does not mean that f is differentiable, in fact, f may not even be continuous. See the exercises for the last section and also for this section.

Proof. We have seen that if f is differentiable, then the partial derivatives exist. Furthermore, the partial derivatives are the entries of the matrix of $f'(x)$. So if $f'\colon U \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous, then the entries are continuous, hence the partial derivatives are continuous.

To prove the opposite direction, suppose the partial derivatives exist and are continuous. Fix $x \in U$. If we show that $f'(x)$ exists we are done, because the entries of the matrix $f'(x)$ are then the partial derivatives and if the entries are continuous functions, the matrix valued function f' is continuous.

Let us do induction on dimension. First let us note that the conclusion is true when $n = 1$. In this case the derivative is just the regular derivative (exercise: you should check that the fact that the function is vector valued is not a problem).

Suppose the conclusion is true for $\mathbb{R}^{n-1}$, that is, if we restrict to the first $n-1$ variables, the conclusion is true. It is easy to see that the first $n-1$ partial derivatives of f restricted to the set where the last coordinate is fixed are the same as those for f. In the following we think of $\mathbb{R}^{n-1}$ as a subset of $\mathbb{R}^n$, that is the set in $\mathbb{R}^n$ where $x_n = 0$. Let

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}, \qquad A_1 = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_{n-1}}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_{n-1}}(x) \end{bmatrix}, \qquad v = \begin{bmatrix} \frac{\partial f_1}{\partial x_n}(x) \\ \vdots \\ \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}.$$

Let $\varepsilon > 0$ be given. Let $\delta > 0$ be such that for any $k \in \mathbb{R}^{n-1}$ with $\|k\| < \delta$ we have

$$\frac{\|f(x+k) - f(x) - A_1 k\|}{\|k\|} < \varepsilon.$$

By continuity of the partial derivatives, suppose δ is small enough so that

$$\left| \frac{\partial f_j}{\partial x_n}(x+h) - \frac{\partial f_j}{\partial x_n}(x) \right| < \varepsilon,$$

for all j and all h with $\|h\| < \delta$.

Let $h = h_1 + te_n$ be a vector in $\mathbb{R}^n$ where $h_1 \in \mathbb{R}^{n-1}$ such that $\|h\| < \delta$. Then $\|h_1\| \leq \|h\| < \delta$. Note that $Ah = A_1 h_1 + tv$.

$$\begin{aligned}
\|f(x+h) - f(x) - Ah\| &= \|f(x+h_1+te_n) - f(x+h_1) - tv + f(x+h_1) - f(x) - A_1 h_1\| \\
&\leq \|f(x+h_1+te_n) - f(x+h_1) - tv\| + \|f(x+h_1) - f(x) - A_1 h_1\| \\
&\leq \|f(x+h_1+te_n) - f(x+h_1) - tv\| + \varepsilon\|h_1\|.
\end{aligned}$$

As all the partial derivatives exist then by the mean value theorem for each j there is some $\theta_j \in [0,t]$ (or $[t,0]$ if $t < 0$), such that

$$f_j(x+h_1+te_n) - f_j(x+h_1) = t\frac{\partial f_j}{\partial x_n}(x+h_1+\theta_j e_n).$$

Note that if $\|h\| < \delta$ then $\|h_1 + \theta_j e_n\| \leq \|h\| < \delta$. So to finish the estimate

$$\begin{aligned}
\|f(x+h) - f(x) - Ah\| &\leq \|f(x+h_1+te_n) - f(x+h_1) - tv\| + \varepsilon\|h_1\| \\
&\leq \sqrt{\sum_{j=1}^{m}\left(t\frac{\partial f_j}{\partial x_n}(x+h_1+\theta_j e_n) - t\frac{\partial f_j}{\partial x_n}(x)\right)^2} + \varepsilon\|h_1\| \\
&\leq \sqrt{m}\,\varepsilon\,|t| + \varepsilon\|h_1\| \\
&\leq (\sqrt{m}+1)\varepsilon\|h\|.
\end{aligned}$$

□

8.4.3 Exercises

***Exercise* 8.4.1:** *Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ as*

$$f(x,y) := \begin{cases} (x^2+y^2)\sin\big((x^2+y^2)^{-1}\big) & \text{if } (x,y) \neq (0,0), \\ 0 & \text{else.} \end{cases}$$

Show that f is differentiable at the origin, but that it is not continuously differentiable.

***Exercise* 8.4.2:** *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the function from Exercise 8.3.5, that is,*

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

Compute the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ at all points and show that these are not continuous functions.

***Exercise* 8.4.3:** *Let $B(0,1) \subset \mathbb{R}^2$ be the unit ball (disc), that is, the set given by $x^2 + y^2 < 1$. Suppose $f\colon B(0,1) \to \mathbb{R}$ is a differentiable function such that $|f(0,0)| \leq 1$, and $\left|\frac{\partial f}{\partial x}\right| \leq 1$ and $\left|\frac{\partial f}{\partial y}\right| \leq 1$ for all points in $B(0,1)$.*
a) Find an $M \in \mathbb{R}$ such that $\|f'(x,y)\| \leq M$ for all $(x,y) \in B(0,1)$.
b) Find a $B \in \mathbb{R}$ such that $|f(x,y)| \leq B$ for all $(x,y) \in B(0,1)$.

***Exercise* 8.4.4:** *Define $\varphi\colon [0,2\pi] \to \mathbb{R}^2$ by $\varphi(t) = \big(\sin(t), \cos(t)\big)$. Compute $\varphi'(t)$ for all t. Compute $\|\varphi'(t)\|$ for all t. Notice that $\varphi'(t)$ is never zero, yet $\varphi(0) = \varphi(2\pi)$, therefore, Rolle's theorem is not true in more than one dimension.*

***Exercise* 8.4.5:** *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a function such that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist at all points and there exists an $M \in \mathbb{R}$ such that $\left|\frac{\partial f}{\partial x}\right| \leq M$ and $\left|\frac{\partial f}{\partial y}\right| \leq M$ at all points. Show that f is continuous.*

***Exercise* 8.4.6:** *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a function and $M \in R$, such that for every $(x,y) \in \mathbb{R}^2$, the function $g(t) := f(xt, yt)$ is differentiable and $|g'(t)| \leq M$.*
a) Show that f is continuous at $(0,0)$.
b) Find an example of such an f which is not continuous at every other point of $\mathbb{R}^2$ (Hint: Think back to how did we construct a nowhere continuous function on $[0,1]$).

8.5 Inverse and implicit function theorem

Note: 2–3 lectures

To prove the inverse function theorem we use the contraction mapping principle we have seen in chapter 7 and that we have used to prove Picard's theorem. Recall that a mapping $f\colon X \to X'$ between two metric spaces (X,d) and (X',d') is called a contraction if there exists a $k<1$ such that

$$d'\bigl(f(x),f(y)\bigr) \leq kd(x,y) \quad \text{for all } x,y \in X.$$

The contraction mapping principle says that if $f\colon X \to X$ is a contraction and X is a complete metric space, then there exists a unique fixed point, that is, there exists a unique $x \in X$ such that $f(x)=x$.

Intuitively if a function is differentiable, then it locally "behaves like" the derivative (which is a linear function). The idea of the inverse function theorem is that if a function is differentiable and the derivative is invertible, the function is (locally) invertible.

Theorem 8.5.1 (Inverse function theorem)**.** *Let $U \subset \mathbb{R}^n$ be a set and let $f\colon U \to \mathbb{R}^n$ be a continuously differentiable function. Also suppose $p \in U$, $f(p)=q$, and $f'(p)$ is invertible (that is, $J_f(p) \neq 0$). Then there exist open sets $V,W \subset \mathbb{R}^n$ such that $p \in V \subset U$, $f(V)=W$ and $f|_V$ is one-to-one and onto. Furthermore, the inverse $g(y)=(f|_V)^{-1}(y)$ is continuously differentiable and*

$$g'(y) = \bigl(f'(x)\bigr)^{-1}, \qquad \textit{for all } x \in V,\ y=f(x).$$

Proof. Write $A=f'(p)$. As f' is continuous, there exists an open ball V around p such that

$$\|A-f'(x)\| < \frac{1}{2\|A^{-1}\|} \qquad \text{for all } x \in V.$$

Note that $f'(x)$ is invertible for all $x \in V$.

Given $y \in \mathbb{R}^n$ we define $\varphi_y\colon C \to \mathbb{R}^n$

$$\varphi_y(x) = x + A^{-1}\bigl(y-f(x)\bigr).$$

As A^{-1} is one-to-one, then $\varphi_y(x)=x$ (x is a fixed point) if only if $y-f(x)=0$, or in other words $f(x)=y$. Using chain rule we obtain

$$\varphi_y'(x) = I - A^{-1}f'(x) = A^{-1}\bigl(A-f'(x)\bigr).$$

So for $x \in V$ we have

$$\|\varphi_y'(x)\| \leq \|A^{-1}\|\|A-f'(x)\| < 1/2.$$

As V is a ball it is convex, and hence

$$\|\varphi_y(x_1)-\varphi_y(x_2)\| \leq \frac{1}{2}\|x_1-x_2\| \qquad \text{for all } x_1,x_2 \in V.$$

In other words φ_y is a contraction defined on V, though we so far do not know what is the range of φ_y. We cannot apply the fixed point theorem, but we can say that φ_y has at most one fixed point (note proof of uniqueness in the contraction mapping principle). That is, there exists at most one $x \in V$ such that $f(x) = y$, and so $f|_V$ is one-to-one.

Let $W = f(V)$. We need to show that W is open. Take a $y_1 \in W$, then there is a unique $x_1 \in V$ such that $f(x_1) = y_1$. Let $r > 0$ be small enough such that the closed ball $C(x_1, r) \subset V$ (such $r > 0$ exists as V is open).

Suppose y is such that

$$\|y - y_1\| < \frac{r}{2\|A^{-1}\|}.$$

If we show that $y \in W$, then we have shown that W is open. Define $\varphi_y(x) = x + A^{-1}\big(y - f(x)\big)$ as before. If $x \in C(x_1, r)$, then

$$\begin{aligned}
\|\varphi_y(x) - x_1\| &\leq \|\varphi_y(x) - \varphi_y(x_1)\| + \|\varphi_y(x_1) - x_1\| \\
&\leq \frac{1}{2}\|x - x_1\| + \|A^{-1}(y - y_1)\| \\
&\leq \frac{1}{2}r + \|A^{-1}\|\|y - y_1\| \\
&< \frac{1}{2}r + \|A^{-1}\|\frac{r}{2\|A^{-1}\|} = r.
\end{aligned}$$

So φ_y takes $C(x_1, r)$ into $B(x_1, r) \subset C(x_1, r)$. It is a contraction on $C(x_1, r)$ and $C(x_1, r)$ is complete (closed subset of $\mathbb{R}^n$ is complete). Apply the contraction mapping principle to obtain a fixed point x, i.e. $\varphi_y(x) = x$. That is $f(x) = y$. So $y \in f\big(C(x_1, r)\big) \subset f(V) = W$. Therefore W is open.

Next we need to show that g is continuously differentiable and compute its derivative. First let us show that it is differentiable. Let $y \in W$ and $k \in \mathbb{R}^n$, $k \neq 0$, such that $y + k \in W$. Then there are unique $x \in V$ and $h \in \mathbb{R}^n$, $h \neq 0$ and $x + h \in V$, such that $f(x) = y$ and $f(x + h) = y + k$ as $f|_V$ is a one-to-one and onto mapping of V onto W. In other words, $g(y) = x$ and $g(y + k) = x + h$. We can still squeeze some information from the fact that φ_y is a contraction.

$$\varphi_y(x + h) - \varphi_y(x) = h + A^{-1}\big(f(x) - f(x + h)\big) = h - A^{-1}k.$$

So

$$\|h - A^{-1}k\| = \|\varphi_y(x + h) - \varphi_y(x)\| \leq \frac{1}{2}\|x + h - x\| = \frac{\|h\|}{2}.$$

By the inverse triangle inequality $\|h\| - \|A^{-1}k\| \leq \frac{1}{2}\|h\|$ so

$$\|h\| \leq 2\|A^{-1}k\| \leq 2\|A^{-1}\|\|k\|.$$

In particular as k goes to 0, so does h.

As $x \in V$, then $f'(x)$ is invertible. Let $B = \big(f'(x)\big)^{-1}$, which is what we think the derivative of g at y is. Then

$$\begin{aligned}
\frac{\|g(y+k)-g(y)-Bk\|}{\|k\|} &= \frac{\|h-Bk\|}{\|k\|} \\
&= \frac{\|h-B\big(f(x+h)-f(x)\big)\|}{\|k\|} \\
&= \frac{\|B\big(f(x+h)-f(x)-f'(x)h\big)\|}{\|k\|} \\
&\leq \|B\| \frac{\|h\|}{\|k\|} \frac{\|f(x+h)-f(x)-f'(x)h\|}{\|h\|} \\
&\leq 2\|B\|\|A^{-1}\| \frac{\|f(x+h)-f(x)-f'(x)h\|}{\|h\|}.
\end{aligned}$$

As k goes to 0, so does h. So the right hand side goes to 0 as f is differentiable, and hence the left hand side also goes to 0. And B is precisely what we wanted $g'(y)$ to be.

We have g is differentiable, let us show it is $C^1(W)$. Now, $g\colon W \to V$ is continuous (it is differentiable), f' is a continuous function from V to $L(\mathbb{R}^n)$, and $X \to X^{-1}$ is a continuous function. $g'(y) = \big(f'\big(g(y)\big)\big)^{-1}$ is the composition of these three continuous functions and hence is continuous. □

Corollary 8.5.2. *Suppose $U \subset \mathbb{R}^n$ is open and $f\colon U \to \mathbb{R}^n$ is a continuously differentiable mapping such that $f'(x)$ is invertible for all $x \in U$. Then given any open set $V \subset U$, $f(V)$ is open. (f is an open mapping).*

Proof. Without loss of generality, suppose $U = V$. For each point $y \in f(V)$, we pick $x \in f^{-1}(y)$ (there could be more than one such point), then by the inverse function theorem there is a neighbourhood of x in V that maps onto an neighbourhood of y. Hence $f(V)$ is open. □

Example 8.5.3: The theorem, and the corollary, is not true if $f'(x)$ is not invertible for some x. For example, the map $f(x,y) = (x,xy)$, maps $\mathbb{R}^2$ onto the set $\mathbb{R}^2 \setminus \{(0,y) : y \neq 0\}$, which is neither open nor closed. In fact $f^{-1}(0,0) = \{(0,y) : y \in \mathbb{R}\}$. This bad behavior only occurs on the y-axis, everywhere else the function is locally invertible. If we avoid the y-axis, f is even one-to-one.

Example 8.5.4: Also note that just because $f'(x)$ is invertible everywhere does not mean that f is one-to-one globally. It is "locally" one-to-one but perhaps not "globally." For an example, take the map $f\colon \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}^2$ defined by $f(x,y) = (x^2-y^2, 2xy)$. It is left to student to show that f is differentiable and the derivative is invertible

On the other hand, the mapping is 2-to-1 globally. For every (a,b) that is not the origin, there are exactly two solutions to $x^2-y^2 = a$ and $2xy = b$. We leave it to the student to show that there is at least one solution, and then notice that replacing x and y with $-x$ and $-y$ we obtain another solution.

The invertibility of the derivative is not a necessary condition, just sufficient, for having a continuous inverse and being an open mapping. For example the function $f(x) = x^3$ is an open mapping from $\mathbb{R}$ to $\mathbb{R}$ and is globally one-to-one with a continuous inverse, although the inverse is not differentiable at $x = 0$.

8.5.1 Implicit function theorem

The inverse function theorem is really a special case of the implicit function theorem which we prove next. Although somewhat ironically we prove the implicit function theorem using the inverse function theorem. What we were showing in the inverse function theorem was that the equation $x - f(y) = 0$ was solvable for y in terms of x if the derivative in terms of y was invertible, that is if $f'(y)$ was invertible. That is there was locally a function g such that $x - f\big(g(x)\big) = 0$.

OK, so how about we look at the equation $f(x,y) = 0$. Obviously this is not solvable for y in terms of x in every case. For example, when $f(x,y)$ does not actually depend on y. For a slightly more complicated example, notice that $x^2 + y^2 - 1 = 0$ defines the unit circle, and we can locally solve for y in terms of x when 1) we are near a point which lies on the unit circle and 2) when we are not at a point where the circle has a vertical tangency, or in other words where $\frac{\partial f}{\partial y} = 0$.

To make things simple we fix some notation. We let $(x,y) \in \mathbb{R}^{n+m}$ denote the coordinates $(x_1,\ldots,x_n,y_1,\ldots,y_m)$. A linear transformation $A \in L(\mathbb{R}^{n+m},\mathbb{R}^m)$ can then always be written as $A = [A_x\ A_y]$ so that $A(x,y) = A_x x + A_y y$, where $A_x \in L(\mathbb{R}^n,\mathbb{R}^m)$ and $A_y \in L(\mathbb{R}^m)$.

Proposition 8.5.5. *Let $A = [A_x\ A_y] \in L(\mathbb{R}^{n+m},\mathbb{R}^m)$ and suppose A_y is invertible. If $B = -(A_y)^{-1}A_x$, then*

$$0 = A(x,Bx) = A_x x + A_y Bx.$$

The proof is obvious. We simply solve and obtain $y = Bx$. Let us show that the same can be done for C^1 functions.

Theorem 8.5.6 (Implicit function theorem)**.** *Let $U \subset \mathbb{R}^{n+m}$ be an open set and let $f\colon U \to \mathbb{R}^m$ be a $C^1(U)$ mapping. Let $(p,q) \in U$ be a point such that $f(p,q) = 0$ and such that*

$$\frac{\partial(f_1,\ldots,f_m)}{\partial(y_1,\ldots,y_m)}(p,q) \neq 0.$$

Then there exists an open set $W \subset \mathbb{R}^n$ with $p \in W$, an open set $W' \subset \mathbb{R}^m$ with $q \in W'$, with $W \times W' \subset U$, and a $C^1(W)$ mapping $g\colon W \to W'$, with $g(p) = q$, and for all $x \in W$, the point $g(x)$ is the unique point in W' such that

$$f\big(x,g(x)\big) = 0.$$

Furthermore, if $[A_x\ A_y] = f'(p,q)$, then

$$g'(p) = -(A_y)^{-1}A_x.$$

The condition $\frac{\partial(f_1,\ldots,f_m)}{\partial(y_1,\ldots,y_m)}(p,q) = \det(A_y) \neq 0$ simply means that A_y is invertible.

Proof. Define $F\colon U \to \mathbb{R}^{n+m}$ by $F(x,y) := \big(x, f(x,y)\big)$. It is clear that F is C^1, and we want to show that the derivative at (p,q) is invertible.

Let us compute the derivative. We know that

$$\frac{\|f(p+h,q+k) - f(p,q) - A_x h - A_y k\|}{\|(h,k)\|}$$

goes to zero as $\|(h,k)\| = \sqrt{\|h\|^2 + \|k\|^2}$ goes to zero. But then so does

$$\frac{\|\big(h, f(p+h,q+k) - f(p,q)\big) - (h, A_x h + A_y k)\|}{\|(h,k)\|} = \frac{\|f(p+h,q+k) - f(p,q) - A_x h - A_y k\|}{\|(h,k)\|}.$$

So the derivative of F at (p,q) takes (h,k) to $(h, A_x h + A_y k)$. If $(h, A_x h + A_y k) = (0,0)$, then $h = 0$, and so $A_y k = 0$. As A_y is one-to-one, then $k = 0$. Therefore $F'(p,q)$ is one-to-one or in other words invertible and we apply the inverse function theorem.

That is, there exists some open set $V \subset \mathbb{R}^{n+m}$ with $(p,0) \in V$, and an inverse mapping $G\colon V \to \mathbb{R}^{n+m}$, that is $F\big(G(x,s)\big) = (x,s)$ for all $(x,s) \in V$ (where $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^m$). Write $G = (G_1, G_2)$ (the first n and the second m components of G). Then

$$F\big(G_1(x,s), G_2(x,s)\big) = \big(G_1(x,s), f(G_1(x,s), G_2(x,s))\big) = (x,s).$$

So $x = G_1(x,s)$ and $f\big(G_1(x,s), G_2(x,s)\big) = f\big(x, G_2(x,s)\big) = s$. Plugging in $s = 0$ we obtain

$$f\big(x, G_2(x,0)\big) = 0.$$

The set $G(V)$ contains a whole neighbourhood of the point (p,q) and therefore there are some open The set V is open and hence there exist some open sets $\widetilde{W}$ and W' such that $\widetilde{W} \times W' \subset G(V)$ with $p \in \widetilde{W}$ and $q \in W'$. Then take $W = \{x \in \widetilde{W} : G_2(x,0) \in W'\}$. The function that takes x to $G_2(x,0)$ is continuous and therefore W is open. We define $g\colon W \to \mathbb{R}^m$ by $g(x) := G_2(x,0)$ which is the g in the theorem. The fact that $g(x)$ is the unique point in W' follows because $W \times W' \subset G(V)$ and G is one-to-one and onto $G(V)$.

Next differentiate

$$x \mapsto f\big(x, g(x)\big),$$

at p, which should be the zero map. The derivative is done in the same way as above. We get that for all $h \in \mathbb{R}^n$

$$0 = A\big(h, g'(p)h\big) = A_x h + A_y g'(p)h,$$

and we obtain the desired derivative for g as well. $\square$

In other words, in the context of the theorem we have m equations in $n+m$ unknowns.

$$\begin{gathered} f_1(x_1,\ldots,x_n,y_1,\ldots,y_m)=0 \\ \vdots \\ f_m(x_1,\ldots,x_n,y_1,\ldots,y_m)=0 \end{gathered}$$

And the condition guaranteeing a solution is that this is a C^1 mapping (that all the components are C^1, or in other words all the partial derivatives exist and are continuous), and the matrix

$$\begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{bmatrix}$$

is invertible at (p,q).

Example 8.5.7: Consider the set $x^2+y^2-(z+1)^3=-1$, $e^x+e^y+e^z=3$ near the point $(0,0,0)$. The function we are looking at is

$$f(x,y,z)=(x^2+y^2-(z+1)^3+1, e^x+e^y+e^z-3).$$

We find that

$$f'=\begin{bmatrix} 2x & 2y & -3(z+1)^2 \\ e^x & e^y & e^z \end{bmatrix}.$$

The matrix

$$\begin{bmatrix} 2(0) & -3(0+1)^2 \\ e^0 & e^0 \end{bmatrix}=\begin{bmatrix} 0 & -3 \\ 1 & 1 \end{bmatrix}$$

is invertible. Hence near $(0,0,0)$ we can find y and z as C^1 functions of x such that for x near 0 we have

$$x^2+y(x)^2-\big(z(x)+1\big)^3=-1, \qquad e^x+e^{y(x)}+e^{z(x)}=3.$$

The theorem does not tell us how to find $y(x)$ and $z(x)$ explicitly, it just tells us they exist. In other words, near the origin the set of solutions is a smooth curve in $\mathbb{R}^3$ that goes through the origin.

Note that there are versions of the theorem for arbitrarily many derivatives. If f has k continuous derivatives, then the solution also has k derivatives.

8.5.2 Exercises

Exercise* 8.5.1: Let $C=\{(x,y)\in\mathbb{R}^2 : x^2+y^2=1\}$.*
a) Solve for y in terms of x near $(0,1)$.
b) Solve for y in terms of x near $(0,-1)$.
c) Solve for x in terms of y near $(-1,0)$.

***Exercise* 8.5.2:** *Define $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x,y) := \big(x, y+h(x)\big)$ for some continuously differentiable function h of one variable.*
a) Show that f is one-to-one and onto.
b) Compute f'.
c) Show that f' is invertible at all points, and compute its inverse.

***Exercise* 8.5.3:** *Define $f\colon \mathbb{R}^2 \to \mathbb{R}^2 \setminus \{(0,0)\}$ by $f(x,y) := \big(e^x\cos(y), e^x\sin(y)\big)$.*
a) Show that f is onto.
b) Show that f' is invertible at all points.
c) Show that f is not one-to-one, in fact for every $(a,b) \in \mathbb{R}^2 \setminus \{(0,0)\}$, there exist infinitely many different points $(x,y) \in \mathbb{R}^2$ such that $f(x,y) = (a,b)$.
Therefore, invertible derivative at every point does not mean that f is invertible globally.

***Exercise* 8.5.4:** *Find a map $f\colon \mathbb{R}^n \to \mathbb{R}^n$ that is one-to-one, onto, continuously differentiable, but $f'(0) = 0$. Hint: Generalize $f(x) = x^3$ from one to n dimensions.*

***Exercise* 8.5.5:** *Consider $z^2 + xz + y = 0$ in $\mathbb{R}^3$. Find an equation $D(x,y) = 0$, such that if $D(x_0,y_0) \neq 0$ and $z^2 + x_0 z + y_0 = 0$ for some $z \in \mathbb{R}$, then for points near (x_0,y_0) there exist exactly two distinct continuously differentiable functions $r_1(x,y)$ and $r_2(x,y)$ such that $z = r_1(x,y)$ and $z = r_2(x,y)$ solve $z^2 + xz + y = 0$. Do you recognize the expression D from algebra?*

***Exercise* 8.5.6:** *Suppose $f\colon (a,b) \to \mathbb{R}^2$ is continuously differentiable and $\frac{\partial f}{\partial x}(t) \neq 0$ for all $t \in (a,b)$. Prove that there exists an interval (c,d) and a continuously differentiable function $g\colon (c,d) \to \mathbb{R}$ such that $(x,y) \in f\big((a,b)\big)$ if and only if $x \in (c,d)$ and $y = g(x)$. In other words, the set $f\big((a,b)\big)$ is a graph of g.*

***Exercise* 8.5.7:** *Define $f\colon \mathbb{R}^2 \to \mathbb{R}^2$*

$$f(x,y) := \begin{cases} \left(x^2 \sin\left(1/x\right) + \frac{x}{2}, y\right) & \text{if } x \neq 0, \\ (0,y) & \text{if } x = 0. \end{cases}$$

a) Show that f is differentiable everywhere.
b) Show that $f'(0,0)$ is invertible.
c) Show that f is not one-to-one in any neighbourhood of the origin (it is not locally invertible, that is, the inverse theorem does not work).
d) Show that f is not continuously differentiable.

Exercise* 8.5.8** (Polar coordinates):*** *Define a mapping $F(r,\theta) := \big(r\cos(\theta), r\sin(\theta)\big)$.*
a) Show that F is continuously differentiable (for all $(r,\theta) \in \mathbb{R}^2$).
b) Compute $F'(0,\theta)$ for any θ.
c) Show that if $r \neq 0$, then $F'(r,\theta)$ is invertible, therefore an inverse of F exists locally as long as $r \neq 0$.
d) Show that $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is onto, and for each point $(x,y) \in \mathbb{R}^2$, the set $F^{-1}(x,y)$ is infinite.
e) Show that $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an open map, despite not satisfying the condition of the inverse function theorem.
f) Show that $F|_{(0,\infty)\times[0,2\pi)}$ is one to one and onto $\mathbb{R}^2 \setminus \{(0,0)\}$.

8.6 Higher order derivatives

Note: less than 1 lecture

Let $U : \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$ a function. Denote by $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ our coordinates. Suppose $\frac{\partial f}{\partial x_j}$ exists everywhere in U, then we note that it is also a function $\frac{\partial f}{\partial x_j} : U \to \mathbb{R}$. Therefore it makes sense to talk about its partial derivatives. We denote the partial derivative of $\frac{\partial f}{\partial x_j}$ with respect to x_k by

$$\frac{\partial^2 f}{\partial x_k \partial x_j} := \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_k}.$$

If $k = j$ then we write $\frac{\partial^2 f}{\partial x_j^2}$ for simplicity.

We define higher order derivatives inductively. Suppose $j_1, j_2, \ldots, j_\ell$ are integers between 1 and n, and suppose

$$\frac{\partial^{\ell-1} f}{\partial x_{j_{\ell-1}} \partial x_{j_{\ell-2}} \cdots \partial x_{j_1}}$$

exists and is differentiable in the variable x_{j_ℓ}, then the partial derivative with respect to that variable is denoted by

$$\frac{\partial^\ell f}{\partial x_{j_\ell} \partial x_{j_{\ell-1}} \cdots \partial x_{j_1}} := \frac{\partial \left(\frac{\partial^{\ell-1} f}{\partial x_{j_{\ell-1}} \partial x_{j_{\ell-2}} \cdots \partial x_{j_1}} \right)}{\partial x_{j_\ell}}.$$

Such a derivative is called a *partial derivative of order* ℓ.

Remark that sometimes the notation $f_{x_j x_k}$ is used for $\frac{\partial^2 f}{\partial x_k \partial x_j}$. This notation swaps the order of derivatives, which may be important.

Definition 8.6.1. If $U \subset \mathbb{R}^n$ is an open set and $f : U \to \mathbb{R}$ a function. We say f is *k-times continuously differentiable function*, or a C^k function, if all partial derivatives of all orders up to and including order k exist and are continuous.

So a continuously differentiable, or C^1, function is one where all partial derivatives exist and are continuous, which agrees with our previous definition due to Proposition 8.4.6. We could have required only that the kth order partial derivatives exist and are continuous, as the existence of lower order derivatives is clearly necessary to even define kth order partial derivatives, and these lower order derivatives will be continuous as they will be differentiable functions.

When the partial derivatives are continuous, we can swap their order.

Proposition 8.6.2. *Suppose $U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}$ is a C^2 function, and j and k are two integers between* 1 *and n. Then*

$$\frac{\partial^2 f}{\partial x_k \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_k}.$$

Proof. Fix a point $p \in U$, and let e_j and e_k be the standard basis vectors and let s and t be two small nonzero real numbers. We pick s and t small enough so that $p + s_0 e_j + t_0 e_k \in U$ for all s_0 and t_0 with $|s_0| \leq |s|$ and $|t_0| \leq |t|$. This is possible since U is open and so contains a small ball (or a box if you wish).

Using the mean value theorem on the partial derivative in x_k of the function $f(p+se_j) - f(p)$, we find a t_0 between 0 and t such that

$$\frac{f(p+se_j+te_k) - f(p+te_k) - f(p+se_j) + f(p)}{t} = \frac{\partial f}{\partial x_k}(p+se_j+t_0e_k) - \frac{\partial f}{\partial x_k}(p+t_0e_k).$$

Next there exists a number s_0 between 0 and s such that

$$\frac{\frac{\partial f}{\partial x_k}(p+se_j+t_0e_k) - \frac{\partial f}{\partial x_k}(p+t_0e_k)}{s} = \frac{\partial^2 f}{\partial x_j \partial x_k}(p+s_0e_j+t_0e_k).$$

In other words

$$g(s,t) := \frac{f(p+se_j+te_k) - f(p+te_k) - f(p+se_j) + f(p)}{st} = \frac{\partial^2 f}{\partial x_j \partial x_k}(p+s_0e_j+t_0e_k).$$

Taking a limit as $(s,t) \in \mathbb{R}^2$ goes to zero we find that (s_0,t_0) also goes to zero and by continuity of the second partial derivatives we find that

$$\lim_{(s,t)\to 0} g(s,t) = \frac{\partial^2 f}{\partial x_j \partial x_k}(p).$$

We now reverse the ordering, starting with the function $f(p+te_k) - f(p)$ we find an s_1 between 0 and s such that

$$\frac{f(p+te_k+se_j) - f(p+se_j) - f(p+te_k) + f(p)}{s} = \frac{\partial f}{\partial x_j}(p+te_k+s_1e_j) - \frac{\partial f}{\partial x_j}(p+s_1e_j).$$

And we find a t_1 between 0 and t

$$\frac{\frac{\partial f}{\partial x_j}(p+te_k+s_1e_j) - \frac{\partial f}{\partial x_j}(p+s_1e_j)}{t} = \frac{\partial^2 f}{\partial x_k \partial x_j}(p+t_1e_k+s_1e_j).$$

Again we find that $g(s,t) = \frac{\partial^2 f}{\partial x_k \partial x_j}(p+t_1e_k+s_1e_j)$ and therefore

$$\lim_{(s,t)\to 0} g(s,t) = \frac{\partial^2 f}{\partial x_k \partial x_j}(p).$$

And therefore the two partial derivatives are equal. □

The proposition does not hold if the derivatives are not continuous. See the exercises. Notice also that we did not really need a C^2 function we only needed the two second order partial derivatives involved to be continuous functions.

8.6.1 Exercises

Exercise* 8.6.1: Suppose $f\colon U \to \mathbb{R}$ is a C^2 function for some open $U \subset \mathbb{R}^n$ and $p \in U$. Use the proof of Proposition 8.6.2 to find an expression in terms of just the values of f (analogue of the difference quotient for the first derivative), whose limit is $\frac{\partial^2 f}{\partial x_j \partial x_k}(p)$.*

Exercise* 8.6.2: Define*

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

Show that
a) The first order partial derivatives exist and are continuous.
b) The partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist, but are not continuous at the origin, and $\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0)$.

Exercise* 8.6.3: Suppose $f\colon U \to \mathbb{R}$ is a C^k function for some open $U \subset \mathbb{R}^n$ and $p \in U$. Suppose $j_1, j_2, \ldots, j_k$ are integers between 1 and n, and suppose $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_k)$ is a permutation of $(1, 2, \ldots, k)$. Prove*

$$\frac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(p) = \frac{\partial^k f}{\partial x_{j_{\sigma_k}} \partial x_{j_{\sigma_{k-1}}} \cdots \partial x_{j_{\sigma_1}}}(p).$$

Exercise* 8.6.4: Suppose $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ be a C^k function such that $\varphi(0,\theta) = \varphi(0,\psi)$ for all $\theta, \psi \in \mathbb{R}$ and $\varphi(r,\theta) = \varphi(r,\theta+2\pi)$ for all $r, \theta \in \mathbb{R}$. Let $F(r,\theta) = \big(r\cos(\theta), r\sin(\theta)\big)$ from Exercise 8.5.8. Show that a function $g\colon \mathbb{R}^2 \to \mathbb{R}$, given $g(x,y) := \varphi\big(F^{-1}(x,y)\big)$ is well defined (notice that $F^{-1}(x,y)$ can only be defined locally), and when restricted to $\mathbb{R}^2 \setminus \{0\}$ it is a C^k function.*

Chapter 9

One dimensional integrals in several variables

9.1 Differentiation under the integral

Note: less than 1 lecture

Let $f(x,y)$ be a function of two variables and define

$$g(y) := \int_a^b f(x,y)\,dx.$$

Suppose f is differentiable in y. The question we ask is when can we "differentiate under the integral", that is, when is it true that g is differentiable and its derivative

$$g'(y) \stackrel{?}{=} \int_a^b \frac{\partial f}{\partial y}(x,y)\,dx.$$

Differentiation is a limit and therefore we are really asking when do the two limitting operations of integration and differentiation commute. As we have seen, this is not always possible, some sort of uniformity is necessary. In particular, the first question we would face is the integrability of $\frac{\partial f}{\partial y}$, but the formula can fail even if $\frac{\partial f}{\partial y}$ is integrable for all y.

Let us prove a simple, but the most useful version of this theorem.

Theorem 9.1.1 (Leibniz integral rule)**.** *Suppose $f\colon [a,b]\times[c,d]\to\mathbb{R}$ is a continuous function, such that $\frac{\partial f}{\partial y}$ exists for all $(x,y)\in[a,b]\times[c,d]$ and is continuous. Define*

$$g(y) := \int_a^b f(x,y)\,dx.$$

Then $g\colon [c,d] \to \mathbb{R}$ *is differentiable and*

$$g'(y) = \int_a^b \frac{\partial f}{\partial y}(x,y)\,dx.$$

The continuity requirements for f and $\frac{\partial f}{\partial y}$ can be weakened, but not dropped outright. The main point is for $\frac{\partial f}{\partial y}$ to exist and be continuous for a small interval in the y direction. In applications, the $[c,d]$ can be a small interval around the point where you need to differentiate.

Proof. Fix $y \in [c,d]$ and let $\varepsilon > 0$ be given. As $\frac{\partial f}{\partial y}$ is continuous on $[a,b] \times [c,d]$ it is uniformly continuous. In particular, there exists $\delta > 0$ such that whenever $y_1 \in [c,d]$ with $|y_1 - y| < \delta$ and all $x \in [a,b]$ we have

$$\left| \frac{\partial f}{\partial y}(x,y_1) - \frac{\partial f}{\partial y}(x,y) \right| < \varepsilon.$$

Suppose h is such that $y+h \in [c,d]$ and $|h| < \delta$. Fix x for a moment and apply mean value theorem to find a y_1 between y and $y+h$ such that

$$\frac{f(x,y+h) - f(x,y)}{h} = \frac{\partial f}{\partial y}(x,y_1).$$

If $|h| < \delta$ then

$$\left| \frac{f(x,y+h) - f(x,y)}{h} - \frac{\partial f}{\partial y}(x,y) \right| = \left| \frac{\partial f}{\partial y}(x,y_1) - \frac{\partial f}{\partial y}(x,y) \right| < \varepsilon.$$

This argument worked for every $x \in [a,b]$. Therefore, as a function of x

$$x \mapsto \frac{f(x,y+h) - f(x,y)}{h} \qquad \text{converges uniformly to} \qquad x \mapsto \frac{\partial f}{\partial y}(x,y) \qquad \text{as } h \to 0.$$

We only defined uniform convergence for sequences although the idea is the same. If you wish you can replace h with $1/n$ above and let $n \to \infty$.

Now consider the difference quotient

$$\frac{g(y+h) - g(y)}{h} = \frac{\int_a^b f(x,y+h)\,dx - \int_a^b f(x,y)\,dx}{h} = \int_a^b \frac{f(x,y+h) - f(x,y)}{h}\,dx.$$

Uniform convergence can be taken underneath the integral and therefore

$$\lim_{h\to 0} \frac{g(y+h) - g(y)}{h} = \int_a^b \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h}\,dx = \int_a^b \frac{\partial f}{\partial y}(x,y)\,dx. \qquad \square$$

Example 9.1.2: Let

$$f(y) = \int_0^1 \sin(x^2 - y^2)\, dx.$$

Then

$$f'(y) = \int_0^1 -2y\cos(x^2 - y^2)\, dx.$$

Example 9.1.3: Suppose we start with

$$\int_0^1 \frac{x-1}{\ln(x)}\, dx.$$

The function under the integral extends to be continuous on $[0,1]$, and hence the integral exists, see exercise below. Trouble is finding it. Introduce a parameter y and define a function:

$$g(y) := \int_0^1 \frac{x^y - 1}{\ln(x)}\, dx.$$

The function $\frac{x^y-1}{\ln(x)}$ also extends to a continuous function of x and y for $(x,y) \in [0,1] \times [0,1]$. Therefore g is a continuous function of on $[0,1]$. In particular $g(0) = 0$. For any $\varepsilon > 0$, the y derivative of the integrand, x^y, is continuous on $[0,1] \times [\varepsilon, 1]$. Therefore, for $y > 0$ we may differentiate under the integral sign

$$g'(y) = \int_0^1 \frac{\ln(x)x^y}{\ln(x)}\, dx = \int_0^1 x^y\, dx = \frac{1}{y+1}.$$

We need to figure out $g(1)$, knowing $g'(y) = \frac{1}{y+1}$ and $g(0) = 0$. By elementary calculus we find $g(1) = \int_0^1 g'(y)\, dy = \ln(2)$. Therefore

$$\int_0^1 \frac{x-1}{\ln(x)}\, dx = \ln(2).$$

Exercise* 9.1.1: Prove the two statements that were asserted in the example.*
a) Prove $\frac{x-1}{\ln(x)}$ *extends to a continuous function of* $[0,1]$.
b) Prove $\frac{x^y-1}{\ln(x)}$ *extends to be a continuous function on* $[0,1] \times [0,1]$.

9.1.1 Exercises

Exercise* 9.1.2: Suppose $h\colon \mathbb{R} \to \mathbb{R}$ is a continuous function. Suppose $g\colon \mathbb{R} \to \mathbb{R}$ is which is continuously differentiable and compactly supported. That is there exists some $M > 0$ such that $g(x) = 0$ whenever $|x| \geq M$. Define*

$$f(x) := \int_{-\infty}^{\infty} h(y)g(x-y)\, dy.$$

Show that f is differentiable.

Exercise* 9.1.3: Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is an infinitely differentiable function (all derivatives exist) such that $f(0) = 0$. Then show that there exists another infinitely differentiable function $g(x)$ such that $f(x) = xg(x)$. Finally show that if $f'(0) \neq 0$ then $g(0) \neq 0$. Hint: first write $f(x) = \int_0^x f'(s)ds$ and then rewrite the integral to go from 0 to 1.*

Exercise* 9.1.4: Compute $\int_0^1 e^{tx}\,dx$. Derive the formula for $\int_0^1 x^n e^x\,dx$ not using itnegration by parts, but by differentiation underneath the integral.*

Exercise* 9.1.5: Let $U \subset \mathbb{R}^n$ be an open set and suppose $f(x, y_1, y_2, \ldots, y_n)$ is a continuous function defined on $[0,1] \times U \subset \mathbb{R}^{n+1}$. Suppose $\frac{\partial f}{\partial y_1}, \frac{\partial f}{\partial y_2}, \ldots, \frac{\partial f}{\partial y_n}$ exist and are continuous on $[0,1] \times U$. Then prove that $F\colon U \to \mathbb{R}$ defined by*

$$F(y_1, y_2, \ldots, y_n) := \int_0^1 f(x, y_1, y_2, \ldots, y_n)\,dx$$

is continuously differentiable.

Exercise* 9.1.6: Show the following counterexample: Let*

$$f(x,y) := \begin{cases} \frac{xy^3}{(x^2+y^2)^2} & \text{if } x \neq 0 \text{ or } y \neq 0, \\ 0 & \text{if } x = 0 \text{ and } y = 0. \end{cases}$$

a) Prove that for any fixed y the function $x \mapsto f(x,y)$ is Riemann integrable on $[0,1]$ and

$$g(y) = \int_0^1 f(x,y)\,dx = \frac{y}{2y^2+2}.$$

Therefore $g'(y)$ exists and we get the continuous function

$$g'(y) = \frac{1-y^2}{2(y^2+1)^2}.$$

b) Prove $\frac{\partial f}{\partial y}$ exists at all x and y and compute it.
c) Show that for all y

$$\int_0^1 \frac{\partial f}{\partial y}(x,y)\,dx$$

exists but

$$g'(0) \neq \int_0^1 \frac{\partial f}{\partial y}(x,0)\,dx.$$

Exercise* 9.1.7: Show the following counterexample: Let*

$$f(x,y) := \begin{cases} xy^2 \sin\left(\frac{1}{x^3 y}\right) & \text{if } x \neq 0 \text{ and } y \neq 0, \\ 0 & \text{if } x = 0 \text{ or } y = 0. \end{cases}$$

a) Prove f is continuous on $[0,1] \times [a,b]$ for any interval $[a,b]$. Therefore the following function is well defined on $[a,b]$

$$g(y) = \int_0^1 f(x,y)\,dx.$$

b) Prove $\frac{\partial f}{\partial y}$ exists for all (x,y) in $[0,1] \times [a,b]$, but is not continuous.
c) Show that $\int_0^1 \frac{\partial f}{\partial y}(x,y)\,dx$ does not exist if $y \neq 0$ even if we take improper integrals.

9.2 Path integrals

Note: 2–3 lectures

9.2.1 Piecewise smooth paths

Definition 9.2.1. A continuously differentiable function $\gamma\colon [a,b] \to \mathbb{R}^n$ is called a *smooth path* or a *continuously differentiable path** if γ is continuously differentiable and $\gamma'(t) \neq 0$ for all $t \in [a,b]$.

The function γ is called a *piecewise smooth path* or a *piecewise continuously differentiable path* if there exist finitely many points $t_0 = a < t_1 < t_2 < \cdots < t_k = b$ such that the restriction of the function $\gamma|_{[t_{j-1},t_j]}$ is smooth path.

We say γ is a *simple path* if $\gamma|_{(a,b)}$ is a one-to-one function. A γ is a *closed path* if $\gamma(a) = \gamma(b)$, that is if the path starts and ends in the same point.

Since γ is a function of one variable, we have seen before that treating $\gamma'(t)$ as a matrix is equivalent to treating it as a vector since it is an $n \times 1$ matrix, that is, a column vector. In fact, by an earlier exercise, even the operator norm of $\gamma'(t)$ is equal to the euclidean norm. Therefore, we will write $\gamma'(t)$ as a vector as is usual, and then $\gamma'(t)$ is just the vector of the derivatives of its components, so if $\gamma(t) = \bigl(\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t)\bigr)$, then $\gamma'(t) = \bigl(\gamma_1'(t), \gamma_2'(t), \ldots, \gamma_n'(t)\bigr)$.

One can often get by with only smooth paths, but for computations, the simplest paths to write down are often piecewise smooth. Note that a piecewise smooth is automatically continuous (exercise).

Generally, it is the direct image $\gamma\bigl([a,b]\bigr)$ that is what we are interested in, although how we parametrize it with γ is also important to some degree. We informally talk about a curve, and often we really mean the set $\gamma\bigl([a,b]\bigr)$, just as before depending on context.

Example 9.2.2: Let $\gamma\colon [0,4] \to \mathbb{R}^2$ be defined by

$$\gamma(t) := \begin{cases} (t,0) & \text{if } t \in [0,1], \\ (1,t-1) & \text{if } t \in (1,2], \\ (3-t,1) & \text{if } t \in (2,3], \\ (0,4-t) & \text{if } t \in (3,4]. \end{cases}$$

Then the reader can check that the path is the unit square traversed counterclockwise. We can check that for example $\gamma|_{[1,2]}(t) = (1,t-1)$ and therefore $(\gamma|_{[1,2]})'(t) = (0,1) \neq 0$. It is good to notice at this point that $(\gamma|_{[1,2]})'(1) = (0,1)$, $(\gamma|_{[0,1]})'(1) = (1,0)$, and $\gamma'(1)$ does not exist. That is, at the corners γ is of course not differentiable, even though the restrictions are differentiable and the derivative depends on which restriction you take.

*The word "smooth" is used sometimes for continuously differentiable and sometimes for infinitely differentiable functions in the literature.

Example 9.2.3: The condition that $\gamma'(t) \neq 0$ means that the image of γ has no "corners" where γ is continuously differentiable. For example, take the function

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

It is left for the reader to check that γ is continuously differentiable, yet the image $\gamma(\mathbb{R}) = \{(x,y) \in \mathbb{R}^2 : (x,y) = (s,0) \text{ or } (x,y) = (0,s) \text{ for some } s \geq 0\}$ has a "corner" at the origin. And that is because $\gamma'(0) = (0,0)$. More complicated examples with even infinitely many corners exist, see the exercises.

The condition that $\gamma'(t) \neq 0$ even at the endpoints guarantees not only no corners, but also that the path ends nicely, that is, can always extend a little bit past the endpoints. Again, see the exercies.

Example 9.2.4: A graph of a continuously differentiable function $f\colon [a,b] \to \mathbb{R}$ is a smooth path. That is, define $\gamma\colon [a,b] \to \mathbb{R}^2$ by

$$\gamma(t) := \bigl(t, f(t)\bigr).$$

Then $\gamma'(t) = \bigl(1, f'(t)\bigr)$, which is never zero.

There are other ways of parametrizing the path. That is, having a different path with the same image. For example, the function that takes t to $(1-t)a + tb$, takes the interval $[0,1]$ to $[a,b]$. So let $\alpha\colon [0,1] \to \mathbb{R}^2$ be defined by

$$\alpha(t) := \bigl((1-t)a + tb, f((1-t)a + tb)\bigr).$$

Then $\alpha'(t) = \bigl(b-a, (b-a)f'((1-t)a+tb)\bigr)$, which is never zero. Furthermore as sets $\alpha\bigl([0,1]\bigr) = \gamma\bigl([a,b]\bigr) = \{(x,y) \in \mathbb{R}^2 : x \in [a,b] \text{ and } f(x) = y\}$, which is just the graph of f.

The last example leads us to a definition.

Definition 9.2.5. Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a smooth path and $h\colon [c,d] \to [a,b]$ a continuously differentiable bijective function such that $h'(t) \neq 0$ for all $t \in [c,d]$. Then the composition $\gamma \circ h$ is called a *smooth reparametrization* of γ.

Let γ be a piecewise smooth path, and h be a piecewise smooth bijective function. Then the composition $\gamma \circ h$ is called a *piecewise smooth reparametrization* of γ.

If h is strictly increasing, then h is said to *preserve orientation*. If h does not preserve orientation, then h is said to *reverse orientation*.

A reparametrization is another path for the same set. That is, $(\gamma \circ h)\bigl([c,d]\bigr) = \gamma\bigl([a,b]\bigr)$.

Let us remark that for h, piecewise smooth means that there is some partition $t_0 = c < t_1 < t_2 < \cdots < t_k = d$, such that $h|_{[t_{j-1},t_j]}$ is continuously differentiable and $(h|_{[t_{j-1},t_j]})'(t) \neq 0$ for all $t \in [t_{j-1}, t_j]$. Since h is bijective, it is either strictly increasing or strictly decreasing. Therefore either $(h|_{[t_{j-1},t_j]})'(t) > 0$ for all t or $(h|_{[t_{j-1},t_j]})'(t) < 0$ for all t.

Proposition 9.2.6. *If $\gamma\colon [a,b] \to \mathbb{R}^n$ is a piecewise smooth path, and $\gamma \circ h\colon [c,d] \to \mathbb{R}^n$ is a piecewise smooth reparametrization, then $\gamma \circ h$ is a piecewise smooth path.*

Proof. Let us assume that h preserves orientation, that is, h is strictly increasing. If $h\colon [c,d] \to [a,b]$ gives a piecewise smooth reparametrization then for some partition $r_0 = c < r_1 < r_2 < \cdots < r_\ell = d$, we have $h|_{[t_{j-1},t_j]}$ is continuously differentiable with positive derivative.

Let $t_0 = a < t_1 < t_2 < \cdots < t_k = b$ be the partition from the definition of piecewise smooth for γ together with the points $\{h(r_0), h(r_1), h(r_2), \ldots, h(r_\ell)\}$. Let $s_j := h^{-1}(t_j)$. Then $s_0 = c < s_1 < s_2 < \cdots < s_k = d$. For $t \in [s_{j-1}, s_j]$ notice that $h(t) \in [t_{j-1}, t_j]$, $h|_{[s_{j-1},s_j]}$ is continuously differentiable, and $\varphi|_{[t_{j-1},t_j]}$ is also continuously differentiable. Then

$$(\gamma \circ h)|_{[s_{j-1},s_j]}(t) = \gamma|_{[t_{j-1},t_j]}\big(h|_{[s_{j-1},s_j]}(t)\big).$$

The function $(\gamma \circ h)|_{[s_{j-1},s_j]}$ is therefore continuously differentiable and by the chain rule

$$\big((\gamma \circ h)|_{[s_{j-1},s_j]}\big)'(t) = \big(\gamma|_{[t_{j-1},t_j]}\big)'\big(h(t)\big)\big(h|_{[s_{j-1},s_j]}\big)'(t) \neq 0.$$

Therefore $\gamma \circ h$ is a piecewise smooth path. The case for an orientation reversing h is left as an exercise. □

If two paths are simple and their images are the same, it is left as an exercise that there exists a reparametrization.

9.2.2 Path integral of a one-form

Definition 9.2.7. If $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ are our coordinates, and given n real-valued continuous functions $f_1, f_2, \ldots, f_n$ defined on some set $S \subset \mathbb{R}^n$ we define a so-called *one-form*:

$$\omega = \omega_1 dx_1 + \omega_2 dx_2 + \cdots \omega_n dx_n.$$

We could represent ω as a continuous function from S to $\mathbb{R}^n$, although it is better to think of it as a different object.

Example 9.2.8: For example,

$$\omega(x,y) = \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy$$

is a one-form defined on $\mathbb{R}^2 \setminus \{(0,0)\}$.

Definition 9.2.9. Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a smooth path and

$$\omega = \omega_1 dx_1 + \omega_2 dx_2 + \cdots \omega_n dx_n,$$

a one-form defined on the direct image $\gamma\big([a,b]\big)$. Let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ be the components of γ. Define:

$$\begin{aligned}\int_\gamma \omega &:= \int_a^b \Big(\omega_1\big(\gamma(t)\big)\gamma_1'(t) + \omega_2\big(\gamma(t)\big)\gamma_2'(t) + \cdots + \omega_n\big(\gamma(t)\big)\gamma_n'(t)\Big)\,dt \\ &= \int_a^b \left(\sum_{j=1}^n \omega_j\big(\gamma(t)\big)\gamma_j'(t)\right)dt.\end{aligned}$$

If γ is piecewise smooth, take the corresponding partition $t_0 = a < t_1 < t_2 < \ldots < t_k = b$, where we assume the partition is the minimal one, that is γ is not differentiable at $t_2, t_3, \ldots, t_{k-1}$. Each $\gamma|_{[t_{j-1},t_j]}$ is a smooth path and we define

$$\int_\gamma \omega := \int_{\gamma|_{[t_0,t_1]}} \omega + \int_{\gamma|_{[t_1,t_2]}} \omega + \cdots + \int_{\gamma|_{[t_{n-1},t_n]}} \omega.$$

The notation makes sense from the formula you remember from calculus, let us state it somewhat informally: if $x_j(t) = \gamma_j(t)$, then $dx_j = \gamma_j'(t)dt$.

Paths can be cut up or concatenated as follows. The proof is a direct application of the additivity of the Riemann integral, and is left as an exercise. The proposition also justifies why we defined the integral over a piecewise smooth path in the way we did, and it further justifies that we may as well have taken any partition not just the minimal one in the definition.

Proposition 9.2.10. *Let $\gamma\colon [a,c] \to \mathbb{R}^n$ be a piecewise smooth path. For some $b \in (a,c)$, define the piecewise smooth paths $\alpha = \gamma|_{[a,b]}$ and $\beta = \gamma|_{[b,c]}$. For a one-form ω defined on the image of γ we have*

$$\int_\gamma \omega = \int_\alpha \omega + \int_\beta \omega.$$

Example 9.2.11: Let the one-form ω and the path $\gamma\colon [0,2\pi] \to \mathbb{R}^2$ be defined by

$$\omega(x,y) := \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy, \qquad \gamma(t) := \big(\cos(t), \sin(t)\big).$$

Then

$$\begin{aligned}\int_\gamma \omega &= \int_0^{2\pi} \left(\frac{-\sin(t)}{\big(\cos(t)\big)^2 + \big(\sin(t)\big)^2}\big(-\sin(t)\big) + \frac{\cos(t)}{\big(\cos(t)\big)^2 + \big(\sin(t)\big)^2}\big(\cos(t)\big)\right)dt \\ &= \int_0^{2\pi} 1\,dt = 2\pi.\end{aligned}$$

Next, let us parametrize the same curve as $\alpha\colon [0,1] \to \mathbb{R}^2$ defined by $\alpha(t) := \big(\cos(2\pi t), \sin(2\pi t)\big)$,

that is α is a smooth reparametrization of γ. Then

$$\begin{aligned}\int_\alpha \omega &= \int_0^1 \left(\frac{-\sin(2\pi t)}{\big(\cos(2\pi t)\big)^2 + \big(\sin(2\pi t)\big)^2} \big(-2\pi \sin(2\pi t)\big) \right. \\ &\qquad \left. + \frac{\cos(2\pi t)}{\big(\cos(2\pi t)\big)^2 + \big(\sin(2\pi t)\big)^2} \big(2\pi \cos(2\pi t)\big) \right) dt \\ &= \int_0^1 2\pi \, dt = 2\pi.\end{aligned}$$

Now let us reparametrize with $\beta\colon [0,2\pi] \to \mathbb{R}^2$ as $\beta(t) := \big(\cos(-t), \sin(-t)\big)$. Then

$$\begin{aligned}\int_\beta \omega &= \int_0^{2\pi} \left(\frac{-\sin(-t)}{\big(\cos(-t)\big)^2 + \big(\sin(-t)\big)^2} \big(\sin(-t)\big) + \frac{\cos(-t)}{\big(\cos(-t)\big)^2 + \big(\sin(-t)\big)^2} \big(-\cos(-t)\big) \right) dt \\ &= \int_0^{2\pi} (-1) \, dt = -2\pi.\end{aligned}$$

Now, α was an orientation preserving reparametrization of γ, and the integral was the same. On the other hand β is an orientation reversing reparametrization and the integral was minus the original.

The previous example is not a fluke. The path integral does not depend on the parametrization of the curve, the only thing that matters is the direction in which the curve is traversed.

Proposition 9.2.12. *Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a piecewise smooth path and $\gamma \circ h\colon [c,d] \to \mathbb{R}^n$ a piecewise smooth reparametrization. Suppose ω is a one-form defined on the set $\gamma\big([a,b]\big)$. Then*

$$\int_{\gamma \circ h} \omega = \begin{cases} \int_\gamma \omega & \text{if } h \text{ preserves orientation,} \\ -\int_\gamma \omega & \text{if } h \text{ reverses orientation.} \end{cases}$$

Proof. Assume first that γ and h are both smooth. Write the one form as $\omega = \omega_1 dx_1 + \omega_2 dx_2 + \cdots + \omega_n dx_n$. Suppose first that h is orientation preserving. Using the definition of the path integral and the change of variables formula for the Riemann integral,

$$\begin{aligned}\int_\gamma \omega &= \int_a^b \left(\sum_{j=1}^n \omega_j\big(\gamma(t)\big) \gamma_j'(t) \right) dt \\ &= \int_c^d \left(\sum_{j=1}^n \omega_j\Big(\gamma\big(h(\tau)\big)\Big) \gamma_j'\big(h(\tau)\big) \right) h'(\tau) \, d\tau \\ &= \int_c^d \left(\sum_{j=1}^n \omega_j\Big(\gamma\big(h(\tau)\big)\Big) (\gamma_j \circ h)'(\tau) \right) d\tau = \int_{\gamma \circ h} \omega.\end{aligned}$$

If h is orientation reversing it will swap the order of the limits on the integral introducing a minus sign. The details, along with finishing the proof for piecewise smooth paths is left to the reader as Exercise 9.2.4. $\square$

Due to this proposition (and the exercises), if we have a set $\Gamma \subset \mathbb{R}^n$ that is the image of a simple piecewise smooth path $\gamma\big([a,b]\big)$, then if we somehow indicate the orientation, that is, which direction we traverse the curve, in other words where we start and where we finish. Then we just write

$$\int_\Gamma \omega,$$

without mentioning the specific γ. Furthermore, for a simple closed path, it does not even matter where we start the parametrization. See the exercises.

Recall that *simple* means that γ restricted to (a,b) is one-to-one, that is, it is one-to-one except perhaps at the endpoints. We also often relax the simple path condition a little bit. For example, as long as $\gamma\colon [a,b] \to \mathbb{R}^n$ is one-to-one except at finitely many points. That is, there are only finitely many points $p \in \mathbb{R}^n$ such that $\gamma^{-1}(p)$ is more than one point. See the exercises. The issue about the injectivity problem is illustrated by the following example.

Example 9.2.13: Suppose $\gamma\colon [0,2\pi] \to \mathbb{R}^2$ is given by $\gamma(t) := \big(\cos(t), \sin(t)\big)$ and $\beta\colon [0,2\pi] \to \mathbb{R}^2$ is given by $\beta(t) := \big(\cos(2t), \sin(2t)\big)$. Notice that $\gamma\big([0,2\pi]\big) = \beta\big([0,2\pi]\big)$, and we travel around the same curve, the unit circle. But γ goes around the unit circle once in the counter clockwise direction, and β goes around the unit circle twice (in the same direction). Then

$$\begin{aligned}
\int_\gamma -y\,dx + x\,dy &= \int_0^{2\pi} \Big(\big(-\sin(t)\big)\big(-\sin(t)\big) + \cos(t)\cos(t) \Big) dt = 2\pi, \\
\int_\beta -y\,dx + x\,dy &= \int_0^{2\pi} \Big(\big(-\sin(2t)\big)\big(-2\sin(2t)\big) + \cos(t)\big(2\cos(t)\big) \Big) dt = 4\pi.
\end{aligned}$$

It is sometimes convenient to define a path integral over $\gamma\colon [a,b] \to \mathbb{R}^n$ that is not a path. We define

$$\int_\gamma \omega := \int_a^b \left(\sum_{j=1}^n \omega_j\big(\gamma(t)\big)\gamma_j'(t) \right) dt$$

for any γ which is continuously differentiable. A case which comes up naturally is when γ is constant. In this case $\gamma'(t) = 0$ for all t and $\gamma\big([a,b]\big)$ is a single point, which we regard as a "curve" of length zero. Then, $\int_\gamma \omega = 0$.

9.2.3 Line integral of a function

Sometimes we wish to simply integrate a function against the so-called arc-length measure.

Definition 9.2.14. Suppose $\gamma\colon [a,b] \to \mathbb{R}^n$ is a smooth path, and f is a continuous function defined on the image $\gamma\big([a,b]\big)$. Then define

$$\int_\gamma f\,ds := \int_a^b f\big(\gamma(t)\big)\|\gamma'(t)\|\,dt.$$

The definition for a piecewise smooth path is similar as before and is left to the reader.

The geometric idea of this integral is to find the "area under the graph of a function" as we move around the path γ. The line integral of a function is also independent of the parametrization, and in this case, the orientation does not matter.

Proposition 9.2.15. *Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a piecewise smooth path and $\gamma \circ h\colon [c,d] \to \mathbb{R}^n$ a piecewise smooth reparametrization. Suppose f is a continuous function defined on the set $\gamma\big([a,b]\big)$. Then*

$$\int_{\gamma\circ h} f\,ds = \int_\gamma f\,ds.$$

Proof. Suppose first that h is orientation preserving and γ and h are both smooth. Then as before

$$\begin{aligned}
\int_\gamma f\,ds &= \int_a^b f(t)\|\gamma'(t)\|\,dt \\
&= \int_c^d f\big(h(\tau)\big)\|\gamma'\big(h(\tau)\big)\|h'(\tau)\,d\tau \\
&= \int_c^d f\big(h(\tau)\big)\|\gamma'\big(h(\tau)\big)h'(\tau)\|\,d\tau \\
&= \int_c^d f\big(h(\tau)\big)\|(\gamma\circ h)'(\tau)\|\,d\tau \\
&= \int_{\gamma\circ h} f\,ds.
\end{aligned}$$

If h is orientation reversing it will swap the order of the limits on the integral but you also have to introduce a minus sign in order to take h' inside the norm. The details, along with finishing the proof for piecewise smooth paths is left to the reader as Exercise 9.2.5. □

Similarly as before, because of this proposition (and the exercises), if γ is simple, it does not matter which parametrization we use. Therefore, if $\Gamma = \gamma\big([a,b]\big)$ we can simply write

$$\int_\Gamma f\,ds.$$

In this case we also do not need to worry about orientation, either way we get the same thing.

Example 9.2.16: Let $f(x,y) = x$. Let $C \subset \mathbb{R}^2$ be half of the unit circle for $x \geq 0$. We wish to compute

$$\int_C f\,ds.$$

Parametrize C by $\gamma\colon [-\pi/2, \pi/2] \to \mathbb{R}^2$ given by $\gamma(t) = \bigl(\cos(t), \sin(t)\bigr)$. Then $\gamma'(t) = \bigl(-\sin(t), \cos(t)\bigr)$, and

$$\int_C f\,ds = \int_\gamma f\,ds = \int_{-\pi/2}^{\pi/2} \cos(t)\sqrt{\bigl(-\sin(t)\bigr)^2 + \bigl(\cos(t)\bigr)^2}\,dt = \int_{-\pi/2}^{\pi/2} \cos(t)\,dt = 2.$$

Definition 9.2.17. Suppose $\Gamma \subset \mathbb{R}^n$ is parametrized by a simple piecewise smooth path $\gamma\colon [a,b] \to \mathbb{R}^n$, that is $\gamma\bigl([a,b]\bigr) = \Gamma$. The we define the *length* by

$$\ell(\Gamma) := \int_\Gamma ds = \int_\gamma ds = \int_a^b \|\gamma'(t)\|\,dt.$$

Example 9.2.18: Let $x, y \in \mathbb{R}^n$ be two points and write $[x,y]$ as the straight line segment between the two points x and y. We parametrize $[x,y]$ by $\gamma(t) := (1-t)x + ty$ for t running between 0 and 1. We find $\gamma'(t) = y - x$ and therefore

$$\ell\bigl([x,y]\bigr) = \int_{[x,y]} ds = \int_0^1 \|y-x\|\,dt = \|y-x\|.$$

So the length of $[x,y]$ is the distance between x and y in the euclidean metric.

A simple piecewise smooth path $\gamma\colon [0,r] \to \mathbb{R}^n$ is said to be an *arc-length parametrization* if

$$\ell\bigl(\gamma\bigl([0,t]\bigr)\bigr) = \int_0^t \|\gamma'(\tau)\|\,d\tau = t.$$

You can think of such a parametrization as moving around your curve at speed 1.

9.2.4 Exercises

Exercise* 9.2.1*: *Show that if $\varphi\colon [a,b] \to \mathbb{R}^n$ is piecewise smooth as we defined it, then φ is a continuous function.*

Exercise* 9.2.2*: *Finish the proof of Proposition 9.2.6 for orientation reversing reparametrizations.*

Exercise* 9.2.3*: *Prove Proposition 9.2.10.*

Exercise* 9.2.4*: *Finish the proof of Proposition 9.2.12 for a) orientation reversing reparametrizations, and b) piecewise smooth paths and reparametrizations.*

***Exercise* 9.2.5:** *Finish the proof of Proposition 9.2.15 for a) orientation reversing reparametrizations, and b) piecewise smooth paths and reparametrizations.*

***Exercise* 9.2.6:** *Suppose $\gamma\colon [a,b] \to \mathbb{R}^n$ is a piecewise smooth path, and f is a continuous function defined on the image $\gamma\big([a,b]\big)$. Provide a definition of $\int_\gamma f\,ds$.*

***Exercise* 9.2.7:** *Directly using the definitions:*
a) Compute the arc-length of the unit square from Example 9.2.2 using the given parametrization.
b) Compute the arc-length of the unit circle using the parametrization $\gamma\colon [0,1] \to \mathbb{R}^2$, $\gamma(t) = \big(\cos(2\pi t), \sin(2\pi t)\big)$.
c) Compute the arc-length of the unit circle using the parametrization $\beta\colon [0,2\pi] \to \mathbb{R}^2$, $\beta(t) = \big(\cos(t), \sin(t)\big)$.

***Exercise* 9.2.8:** *Suppose $\gamma\colon [0,1] \to \mathbb{R}^n$ is a smooth path, and ω is a one-form defined on the image $\gamma\big([a,b]\big)$. For $r \in [0,1]$, let $\gamma_r\colon [0,r] \to \mathbb{R}^n$ be defined as simply the restriction of γ to $[0,r]$. Show that the function $h(r) := \int_{\gamma_r} \omega$ is a continuously differentiable function on $[0,1]$.*

***Exercise* 9.2.9:** *Suppose $\gamma\colon [a,b] \to \mathbb{R}^n$ is a smooth path. Show that there exists an $\varepsilon > 0$ and a smooth function $\tilde{\gamma}\colon (a-\varepsilon, b+\varepsilon) \to \mathbb{R}^n$ with $\tilde{\gamma}(t) = \gamma(t)$ for all $t \in [a,b]$ and $\tilde{\gamma}'(t) \neq 0$ for all $t \in (a-\varepsilon, b+\varepsilon)$. That is, prove that a smooth path always extends some small distance past the end points.*

***Exercise* 9.2.10:** *Suppose $\alpha\colon [a,b] \to \mathbb{R}^n$ and $\beta\colon [c,d] \to \mathbb{R}^n$ are piecewise smooth paths such that $\Gamma := \alpha\big([a,b]\big) = \beta\big([c,d]\big)$. Show that there exist finitely many points $\{p_1, p_2, \ldots, p_k\} \in \Gamma$, such that the sets $\alpha^{-1}\big(\{p_1, p_2, \ldots, p_k\}\big)$ and $\beta^{-1}\big(\{p_1, p_2, \ldots, p_k\}\big)$ are partitions of $[a,b]$ and $[c,d]$, such that on any subinterval the paths are smooth (that is, they are partitions as in the definition of piecewise smooth path).*

***Exercise* 9.2.11:** *a) Suppose $\gamma\colon [a,b] \to \mathbb{R}^n$ and $\alpha\colon [c,d] \to \mathbb{R}^n$ are two smooth paths which are one-to-one and $\gamma\big([a,b]\big) = \alpha\big([c,d]\big)$. Then there exists a smooth reparametrization $h\colon [a,b] \to [c,d]$ such that $\gamma = \alpha \circ h$. Hint: It should be not hard to find some h. The trick is to show it is continuously differentiable with a nonvanishing derivative. You will want to apply the implicit function theorem and it may at first seem the dimensions don't seem to work out.*
b) Prove the same thing as part a, but now for simple closed paths with the further assumption that $\gamma(a) = \gamma(b) = \alpha(c) = \alpha(d)$.
c) Prove parts a) and b) but for piecewise smooth paths, obtaining piecewise smooth reparametrizations. Hint: The trick is to find two partitions such that when restricted to a subinterval of the partition both paths have the same image and are smooth, see the above exercise.

***Exercise* 9.2.12:** *Suppose $\alpha\colon [a,b] \to \mathbb{R}^n$ and $\beta\colon [b,c] \to \mathbb{R}^n$ are piecewise smooth paths with $\alpha(b) = \beta(b)$. Let $\gamma\colon [a,c] \to \mathbb{R}^n$ be defined by*

$$\gamma(t) := \begin{cases} \alpha(t) & \text{if } t \in [a,b], \\ \beta(t) & \text{if } t \in (b,c]. \end{cases}$$

Show that γ is a piecewise smooth path, and that if ω is a one-form defined on the curve given by γ, then

$$\int_\gamma \omega = \int_\alpha \omega + \int_\beta \omega.$$

***Exercise* 9.2.13:** *Suppose $\gamma\colon [a,b] \to \mathbb{R}^n$ and $\beta\colon [c,d] \to \mathbb{R}^n$ are two simple piecewise smooth closed paths. That is $\gamma(a) = \gamma(b)$ and $\beta(c) = \beta(d)$ and the restrictions $\gamma|_{(a,b)}$ and $\beta|_{(c,d)}$ are one-to-one. Suppose $\Gamma = \gamma\big([a,b]\big) = \beta\big([c,d]\big)$ and ω is a one-form defined on $\Gamma \subset \mathbb{R}^n$. Show that either*

$$\int_\gamma \omega = \int_\beta \omega, \qquad \text{or} \qquad \int_\gamma \omega = -\int_\beta \omega.$$

In particular, the notation $\int_\Gamma \omega$ makes sense if we indicate the direction in which the integral is evaluated. Hint: see previous three exercises.

***Exercise* 9.2.14:** *Suppose $\gamma\colon [a,b] \to \mathbb{R}^n$ and $\beta\colon [c,d] \to \mathbb{R}^n$ are two piecewise smooth paths which are one-to-one except at finitely many points. That is, there is at most finitely many points $p \in \mathbb{R}^n$ such that $\gamma^{-1}(p)$ or $\beta^{-1}(p)$ contains more than one point. Suppose $\Gamma = \gamma\big([a,b]\big) = \beta\big([c,d]\big)$ and ω is a one-form defined on $\Gamma \subset \mathbb{R}^n$. Show that either*

$$\int_\gamma \omega = \int_\beta \omega, \qquad \text{or} \qquad \int_\gamma \omega = -\int_\beta \omega.$$

In particular, the notation $\int_\Gamma \omega$ makes sense if we indicate the direction in which the integral is evaluated. Hint: same hint as the last exercise.

***Exercise* 9.2.15:** *Define $\gamma\colon [0,1] \to \mathbb{R}^2$ by $\gamma(t) := \Big(t^3 \sin(1/t), t\big(3t^2 \sin(1/t) - t\cos(1/t)\big)^2\Big)$ for $t \neq 0$ and $\gamma(0) = (0,0)$. Show that:*
a) γ is continuously differentiable on $[0,1]$.
b) Show that there exists an infinite sequence $\{t_n\}$ in $[0,1]$ converging to 0, such that $\gamma'(t_n) = (0,0)$.
c) Show that the points $\gamma(t_n)$ lie on the line $y = 0$ and such that the x-coordinate of $\gamma(t_n)$ alternates between positive and negative (if they do not alternate you only found a subsequence and you need to find them all).
d) Show that there is no piecewise smooth α whose image equals $\gamma\big([0,1]\big)$. Hint: look at part c) and show that α' must be zero where it reaches the origin.
e) (Computer) if you know a plotting software that allows you to plot parametric curves, make a plot of the curve, but only for t in the range $[0,0.1]$ otherwise you will not see the behavior. In particular you should notice that $\gamma\big([0,1]\big)$ has infinitely many "corners" near the origin.

9.3 Path independence

Note: 2 lectures

9.3.1 Path independent integrals

Let $U \subset \mathbb{R}^n$ be a set and ω a one-form defined on U, The integral of ω is said to be *path independent* if for any two points $x, y \in U$ and any two piecewise smooth paths $\gamma\colon [a,b] \to U$ and $\beta\colon [c,d] \to U$ such that $\gamma(a) = \beta(c) = x$ and $\gamma(b) = \beta(d) = y$ we have

$$\int_\gamma \omega = \int_\beta \omega.$$

In this case we simply write

$$\int_x^y \omega := \int_\gamma \omega = \int_\beta \omega.$$

Not every one-form gives a path independent integral. In fact, most do not.

Example 9.3.1: Let $\gamma\colon [0,1] \to \mathbb{R}^2$ be the path $\gamma(t) = (t,0)$ going from $(0,0)$ to $(1,0)$. Let $\beta\colon [0,1] \to \mathbb{R}^2$ be the path $\beta(t) = \bigl(t,(1-t)t\bigr)$ also going between the same points. Then

$$\begin{aligned}
\int_\gamma y\,dx &= \int_0^1 \gamma_2(t)\gamma_1'(t)\,dt = \int_0^1 0(1)\,dt = 0, \\
\int_\beta y\,dx &= \int_0^1 \beta_2(t)\beta_1'(t)\,dt = \int_0^1 (1-t)t(1)\,dt = \frac{1}{6}.
\end{aligned}$$

So the integral of $y\,dx$ is not path independent. In particular, $\int_{(0,0)}^{(1,0)} y\,dx$ does not make sense.

Definition 9.3.2. Let $U \subset \mathbb{R}^n$ be an open set and $f\colon U \to \mathbb{R}$ a continuously differentiable function. Then the one-form

$$df := \frac{\partial f}{\partial x_1}\,dx_1 + \frac{\partial f}{\partial x_2}\,dx_2 + \cdots + \frac{\partial f}{\partial x_n}\,dx_n$$

is called the *total derivative* of f.

An open set $U \subset \mathbb{R}^n$ is said to be *path connected** if for every two points x and y in U, there exists a piecewise smooth path starting at x and ending at y.

We will leave as an exercise that every connected open set is path connected.

*Normally only a continuous path is used in this definition, but for open sets two two definitions are equivalent. See the exercises.

Proposition 9.3.3. *Let $U \subset \mathbb{R}^n$ be a path connected open set and ω a one-form defined on U. Then*

$$\int_x^y \omega$$

is path independent (for all $x, y \in U$) if and only if there exists a continuously differentiable $f \colon U \to \mathbb{R}$ such that $\omega = df$.

In fact, if such an f exists, then for any two points $x, y \in U$

$$\int_x^y \omega = f(y) - f(x).$$

In other words if we fix $p \in U$, then $f(x) = C + \int_p^x \omega$.

Proof. First suppose that the integral is path independent. Pick $p \in U$ and define

$$f(x) := \int_p^x \omega.$$

Write $\omega = \omega_1 dx_1 + \omega_2 dx_2 + \cdots + \omega_n dx_n$. We wish to show that for every $j = 1, 2, \ldots, n$, the partial derivative $\frac{\partial f}{\partial x_j}$ exists and is equal to ω_j.

Let e_j be an arbitrary standard basis vector. Compute

$$\frac{f(x + he_j) - f(x)}{h} = \frac{1}{h}\left(\int_p^{x+he_j} \omega - \int_p^x \omega\right) = \frac{1}{h}\int_x^{x+he_j} \omega,$$

which follows by Proposition 9.2.10 and path indepdendence as $\int_p^{x+he_j} \omega = \int_p^x \omega + \int_x^{x+he_j} \omega$, because we could have picked a path from p to $x + he_j$ that also happens to pass through x, and then cut this path in two.

Since U is open, suppose h is so small so that all points of distance $|h|$ or less from x are in U. As the integral is path independent, pick the simplest path possible from x to $x + he_j$, that is $\gamma(t) = x + the_j$ for $t \in [0, 1]$. The path is in U. Notice $\gamma'(t) = he_j$ has only one nonzero component and that is the jth component, which is h. Therefore

$$\frac{1}{h}\int_x^{x+he_j} \omega = \frac{1}{h}\int_\gamma \omega = \frac{1}{h}\int_0^1 \omega_j(x + the_j)h\,dt = \int_0^1 \omega_j(x + the_j)\,dt.$$

We wish to take the limit as $h \to 0$. The function ω_j is continuous. So given $\varepsilon > 0$, h can be small enough so that $|\omega(x) - \omega(y)| < \varepsilon$, whenever $\|x - y\| \leq |h|$. Therefore, $\left|\omega_j(x + the_j) - \omega_j(x)\right| < \varepsilon$ for all $t \in [0, 1]$, and we estimate

$$\left|\int_0^1 \omega_j(x + the_j)\,dt - \omega(x)\right| = \left|\int_0^1 \big(\omega_j(x + the_j) - \omega(x)\big)\,dt\right| \leq \varepsilon.$$

That is,

$$\lim_{h\to 0} \frac{f(x+he_j)-f(x)}{h} = \omega_j(x),$$

which is what we wanted that is $df = \omega$. As ω_j are continuous for all j, we find that f has continuous partial derivatives and therefore is continuously differentiable.

For the other direction suppose f exists such that $df = \omega$. Suppose we take a smooth path $\gamma\colon [a,b] \to U$ such that $\gamma(a) = x$ and $\gamma(b) = y$, then

$$\begin{aligned}\int_\gamma df &= \int_a^b \left(\frac{\partial f}{\partial x_1}\big(\gamma(t)\big)\gamma_1'(t) + \frac{\partial f}{\partial x_2}\big(\gamma(t)\big)\gamma_2'(t) + \cdots + \frac{\partial f}{\partial x_n}\big(\gamma(t)\big)\gamma_n'(t) \right) dt \\ &= \int_a^b \frac{d}{dt}\left[f\big(\gamma(t)\big)\right] dt \\ &= f(y) - f(x).\end{aligned}$$

The value of the integral only depends on x and y, not the path taken. Therefore the integral is path independent. We leave checking this for a piecewise smooth path as an exercise to the reader. □

Proposition 9.3.4. *Let $U \subset \mathbb{R}^n$ be a path connected open set and ω a 1-form defined on U. Then $\omega = df$ for some continuously differentiable $f\colon U \to \mathbb{R}$ if and only if*

$$\int_\gamma \omega = 0$$

for every piecewise smooth closed path $\gamma\colon [a,b] \to U$.

Proof. Suppose first that $\omega = df$ and let γ be a piecewise smooth closed path. Then we from above we have that

$$\int_\gamma \omega = f\big(\gamma(b)\big) - f\big(\gamma(a)\big) = 0,$$

because $\gamma(a) = \gamma(b)$ for a closed path.

Now suppose that for every piecewise smooth closed path γ, $\int_\gamma \omega = 0$. Let x, y be two points in U and let $\alpha\colon [0,1] \to U$ and $\beta\colon [0,1] \to U$ be two piecewise smooth paths with $\alpha(0) = \beta(0) = x$ and $\alpha(1) = \beta(1) = y$. Then let $\gamma\colon [0,2] \to U$ be defined by

$$\gamma(t) := \begin{cases} \alpha(t) & \text{if } t \in [0,1], \\ \beta(2-t) & \text{if } t \in (1,2]. \end{cases}$$

This is a piecewise smooth closed path and so

$$0 = \int_\gamma \omega = \int_\alpha \omega - \int_\beta \omega.$$

This follows first by Proposition 9.2.10, and then noticing that the second part is β travelled backwards so that we get minus the β integral. Thus the integral of ω on U is path independent. □

There is a local criterion, a differential equation, that guarantees path independence. That is, under the right condition there exists an *antiderivative* f whose total derivative is the given one form ω. However, since the criterion is local, we only get the result locally. We can define the antiderivative in any so-called *simply connected* domain, which informally is a domain where any path between two points can be "continuously deformed" into any other path between those two points. To make matters simple, the usual way this result is proved is for so-called star-shaped domains.

Definition 9.3.5. Let $U \subset \mathbb{R}^n$ be an open set and $p \in U$. We say U is a *star shaped domain* with respect to p if for any other point $x \in U$, the line segment between p and x is in U, that is, if $(1-t)p + tx \in U$ for all $t \in [0,1]$. If we say simply *star shaped* then U is star shaped with respect to some $p \in U$.

Notice the difference between star shaped and convex. A convex domain is star shaped, but a star shaped domain need not be convex.

Theorem 9.3.6 (Poincarè lemma)**.** *Let $U \subset \mathbb{R}^n$ be a star shaped domain and ω a continuously differentiable one-form defined on U. That is, if*

$$\omega = \omega_1 dx_1 + \omega_2 dx_2 + \cdots + \omega_n dx_n,$$

then $\omega_1, \omega_2, \ldots, \omega_n$ are continuously differentiable functions. Suppose that for every j and k

$$\frac{\partial \omega_j}{\partial x_k} = \frac{\partial \omega_k}{\partial x_j},$$

then there exists a twice continuously differentiable function $f \colon U \to \mathbb{R}$ such that $df = \omega$.

The condition on the derivatives of ω is precisely the condition that the second partial derivatives commute. That is, if $df = \omega$, and f is twice continuously differentiable then

$$\frac{\partial \omega_j}{\partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial \omega_k}{\partial x_j}.$$

The condition is therefore clearly necessary. The lemma says that it is sufficient for a star shaped U.

Proof. Suppose U is star shaped with respect to $y = (y_1, y_2, \ldots, y_n) \in U$.

Given $x = (x_1, x_2, \ldots, x_n) \in U$, define the path $\gamma \colon [0,1] \to U$ as $\gamma(t) := (1-t)y + tx$, so $\gamma'(t) = x - y$. Then let

$$f(x) := \int_\gamma \omega = \int_0^1 \left(\sum_{k=1}^n \omega_k\big((1-t)y + tx\big)(x_k - y_k) \right) dt.$$

We differentiate in x_j under the integral. We can do that since everything, including the partials themselves are continuous.

$$\begin{aligned}
\frac{\partial f}{\partial x_j}(x) &= \int_0^1 \left(\left(\sum_{k=1}^n \frac{\partial \omega_k}{\partial x_j}\big((1-t)y+tx\big)t(x_k-y_k) \right) + \omega_j\big((1-t)y+tx\big) \right) dt \\
&= \int_0^1 \left(\left(\sum_{k=1}^n \frac{\partial \omega_j}{\partial x_k}\big((1-t)y+tx\big)t(x_k-y_k) \right) + \omega_j\big((1-t)y+tx\big) \right) dt \\
&= \int_0^1 \frac{d}{dt}\left[t\omega_j\big((1-t)y+tx\big) \right] dt \\
&= \omega_j(x).
\end{aligned}$$

And this is precisely what we wanted. □

Example 9.3.7: Without some hypothesis on U the theorem is not true. Let

$$\omega(x,y) := \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy$$

be defined on $\mathbb{R}^2 \setminus \{0\}$. It is easy to see that

$$\frac{\partial}{\partial y}\left[\frac{-y}{x^2+y^2}\right] = \frac{\partial}{\partial x}\left[\frac{x}{x^2+y^2}\right].$$

However, there is no $f\colon \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ such that $df = \omega$. We saw in if we integrate from $(1,0)$ to $(1,0)$ along the unit circle, that is $\gamma(t) = \big(\cos(t), \sin(t)\big)$ for $t \in [0, 2\pi]$ we got 2π and not 0 as it should be if the integral is path independent or in other words if there would exist an f such that $df = \omega$.

9.3.2 Vector fields

A common object to integrate is a so-called vector field. That is an assignment of a vector at each point of a domain.

Definition 9.3.8. Let $U \subset \mathbb{R}^n$ be a set. A continuous function $v\colon U \to \mathbb{R}^n$ is called a *vector field*. Write $v = (v_1, v_2, \ldots, v_n)$.

Given a smooth path $\gamma\colon [a,b] \to \mathbb{R}^n$ with $\gamma\big([a,b]\big) \subset U$ we define the path integral of the vectorfield v as

$$\int_\gamma v \cdot d\gamma := \int_a^b v\big(\gamma(t)\big) \cdot \gamma'(t)\, dt,$$

where the dot in the definition is the standard dot product. Again the definition of a piecewise smooth path is done by integrating over each smooth interval and adding the result.

If we unravel the definition we find that

$$\int_\gamma v \cdot d\gamma = \int_\gamma v_1 dx_1 + v_2 dx_2 + \cdots + v_n dx_n.$$

Therefore what we know about integration of one-forms carries over to the integration of vector fields. For example path independence for integration of vector fields is simply that

$$\int_x^y v \cdot d\gamma$$

is path independent (so for any γ) if and only if $v = \nabla f$, that is the gradient of a function. The function f is then called the *potential* for v.

A vector field v whose path integrals are path independent is called a *conservative vector field*. The naming comes from the fact that such vector fields arise in physical systems where a certain quantity, the energy is conserved.

9.3.3 Exercises

Exercise* 9.3.1: Find an $f\colon \mathbb{R}^2 \to \mathbb{R}$ such that $df = xe^{x^2+y^2}dx + ye^{x^2+y^2}dy$.*

Exercise* 9.3.2: Find an $\omega_2\colon \mathbb{R}^2 \to \mathbb{R}$ such that there exists a continuously differentiable $f\colon \mathbb{R}^2 \to \mathbb{R}$ for which $df = e^{xy}dx + \omega_2 dy$.*

Exercise* 9.3.3: Finish the proof of Proposition 9.3.3, that is, we only proved the second direction for a smooth path, not a piecewise smooth path.*

Exercise* 9.3.4: Show that a star shaped domain $U \subset \mathbb{R}^n$ is path connected.*

Exercise* 9.3.5: Show that $U := \mathbb{R}^2 \setminus \{(x,y) \in \mathbb{R}^2 : x \leq 0, y = 0\}$ is star shaped and find all points $(x_0, y_0) \in U$ such that U is star shaped with respect to (x_0, y_0).*

Exercise* 9.3.6: Suppose B_1 and B_2 are two open balls in $\mathbb{R}^n$ with a nonempty intersection. Suppose ω is a continuously differentiable one-form defined on $B_1 \cup B_2$ such that $\frac{\partial \omega_j}{\partial x_k} = \frac{\partial \omega_k}{\partial x_j}$ for all j and k. Prove that there exists a twice continuously differentiable $f\colon U \to \mathbb{R}$ such that $df = \omega$.*
Hint: notice that $B_1 \cap B_2$ is connected (it is convex).

Exercise* 9.3.7** (Hard): Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a simple nonclosed piecewise smooth path (so γ is one-to-one). Suppose ω is a continuously differentiable one-form defined on some open set V with $\gamma\big([a,b]\big) \subset V$ and $\frac{\partial \omega_j}{\partial x_k} = \frac{\partial \omega_k}{\partial x_j}$ for all j and k. Prove that there exists an open set U with $\gamma\big([a,b]\big) \subset U \subset V$ and a twice continuously differentiable function $f\colon U \to \mathbb{R}$ such that $df = \omega$.*
Hint 1: See previous exercise.
Hint 2: $\gamma\big([a,b]\big)$ is compact.
Hint 3: Show that you can cover the curve by finitely many balls in sequence so that the kth ball only intersects the $(k-1)$th ball.

Exercise* 9.3.8: a) Show that a connected open set is path connected. Hint: Start with two points x and y in a connected set U, and let $U_x \subset U$ is the set of points that are reachable by a path from x and similarly for U_y. Show that both sets are open, since they are nonempty ($x \in U_x$ and $y \in U_y$) it must be that $U_x = U_y = U$.*
b) Prove the converse that is, a path connected set $U \subset \mathbb{R}^n$ is connected. Hint: for contradiction assume there exist two open and disjoint nonempty open sets and then assume there is a piecewise smooth (and therefore continuous) path between a point in one to a point in the other.

Exercise* 9.3.9: Usually path connectedness is defined using just continuous paths rather than piecewise smooth paths. Prove that the definitions are equivalent, in other words prove the following statement:*
Suppose $U \subset \mathbb{R}^n$ is such that for any $x, y \in U$, there exists a continuous function $\gamma\colon [a,b] \to U$ such that $\gamma(a) = x$ and $\gamma(b) = y$. Then U is path connected (in other words, then there exists a piecewise smooth path).

Exercise* 9.3.10** (Hard):*** *Take*

$$\omega(x,y) = \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy$$

defined on $\mathbb{R}^2 \setminus \{(0,0)\}$. Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{(0,0)\}$ be a closed piecewise smooth path. Let $R := \{(x,y) \in \mathbb{R}^2 : x \leq 0$ and $y = 0\}$. Suppose $R \cap \gamma\big([a,b]\big)$ is a finite set of k points. Then

$$\int_\gamma \omega = 2\pi\ell$$

for some integer ℓ with $|\ell| \leq k$.
Hint 1: First prove that for a path β that starts and end on R but does not intersect it otherwise, you find that $\int_\beta \omega$ is -2π, 0, or 2π. Hint 2: You proved above that $\mathbb{R}^2 \setminus R$ is star shaped.
Note: The number ℓ is called the winding number *it it measures how many times does γ wind around the origin in the clockwise direction.*

Chapter 10

Multivariable integral

10.1 Riemann integral over rectangles

Note: 2–3 lectures

As in chapter chapter 5, we define the Riemann integral using the Darboux upper and lower integrals. The ideas in this section are very similar to integration in one dimension. The complication is mostly notational. The differences between one and several dimensions will grow more pronounced in the sections following.

10.1.1 Rectangles and partitions

Definition 10.1.1. Let $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ be such that $a_k \leq b_k$ for all k. A set of the form $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is called a *closed rectangle*. In this setting it is sometimes useful to allow $a_k = b_k$, in which case we think of $[a_k, b_k] = \{a_k\}$ as usual. If $a_k < b_k$ for all k, then a set of the form $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$ is called an *open rectangle*.

For an open or closed rectangle $R := [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ or $R := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n) \subset \mathbb{R}^n$, we define the *n-dimensional volume* by

$$V(R) := (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

A *partition* P of the closed rectangle $R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is a finite set of partitions $P_1, P_2, \ldots, P_n$ of the intervals $[a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n]$. We write $P = (P_1, P_2, \ldots, P_n)$. That is, for every k there is an integer ℓ_k and the finite set of numbers $P_k = \{x_{k,0}, x_{k,1}, x_{k,2}, \ldots, x_{k,\ell_k}\}$ such that

$$a_k = x_{k,0} < x_{k,1} < x_{k,2} < \cdots < x_{k,\ell_k - 1} < x_{k,\ell_k} = b_k.$$

Picking a set of n integers $j_1, j_2, \ldots, j_n$ where $j_k \in \{1, 2, \ldots, \ell_k\}$ we get the *subrectangle*

$$[x_{1,j_1-1}\ ,\ x_{1,j_1}] \times [x_{2,j_2-1}\ ,\ x_{2,j_2}] \times \cdots \times [x_{n,j_n-1}\ ,\ x_{n,j_n}].$$

For simplicity, we order the subrectangles somehow and we say $\{R_1, R_2, \ldots, R_N\}$ are the subrectangles corresponding to the partition P of R. More simply, we say they are the subrectangles of P. In other words, we subdivided the original rectangle into many smaller subrectangles. See Figure 10.1. It is not difficult to see that these subrectangles cover our original R, and their volume sums to that of R. That is,

$$R = \bigcup_{j=1}^{N} R_j, \qquad \text{and} \qquad V(R) = \sum_{j=1}^{N} V(R_j).$$

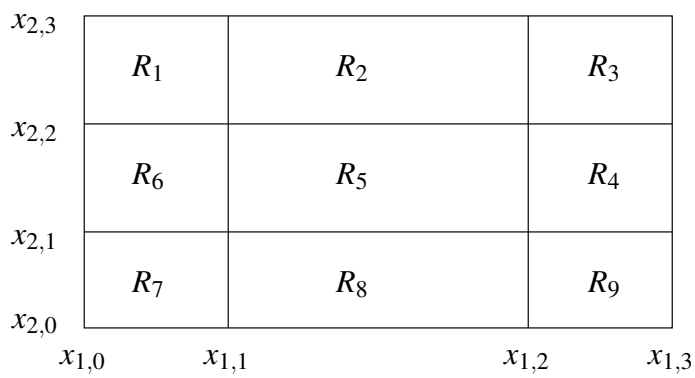

Figure 10.1: Example partition of a rectangle in $\mathbb{R}^2$. The order of the subrectangles is not important.

When

$$R_k = [x_{1,j_1-1}\ ,\ x_{1,j_1}] \times [x_{2,j_2-1}\ ,\ x_{2,j_2}] \times \cdots \times [x_{n,j_n-1}\ ,\ x_{n,j_n}]$$

then

$$V(R_k) = \Delta x_{1,j_1} \Delta x_{2,j_2} \cdots \Delta x_{n,j_n} = (x_{1,j_1} - x_{1,j_1-1})(x_{2,j_2} - x_{2,j_2-1}) \cdots (x_{n,j_n} - x_{n,j_n-1}).$$

Let $R \subset \mathbb{R}^n$ be a closed rectangle and let $f\colon R \to \mathbb{R}$ be a bounded function. Let P be a partition of $[a,b]$ and suppose that there are N subrectangles $R_1, R_2, \ldots, R_N$. Define

$$\begin{aligned}
m_i &:= \inf\{f(x) : x \in R_i\}, \\
M_i &:= \sup\{f(x) : x \in R_i\}, \\
L(P,f) &:= \sum_{i=1}^{N} m_i V(R_i), \\
U(P,f) &:= \sum_{i=1}^{N} M_i V(R_i).
\end{aligned}$$

We call $L(P,f)$ the *lower Darboux sum* and $U(P,f)$ the *upper Darboux sum*.

The indexing in the definition may be complicated, but fortunately we generally do not need to go back directly to the definition often. We start proving facts about the Darboux sums analogous to the one-variable results.

Proposition 10.1.2. *Suppose $R \subset \mathbb{R}^n$ is a closed rectangle and $f \colon R \to \mathbb{R}$ is a bounded function. Let $m, M \in \mathbb{R}$ be such that for all $x \in R$ we have $m \leq f(x) \leq M$. For any partition P of R we have*

$$mV(R) \leq L(P,f) \leq U(P,f) \leq MV(R).$$

Proof. Let P be a partition. Then for all i we have $m \leq m_i$ and $M_i \leq M$. Also $m_i \leq M_i$ for all i. Finally $\sum_{i=1}^{N} V(R_i) = V(R)$. Therefore,

$$mV(R) = m\left(\sum_{i=1}^{N} V(R_i)\right) = \sum_{i=1}^{N} mV(R_i) \leq \sum_{i=1}^{N} m_i V(R_i) \leq$$
$$\leq \sum_{i=1}^{N} M_i V(R_i) \leq \sum_{i=1}^{N} MV(R_i) = M\left(\sum_{i=1}^{N} V(R_i)\right) = MV(R). \quad \square$$

10.1.2 Upper and lower integrals

By Proposition 10.1.2 the set of upper and lower Darboux sums are bounded sets and we can take their infima and suprema. As before, we now make the following definition.

Definition 10.1.3. If $f \colon R \to \mathbb{R}$ is a bounded function on a closed rectangle $R \subset \mathbb{R}^n$. Define

$$\underline{\int_R} f := \sup\{L(P,f) : P \text{ a partition of } R\}, \qquad \overline{\int_R} f := \inf\{U(P,f) : P \text{ a partition of } R\}.$$

We call $\underline{\int}$ the *lower Darboux integral* and $\overline{\int}$ the *upper Darboux integral*.

As in one dimension we have refinements of partitions.

Definition 10.1.4. Let $R \subset \mathbb{R}^n$ be a closed rectangle. Let $P = (P_1, P_2, \ldots, P_n)$ and $\widetilde{P} = (\widetilde{P}_1, \widetilde{P}_2, \ldots, \widetilde{P}_n)$ be partitions of R. We say $\widetilde{P}$ a *refinement* of P if, as sets, $P_k \subset \widetilde{P}_k$ for all $k = 1, 2, \ldots, n$.

It is not difficult to see that if $\widetilde{P}$ is a refinement of P, then subrectangles of P are unions of subrectangles of $\widetilde{P}$. Simply put, in a refinement we take the subrectangles of P, and we cut them into smaller subrectangles. See Figure 10.2.

Proposition 10.1.5. *Suppose $R \subset \mathbb{R}^n$ is a closed rectangle, P is a partition of R and $\widetilde{P}$ is a refinement of P. If $f \colon R \to \mathbb{R}$ be a bounded function, then*

$$L(P,f) \leq L(\widetilde{P},f) \qquad \text{and} \qquad U(\widetilde{P},f) \leq U(P,f).$$

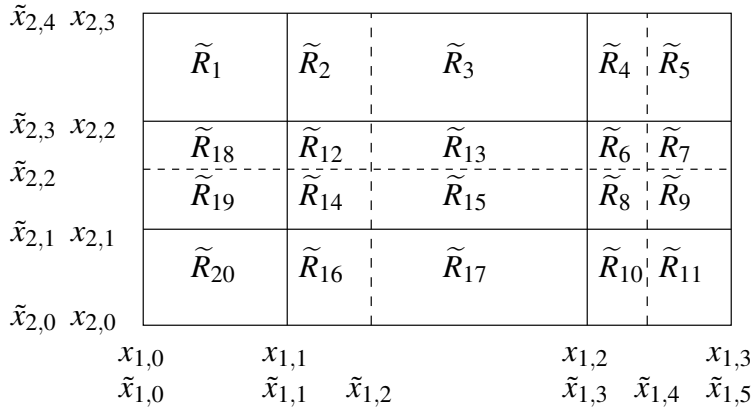

Figure 10.2: Example refinement of a partition. New "cuts" are marked in dashed lines. Do note that the exact order of the new subrectangles does not matter.

Proof. We prove the first inequality, the second follows similarly. Let $R_1, R_2, \ldots, R_N$ be the subrectangles of P and $\widetilde{R}_1, \widetilde{R}_2, \ldots, \widetilde{R}_{\widetilde{N}}$ be the subrectangles of $\widetilde{R}$. Let I_k be the set of all indices j such that $\widetilde{R}_j \subset R_k$. For example, using the exampels in figures 10.1 and 10.2, $I_4 = \{6,7,8,9\}$ and $R_4 = \widetilde{R}_6 \cup \widetilde{R}_7 \cup \widetilde{R}_8 \cup \widetilde{R}_9$. We notice in general that

$$R_k = \bigcup_{j \in I_k} \widetilde{R}_j, \qquad V(R_k) = \sum_{j \in I_k} V(\widetilde{R}_j).$$

Let $m_j := \inf\{f(x) : x \in R_j\}$, and $\widetilde{m}_j := \inf\{f(x) :\in \widetilde{R}_j\}$ as usual. Notice also that if $j \in I_k$, then $m_k \leq \widetilde{m}_j$. Then

$$L(P,f) = \sum_{k=1}^{N} m_k V(R_k) = \sum_{k=1}^{N} \sum_{j \in I_k} m_k V(\widetilde{R}_j) \leq \sum_{k=1}^{N} \sum_{j \in I_k} \widetilde{m}_j V(\widetilde{R}_j) = \sum_{j=1}^{\widetilde{N}} \widetilde{m}_j V(\widetilde{R}_j) = L(\widetilde{P}, f). \qquad \square$$

The key point of this next proposition is that the lower Darboux integral is less than or equal to the upper Darboux integral.

Proposition 10.1.6. *Let $R \subset \mathbb{R}^n$ be a closed rectangle and $f \colon R \to \mathbb{R}$ a bounded function. Let $m, M \in \mathbb{R}$ be such that for all $x \in R$ we have $m \leq f(x) \leq M$. Then*

$$mV(R) \leq \underline{\int_R} f \leq \overline{\int_R} f \leq MV(R). \tag{10.1}$$

Proof. For any partition P, via Proposition 10.1.2,

$$mV(R) \leq L(P,f) \leq U(P,f) \leq MV(R).$$

Taking suprema of $L(P,f)$ and infima of $U(P,f)$ over all P, we obtain the first and the last inequality.

The key inequality in (10.1) is the middle one. Let $P = (P_1, P_2, \ldots, P_n)$ and $Q = (Q_1, Q_2, \ldots, Q_n)$ be partitions of R. Define $\widetilde{P} = (\widetilde{P}_1, \widetilde{P}_2, \ldots, \widetilde{P}_n)$ by letting $\widetilde{P}_k = P_k \cup Q_k$. Then $\widetilde{P}$ is a partition of R as can easily be checked, and $\widetilde{P}$ is a refinement of P and a refinement of Q. By Proposition 10.1.5, $L(P,f) \leq L(\widetilde{P},f)$ and $U(\widetilde{P},f) \leq U(Q,f)$. Therefore,

$$L(P,f) \leq L(\widetilde{P},f) \leq U(\widetilde{P},f) \leq U(Q,f).$$

In other words, for two arbitrary partitions P and Q we have $L(P,f) \leq U(Q,f)$. Via Proposition **??** we obtain

$$\sup\{L(P,f) : P \text{ a partition of } R\} \leq \inf\{U(P,f) : P \text{ a partition of } R\}.$$

In other words $\underline{\int_R} f \leq \overline{\int_R} f$. □

10.1.3 The Riemann integral

We have all we need to define the Riemann integral in n-dimensions over rectangles. Again, the Riemann integral is only defined on a certain class of functions, called the Riemann integrable functions.

Definition 10.1.7. Let $R \subset \mathbb{R}^n$ be a closed rectangle. Let $f \colon R \to \mathbb{R}$ be a bounded function such that

$$\underline{\int_R} f(x)\, dx = \overline{\int_R} f(x)\, dx.$$

Then f is said to be *Riemann integrable*, and we sometimes say simply *integrable*. The set of Riemann integrable functions on R is denoted by $\mathscr{R}(R)$. When $f \in \mathscr{R}(R)$ we define the *Riemann integral*

$$\int_R f := \underline{\int_R} f = \overline{\int_R} f.$$

When the variable $x \in \mathbb{R}^n$ needs to be emphasized we write

$$\int_R f(x)\, dx, \qquad \int_R f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n, \qquad \text{or} \qquad \int_R f(x)\, dV.$$

If $R \subset \mathbb{R}^2$, then often instead of volume we say area, and hence write

$$\int_R f(x)\, dA.$$

Proposition 10.1.6 implies immediately the following proposition.

Proposition 10.1.8. *Let $f \colon R \to \mathbb{R}$ be a Riemann integrable function on a closed rectangle $R \subset \mathbb{R}^n$. Let $m, M \in \mathbb{R}$ be such that $m \leq f(x) \leq M$ for all $x \in R$. Then*

$$mV(R) \leq \int_R f \leq MV(R).$$

Example 10.1.9: A constant function is Riemann integrable. Suppose $f(x) = c$ for all x on R. Then

$$cV(R) \leq \underline{\int_R} f \leq \overline{\int_R} f \leq cV(R).$$

So f is integrable, and furthermore $\int_R f = cV(R)$.

The proofs of linearity and monotonicity are almost completely identical as the proofs from one variable. We therefore leave it as an exercise to prove the next two propositions.

Proposition 10.1.10 (Linearity)**.** *Let $R \subset \mathbb{R}^n$ be a closed rectangle and let f and g be in $\mathscr{R}(R)$ and $\alpha \in \mathbb{R}$.*

(i) αf is in $\mathscr{R}(R)$ and

$$\int_R \alpha f = \alpha \int_R f.$$

(ii) $f + g$ is in $\mathscr{R}(R)$ and

$$\int_R (f+g) = \int_R f + \int_R g.$$

Proposition 10.1.11 (Monotonicity)**.** *Let $R \subset \mathbb{R}^n$ be a closed rectangle, let f and g be in $\mathscr{R}(R)$, and suppose $f(x) \leq g(x)$ for all $x \in R$. Then*

$$\int_R f \leq \int_R g.$$

Checking for integrability using the definition often involves the following technique, as in the single variable case.

Proposition 10.1.12. *Let $R \subset \mathbb{R}^n$ be a closed rectangle and $f \colon R \to \mathbb{R}$ a bounded function. Then $f \in \mathscr{R}(R)$ if and only if for every $\varepsilon > 0$, there exists a partition P of R such that*

$$U(P,f) - L(P,f) < \varepsilon.$$

Proof. First, if f is integrable, then clearly the supremum of $L(P,f)$ and infimum of $U(P,f)$ must be equal and hence the infimum of $U(P,f) - L(P,f)$ is zero. Therefore for every $\varepsilon > 0$ there must be some partition P such that $U(P,f) - L(P,f) < \varepsilon$.

For the other direction, given an $\varepsilon > 0$ find P such that $U(P,f) - L(P,f) < \varepsilon$.

$$\overline{\int_R} f - \underline{\int_R} f \leq U(P,f) - L(P,f) < \varepsilon.$$

As $\overline{\int_R} f \geq \underline{\int_R} f$ and the above holds for every $\varepsilon > 0$, we conclude $\overline{\int_R} f = \underline{\int_R} f$ and $f \in \mathscr{R}(R)$. $\square$

For simplicity if $f\colon S \to \mathbb{R}$ is a function and $R \subset S$ is a closed rectangle, then if the restriction $f|_R$ is integrable we say f is integrable on R, or $f \in \mathscr{R}(R)$ and we write

$$\int_R f := \int_R f|_R.$$

Proposition 10.1.13. *For a closed rectangle $S \subset \mathbb{R}^n$, if $f\colon S \to \mathbb{R}$ is integrable and $R \subset S$ is a closed rectangle, then f is integrable over R.*

Proof. Given $\varepsilon > 0$, we find a partition P of S such that $U(P,f) - L(P,f) < \varepsilon$. By making a refinement of P if necessary, we assume that the endpoints of R are in P. In other words, R is a union of subrectangles of P. The subrectangles of P divide into two collections, ones that are subsets of R and ones whose intersection with the interior of R is empty. Suppose $R_1, R_2 \ldots, R_K$ are the subrectangles that are subsets of R and let $R_{K+1}, \ldots, R_N$ be the rest. Let $\widetilde{P}$ be the partition of R composed of those subrectangles of P contained in R. Using the same notation as before,

$$\begin{aligned}\varepsilon > U(P,f) - L(P,f) &= \sum_{k=1}^{K} (M_k - m_k) V(R_k) + \sum_{k=K+1}^{N} (M_k - m_k) V(R_k) \\ &\geq \sum_{k=1}^{K} (M_k - m_k) V(R_k) = U(\widetilde{P}, f|_R) - L(\widetilde{P}, f|_R).\end{aligned}$$

Therefore, $f|_R$ is integrable. □

10.1.4 Integrals of continuous functions

Although later we will prove a much more general result, it is useful to start with integrability of continuous functions. First we wish to measure the fineness of partitions. In one variable we measured the length of a subinterval, in several variables, we similarly measure the sides of a subrectangle. We say a rectangle $R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ has *longest side at most* α if $b_k - a_k \leq \alpha$ for all $k = 1, 2, \ldots, n$.

Proposition 10.1.14. *If a rectangle $R \subset \mathbb{R}^n$ has longest side at most α. Then for any $x, y \in R$,*

$$\|x - y\| \leq \sqrt{n}\,\alpha.$$

Proof.

$$\begin{aligned}\|x - y\| &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \\ &\leq \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2} \\ &\leq \sqrt{\alpha^2 + \alpha^2 + \cdots + \alpha^2} = \sqrt{n}\,\alpha.\end{aligned}$$

□

Theorem 10.1.15. *Let $R \subset \mathbb{R}^n$ be a closed rectangle and $f\colon R \to \mathbb{R}$ a continuous function, then $f \in \mathscr{R}(R)$.*

Proof. The proof is analogous to the one variable proof with some complications. The set R is a closed and bounded subset of $\mathbb{R}^n$, and hence compact. So f is not just continuous, but in fact uniformly continuous by Proposition **??**. Let $\varepsilon > 0$ be given. Find a $\delta > 0$ such that $\|x-y\| < \delta$ implies $|f(x) - f(y)| < \frac{\varepsilon}{V(R)}$.

Let P be a partition of R, such that longest side of any subrectangle is strictly less than $\frac{\delta}{\sqrt{n}}$. If $x, y \in R_k$ for some subrectangle R_k of P we have, by the proposition above, $\|x-y\| < \sqrt{n}\frac{\delta}{\sqrt{n}} = \delta$. Therefore

$$f(x) - f(y) \leq |f(x) - f(y)| < \frac{\varepsilon}{V(R)}.$$

As f is continuous on R_k, it attains a maximum and a minimum on this subrectangle. Let x be a point where f attains the maximum and y be a point where f attains the minimum. Then $f(x) = M_k$ and $f(y) = m_k$ in the notation from the definition of the integral. Therefore,

$$M_i - m_i = f(x) - f(y) < \frac{\varepsilon}{V(R)}.$$

And so

$$\begin{aligned} U(P,f) - L(P,f) &= \left(\sum_{k=1}^{N} M_k V(R_k)\right) - \left(\sum_{k=1}^{N} m_k V(R_k)\right) \\ &= \sum_{k=1}^{N} (M_k - m_k) V(R_k) \\ &< \frac{\varepsilon}{V(R)} \sum_{k=1}^{N} V(R_k) = \varepsilon. \end{aligned}$$

Via application of Proposition 10.1.12 we find that $f \in \mathscr{R}(R)$. □

10.1.5 Integration of functions with compact support

Let $U \subset \mathbb{R}^n$ be an open set and $f\colon U \to \mathbb{R}$ be a function. We say the *support* of f is the set

$$\operatorname{supp}(f) := \overline{\{x \in U : f(x) \neq 0\}},$$

where the closure is with respect to the subspace topology on U. Recall that if you take the closure with respect to the subspace topology it is the same as $\overline{\{x \in U : f(x) \neq 0\}} \cap U$, taking the closure with respect to the ambient euclidean space $\mathbb{R}^n$. In particular, $\operatorname{supp}(f) \subset U$. That is, the support is the closure (in U) of the set of points where the function is nonzero, its complement in U is open. So for a point $x \in U$, that is not in the support of f, we have that f is constantly zero in a whole neighbourhood of x.

A function f is said to have *compact support* if $\operatorname{supp}(f)$ is a compact set.

Example 10.1.16: Suppose $B(0,1) \subset \mathbb{R}^2$ is the unit disc. The function $f\colon B(0,1) \to \mathbb{R}$ defined by

$$f(x,y) := \begin{cases} 0 & \text{if } \sqrt{x^2+y^2} > 1/2, \\ 1/2 - \sqrt{x^2+y^2} & \text{if } \sqrt{x^2+y^2} \leq 1/2, \end{cases}$$

is continuous on $B(0,1)$ and its support is the smaller closed ball $C(0,1/2)$. As that is a compact set, f has compact support.

Similarly $g\colon B(0,1) \to \mathbb{R}$ defined by

$$g(x,y) := \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } x > 0, \end{cases}$$

is continuous on $B(0,1)$, but its support is the set $\{(x,y) \in B(0,1) : x \geq 0\}$. In particular g is not compactly supported.

We will mostly consider the case when $U = \mathbb{R}^n$. In light of the following exercise, this is not an oversimplification.

Exercise* 10.1.1: Suppose $U \subset \mathbb{R}^n$ is open and $f\colon U \to \mathbb{R}$ is continuous and of compact support. Show that the function $\widetilde{f}\colon \mathbb{R}^n \to \mathbb{R}$*

$$\widetilde{f}(x) := \begin{cases} f(x) & \textit{if } x \in U, \\ 0 & \textit{otherwise,} \end{cases}$$

is continuous.

Example 10.1.17: On the other hand for the unit disc $B(0,1) \subset \mathbb{R}^2$, the function continuous $f\colon B(0,1) \to \mathbb{R}$ defined by $f(x,y) := \sin\bigl(\frac{1}{1-x^2-y^2}\bigr)$, does not have compact support; as f is not constantly zero on neighbourhood of any point in $B(0,1)$, we know that the support is the entire disc $B(0,1)$. The function clearly does not extend as above to a continuous function. In fact it is not difficult to show that it cannot be extended in any way whatsoever to be continuous on all of $\mathbb{R}^2$ (the boundary of the disc is the problem).

Proposition 10.1.18. *Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous function with compact support. If R and S are closed rectangles such that* $\operatorname{supp}(f) \subset R$ *and* $\operatorname{supp}(f) \subset S$*, then*

$$\int_S f = \int_R f.$$

Proof. As f is continuous, it is automatically integrable on the rectangles R, S, and $R \cap S$. Then Exercise 10.1.7 says $\int_S f = \int_{S\cap R} f = \int_R f$. □

Because of this proposition, when $f\colon \mathbb{R}^n \to \mathbb{R}$ has compact support and is integrable over a rectangle R containing the support we write

$$\int f := \int_R f \qquad \text{or} \qquad \int_{\mathbb{R}^n} f := \int_R f.$$

For example if f is continuous and of compact support then $\int_{\mathbb{R}^n} f$ exists.

10.1.6 Exercises

***Exercise* 10.1.2:** *Prove Proposition 10.1.10.*

***Exercise* 10.1.3:** *Suppose R is a rectangle with the length of one of the sides equal to 0. For any bounded function f, show that $f \in \mathscr{R}(R)$ and $\int_R f = 0$.*

***Exercise* 10.1.4:** *Suppose R is a rectangle with the length of one of the sides equal to 0, and suppose S is a rectangle with $R \subset S$. If f is a bounded function such that $f(x) = 0$ for $x \in R \setminus S$, show that $f \in \mathscr{R}(R)$ and $\int_R f = 0$.*

***Exercise* 10.1.5:** *Suppose $f \colon \mathbb{R}^n \to \mathbb{R}$ is such that $f(x) := 0$ if $x \neq 0$ and $f(0) := 1$. Show that f is integrable on $R := [-1,1] \times [-1,1] \times \cdots \times [-1,1]$ directly using the definition, and find $\int_R f$.*

***Exercise* 10.1.6:** *Suppose R is a closed rectangle and $h \colon R \to \mathbb{R}$ is a bounded function such that $h(x) = 0$ if $x \notin \partial R$ (the boundary of R). Let S be any closed rectangle. Show that $h \in \mathscr{R}(S)$ and*

$$\int_S h = 0.$$

Hint: Write h as a sum of functions as in Exercise 10.1.4.

***Exercise* 10.1.7:** *Suppose R and R' are two closed rectangles with $R' \subset R$. Suppose $f \colon R \to \mathbb{R}$ is in $\mathscr{R}(R')$ and $f(x) = 0$ for $x \in R \setminus R'$. Show that $f \in \mathscr{R}(R)$ and*

$$\int_{R'} f = \int_R f.$$

Do this in the following steps.
a) First do the proof assuming that furthermore $f(x) = 0$ whenever $x \in \overline{R \setminus R'}$.
b) Write $f(x) = g(x) + h(x)$ where $g(x) = 0$ whenever $x \in \overline{R \setminus R'}$, and $h(x)$ is zero except perhaps on $\partial R'$. Then show $\int_R h = \int_{R'} h = 0$ (see Exercise 10.1.6).
c) Show $\int_{R'} f = \int_R f$.

***Exercise* 10.1.8:** *Suppose $R' \subset \mathbb{R}^n$ and $R'' \subset \mathbb{R}^n$ are two rectangles such that $R = R' \cup R''$ is a rectangle, and $R' \cap R''$ is rectangle with one of the sides having length 0 (that is $V(R' \cap R'') = 0$). Let $f \colon R \to \mathbb{R}$ be a function such that $f \in \mathscr{R}(R')$ and $f \in \mathscr{R}(R'')$. Show that $f \in \mathscr{R}(R)$ and*

$$\int_R f = \int_{R'} f + \int_{R''} f.$$

Hint: see previous exercise.

***Exercise* 10.1.9:** *Prove a stronger version of Proposition 10.1.18. Suppose $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function with compact support but not necessarily continuous. Prove that if R is a closed rectangle such that $\operatorname{supp}(f) \subset R$ and f is integrable over R, then for any other closed rectangle S with $\operatorname{supp}(f) \subset S$, the function f is integrable over S and $\int_S f = \int_R f$. Hint: See Exercise 10.1.7.*

***Exercise* 10.1.10:** *Suppose R and S are closed rectangles of $\mathbb{R}^n$. Define $f \colon \mathbb{R}^n \to \mathbb{R}$ as $f(x) := 1$ if $x \in R$, and $f(x) := 0$ otherwise. Prove f is integrable over S and compute $\int_S f$. Hint: Consider $S \cap R$.*

Exercise* 10.1.11*: *Let $R = [0,1] \times [0,1] \subset \mathbb{R}^2$.*
a) Suppose $f \colon R \to \mathbb{R}$ is defined by

$$f(x,y) := \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{else.} \end{cases}$$

Show that $f \in \mathscr{R}(R)$ and compute $\int_R f$.
b) Suppose $f \colon R \to \mathbb{R}$ is defined by

$$f(x,y) := \begin{cases} 1 & \text{if } x \in \mathbb{Q} \text{ or } y \in \mathbb{Q}, \\ 0 & \text{else.} \end{cases}$$

Show that $f \notin \mathscr{R}(R)$.

Exercise* 10.1.12*: *Suppose R is a closed rectangle, and suppose S_j are closed rectangles such that $S_j \subset R$ and $S_j \subset S_{j+1}$ for all j. Suppose $f \colon R \to \mathbb{R}$ is bounded and $f \in \mathscr{R}(S_j)$ for all j. Show that $f \in \mathscr{R}(R)$ and*

$$\lim_{j \to \infty} \int_{S_j} f = \int_R f.$$

Exercise* 10.1.13*: *Suppose $f \colon [-1,1] \times [-1,1] \to \mathbb{R}$ is a Riemann integrable function such $f(x) = -f(-x)$. Using the definition prove*

$$\int_{[-1,1]\times[-1,1]} f = 0.$$

10.2 Iterated integrals and Fubini theorem

Note: 1–2 lectures

The Riemann integral in several variables is hard to compute from the definition. For one-dimensional Riemann integral we have the fundamental theorem of calculus and we can compute many integrals without having to appeal to the definition of the integral. We will rewrite a a Riemann integral in several variables into several one-dimensional Riemann integrals by iterating. However, if $f\colon [0,1]^2 \to \mathbb{R}$ is a Riemann integrable function, it is not immediately clear if the three expressions

$$\int_{[0,1]^2} f, \qquad \int_0^1 \int_0^1 f(x,y)\,dx\,dy, \qquad \text{and} \qquad \int_0^1 \int_0^1 f(x,y)\,dy\,dx$$

are equal, or if the last two are even well-defined.

Example 10.2.1: Define

$$f(x,y) := \begin{cases} 1 & \text{if } x = 1/2 \text{ and } y \in \mathbb{Q}, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is Riemann integrable on $R := [0,1]^2$ and $\int_R f = 0$. Furthermore, $\int_0^1 \int_0^1 f(x,y)\,dx\,dy = 0$. However

$$\int_0^1 f(1/2,y)\,dy$$

does not exist, so we cannot even write $\int_0^1 \int_0^1 f(x,y)\,dy\,dx$.

Proof: Let us start with integrability of f. We simply take the partition of $[0,1]^2$ where the partition in the x direction is $\{0, 1/2-\varepsilon, 1/2+\varepsilon, 1\}$ and in the y direction $\{0,1\}$. The subrectangles of the partition are

$$R_1 := [0, 1/2-\varepsilon] \times [0,1], \qquad R_2 := [1/2-\varepsilon, 1/2+\varepsilon] \times [0,1], \qquad R_3 := [1/2+\varepsilon, 1] \times [0,1].$$

We have $m_1 = M_1 = 0$, $m_2 = 0$, $M_2 = 1$, and $m_3 = M_3 = 0$. Therefore,

$$L(P,f) = m_1 V(R_1) + m_2 V(R_2) + m_3 V(R_3) = 0(1/2-\varepsilon) + 0(2\varepsilon) + 0(1/2-\varepsilon) = 0,$$

and

$$U(P,f) = M_1 V(R_1) + M_2 V(R_2) + M_3 V(R_3) = 0(1/2-\varepsilon) + 1(2\varepsilon) + 0(1/2-\varepsilon) = 2\varepsilon.$$

The upper and lower sum are arbitrarily close and the lower sum is always zero, so the function is integrable and $\int_R f = 0$.

For any y, the function that takes x to $f(x,y)$ is zero except perhaps at a single point $x = 1/2$. We know that such a function is integrable and $\int_0^1 f(x,y)\,dx = 0$. Therefore, $\int_0^1 \int_0^1 f(x,y)\,dx\,dy = 0$.

However if $x = 1/2$, the function that takes y to $f(1/2,y)$ is the nonintegrable function that is 1 on the rationals and 0 on the irrationals. See Example **??**.

We will solve this problem of undefined inside integrals by using the upper and lower integrals, which are always defined.

We split the coordinates of $\mathbb{R}^{n+m}$ into two parts. That is, we write the coordinates on $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ as (x,y) where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. For a function $f(x,y)$ we write

$$f_x(y) := f(x,y)$$

when x is fixed and we wish to speak of the function in terms of y. We write

$$f^y(x) := f(x,y)$$

when y is fixed and we wish to speak of the function in terms of x.

Theorem 10.2.2 (Fubini version A)**.** *Let $R \times S \subset \mathbb{R}^n \times \mathbb{R}^m$ be a closed rectangle and $f\colon R \times S \to \mathbb{R}$ be integrable. The functions $g\colon R \to \mathbb{R}$ and $h\colon R \to \mathbb{R}$ defined by*

$$g(x) := \underline{\int_S} f_x \qquad \text{and} \qquad h(x) := \overline{\int_S} f_x$$

are integrable over R and

$$\int_R g = \int_R h = \int_{R\times S} f.$$

In other words

$$\int_{R\times S} f = \int_R \left(\underline{\int_S} f(x,y)\,dy \right) dx = \int_R \left(\overline{\int_S} f(x,y)\,dy \right) dx.$$

If it turns out that f_x is integrable for all x, for example when f is continuous, then we obtain the more familiar

$$\int_{R\times S} f = \int_R \int_S f(x,y)\,dy\,dx.$$

Proof. Any partition of $R \times S$ is a concatenation of a partition of R and a partition of S. That is, write a partition of $R \times S$ as $(P,P') = (P_1,P_2,\ldots,P_n,P'_1,P'_2,\ldots,P'_m)$. where $P = (P_1,P_2,\ldots,P_n)$ and $P' = (P'_1,P'_2,\ldots,P'_m)$ are partitions of R and S respectively. Let $R_1,R_2,\ldots,R_N$ be the subrectangles of P and $R'_1,R'_2,\ldots,R'_K$ be the subrectangles of P'. Then the subrectangles of (P,P') are $R_j \times R'_k$ where $1 \le j \le N$ and $1 \le k \le K$.

Let

$$m_{j,k} := \inf_{(x,y)\in R_j \times R'_k} f(x,y).$$

We notice that $V(R_j \times R'_k) = V(R_j)V(R'_k)$ and hence

$$L\big((P,P'),f\big) = \sum_{j=1}^{N}\sum_{k=1}^{K} m_{j,k} V(R_j \times R'_k) = \sum_{j=1}^{N} \left(\sum_{k=1}^{K} m_{j,k} V(R'_k) \right) V(R_j).$$

If we let

$$m_k(x) := \inf_{y \in R'_k} f(x,y) = \inf_{y \in R'_k} f_x(y),$$

then of course if $x \in R_j$ then $m_{j,k} \leq m_k(x)$. Therefore

$$\sum_{k=1}^{K} m_{j,k} V(R'_k) \leq \sum_{k=1}^{K} m_k(x) V(R'_k) = L(P', f_x) \leq \underline{\int_S} f_x = g(x).$$

As we have the inequality for all $x \in R_j$ we have

$$\sum_{k=1}^{K} m_{j,k} V(R'_k) \leq \inf_{x \in R_j} g(x).$$

We thus obtain

$$L\big((P,P'),f\big) \leq \sum_{j=1}^{N} \left(\inf_{x \in R_j} g(x) \right) V(R_j) = L(P,g).$$

Similarly $U\big((P,P'),f\big) \geq U(P,h)$, and the proof of this inequality is left as an exercise.

Putting this together we have

$$L\big((P,P'),f\big) \leq L(P,g) \leq U(P,g) \leq U(P,h) \leq U\big((P,P'),f\big).$$

And since f is integrable, it must be that g is integrable as

$$U(P,g) - L(P,g) \leq U\big((P,P'),f\big) - L\big((P,P'),f\big),$$

and we can make the right hand side arbitrarily small. As for any partition we have $L\big((P,P'),f\big) \leq L(P,g) \leq U\big((P,P'),f\big)$ we must have that $\int_R g = \int_{R \times S} f$.

Similarly we have

$$L\big((P,P'),f\big) \leq L(P,g) \leq L(P,h) \leq U(P,h) \leq U\big((P,P'),f\big),$$

and hence

$$U(P,h) - L(P,h) \leq U\big((P,P'),f\big) - L\big((P,P'),f\big).$$

So if f is integrable so is h, and as $L\big((P,P'),f\big) \leq L(P,h) \leq U\big((P,P'),f\big)$ we must have that $\int_R h = \int_{R \times S} f$. □

We can also do the iterated integration in opposite order. The proof of this version is almost identical to version A, and we leave it as an exercise to the reader.

Theorem 10.2.3 (Fubini version B)**.** *Let $R \times S \subset \mathbb{R}^n \times \mathbb{R}^m$ be a closed rectangle and $f\colon R \times S \to \mathbb{R}$ be integrable. The functions $g\colon S \to \mathbb{R}$ and $h\colon S \to \mathbb{R}$ defined by*

$$g(y) := \underline{\int_R} f^y \qquad \text{and} \qquad h(y) := \overline{\int_R} f^y$$

are integrable over S and

$$\int_S g = \int_S h = \int_{R\times S} f.$$

That is we also have

$$\int_{R\times S} f = \int_S \left(\underline{\int_R} f(x,y)\,dx \right) dy = \int_S \left(\overline{\int_R} f(x,y)\,dx \right) dy.$$

Next suppose that f_x and f^y are integrable for simplicity. For example, suppose that f is continuous. Then by putting the two versions together we obtain the familiar

$$\int_{R\times S} f = \int_R \int_S f(x,y)\,dy\,dx = \int_S \int_R f(x,y)\,dx\,dy.$$

Often the Fubini theorem is stated in two dimensions for a continuous function $f\colon R \to \mathbb{R}$ on a rectangle $R = [a,b] \times [c,d]$. Then the Fubini theorem states that

$$\int_R f = \int_a^b \int_c^d f(x,y)\,dy\,dx = \int_c^d \int_a^b f(x,y)\,dx\,dy.$$

And the Fubini theorem is commonly thought of as the theorem that allows us to swap the order of iterated integrals.

Repeatedly applying Fubini theorem gets us the following corollary: Let $R := [a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_n,b_n] \subset \mathbb{R}^n$ be a closed rectangle and let $f\colon R \to \mathbb{R}$ be continuous. Then

$$\int_R f = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1,x_2,\ldots,x_n)\,dx_n\,dx_{n-1}\cdots dx_1.$$

Clearly we can also switch the order of integration to any order we please. We can also relax the continuity requirement by making sure that all the intermediate functions are integrable, or by using upper or lower integrals.

10.2.1 Exercises

Exercise* 10.2.1: Compute $\int_0^1 \int_{-1}^1 xe^{xy}\,dx\,dy$ in a simple way.*

Exercise* 10.2.2: Prove the assertion $U\big((P,P'),f\big) \geq U(P,h)$ from the proof of Theorem 10.2.2.*

Exercise* 10.2.3: Prove Theorem 10.2.3.*

***Exercise* 10.2.4:** *Let $R = [a,b] \times [c,d]$ and $f(x,y)$ is an integrable function on R such that such that for any fixed y, the function that takes x to $f(x,y)$ is zero except at finitely many points. Show*

$$\int_R f = 0.$$

***Exercise* 10.2.5:** *Let $R = [a,b] \times [c,d]$ and $f(x,y) := g(x)h(y)$ for two continuous functions $g\colon [a,b] \to \mathbb{R}$ and $h\colon [a,b] \to \mathbb{R}$. Prove*

$$\int_R f = \left(\int_a^b g\right)\left(\int_c^d h\right).$$

***Exercise* 10.2.6:** *Compute*

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2+y^2)^2}\,dx\,dy \qquad \text{and} \qquad \int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2+y^2)^2}\,dy\,dx.$$

You will need to interpret the integrals as improper, that is, the limit of $\int_\varepsilon^1$ as $\varepsilon \to 0$.

***Exercise* 10.2.7:** *Suppose $f(x,y) := g(x)$ where $g\colon [a,b] \to \mathbb{R}$ is Riemann integrable. Show that f is Riemann integrable for any $R = [a,b] \times [c,d]$ and*

$$\int_R f = (d-c)\int_a^b g.$$

***Exercise* 10.2.8:** *Define $f\colon [-1,1] \times [0,1] \to \mathbb{R}$ by*

$$f(x,y) := \begin{cases} x & \text{if } y \in \mathbb{Q}, \\ 0 & \text{else.} \end{cases}$$

Show
a) $\int_0^1 \int_{-1}^1 f(x,y)\,dx\,dy$ exists, but $\int_{-1}^1 \int_0^1 f(x,y)\,dy\,dx$ does not.
b) Compute $\int_{-1}^1 \overline{\int_0^1} f(x,y)\,dy\,dx$ and $\int_{-1}^1 \underline{\int_0^1} f(x,y)\,dy\,dx$.
c) Show f is not Riemann integrable on $[-1,1] \times [0,1]$ (use Fubini).

***Exercise* 10.2.9:** *Define $f\colon [0,1] \times [0,1] \to \mathbb{R}$ by*

$$f(x,y) := \begin{cases} 1/q & \text{if } x \in \mathbb{Q},\ y \in \mathbb{Q}, \text{ and } y = p/q \text{ in lowest terms,} \\ 0 & \text{else.} \end{cases}$$

Show
a) Show f is Riemann integrable on $[0,1] \times [0,1]$.
b) Find $\overline{\int_0^1} f(x,y)\,dx$ and $\underline{\int_0^1} f(x,y)\,dx$ for all $y \in [0,1]$, and show they are unequal for all $y \in \mathbb{Q}$.
c) $\int_0^1 \int_0^1 f(x,y)\,dy\,dx$ exists, but $\int_0^1 \int_0^1 f(x,y)\,dx\,dy$ does not.
Note: By Fubini, $\int_0^1 \overline{\int_0^1} f(x,y)\,dy\,dx$ and $\int_0^1 \underline{\int_0^1} f(x,y)\,dy\,dx$ do exist and equal the integral of f on R.

10.3 Outer measure and null sets

Note: 2 lectures

10.3.1 Outer measure and null sets

Before we characterize all Riemann integrable functions, we need to make a slight detour. We introduce a way of measuring the size of sets in $\mathbb{R}^n$.

Definition 10.3.1. Let $S \subset \mathbb{R}^n$ be a subset. Define the *outer measure* of S as

$$m^*(S) := \inf \sum_{j=1}^{\infty} V(R_j),$$

where the infimum is taken over all sequences $\{R_j\}$ of open rectangles such that $S \subset \bigcup_{j=1}^{\infty} R_j$. In particular S is of *measure zero* or a *null set* if $m^*(S) = 0$.

We will only need measure zero sets and so we focus on these. Note that S is of measure zero if for every $\varepsilon > 0$ there exist a sequence of open rectangles $\{R_j\}$ such that

$$S \subset \bigcup_{j=1}^{\infty} R_j \qquad \text{and} \qquad \sum_{j=1}^{\infty} V(R_j) < \varepsilon. \tag{10.2}$$

Furthermore, if S is measure zero and $S' \subset S$, then S' is of measure zero. We can in fact use the same exact rectangles.

It is sometimes more convenient to use balls instead of rectangles. In fact we can choose balls no bigger than a fixed radius.

Proposition 10.3.2. *Let $\delta > 0$ be given. A set $S \subset \mathbb{R}^n$ is measure zero if and only if for every $\varepsilon > 0$, there exists a sequence of open balls $\{B_j\}$, where the radius of B_j is $r_j < \delta$ such that*

$$S \subset \bigcup_{j=1}^{\infty} B_j \qquad \textit{and} \qquad \sum_{j=1}^{\infty} r_j^n < \varepsilon.$$

Note that the "volume" of B_j is proportional to r_j^n.

Proof. If R is a (closed or open) cube (rectangle with all sides equal) of side s, then R is contained in a closed ball of radius $\sqrt{n}\,s$ by Proposition 10.1.14, and therefore in an open ball of size $2\sqrt{n}\,s$.

Let s be a number that is less than the smallest side of R and also so that $2\sqrt{n}\,s < \delta$. We claim R is contained in a union of closed cubes $C_1, C_2, \ldots, C_k$ of sides s such that

$$\sum_{j=1}^{k} V(C_j) \leq 2^n V(R).$$

It is clearly true (without the 2^n) if R has sides that are integer multiples of s. So if a side is of length $(\ell+\alpha)s$, for $\ell \in \mathbb{N}$ and $0 \leq \alpha < 1$, then $(\ell+\alpha)s \leq 2\ell s$. Increasing the side to $2\ell s$ we obtain a new larger rectangle of volume at most 2^n times larger, but whose sides are multiples of s.

So suppose that there exist $\{R_j\}$ as in the definition such that (10.2) is true. As we have seen above, we can choose closed cubes $\{C_k\}$ with C_k of side s_k as above that cover all the rectangles $\{R_j\}$ and so that

$$\sum_{k=1}^{\infty} s_k^n = \sum_{k=1}^{\infty} V(C_k) \leq 2^n \sum_{j=1}^{\infty} V(R_k) < 2^n \varepsilon.$$

Covering C_k with balls B_k of radius $r_k = 2\sqrt{n}\, s_k$ we obtain

$$\sum_{k=1}^{\infty} r_k^n < 2^{2n} n \varepsilon.$$

And as $S \subset \bigcup_j R_j \subset \bigcup_k C_k \subset \bigcup_k B_k$, we are finished.

Suppose we have the ball condition above for some $\varepsilon > 0$. Without loss of generality assume that all $r_j < 1$. Each B_j is contained a in a cube R_j of side $2r_j$. So $V(R_j) = (2r_j)^n < 2^n r_j$. Therefore

$$S \subset \bigcup_{j=1}^{\infty} R_j \qquad \text{and} \qquad \sum_{j=1}^{\infty} V(R_j) < \sum_{j=1}^{\infty} 2^n r_j < 2^n \varepsilon. \qquad \square$$

The definition of outer measure could have been done with open balls as well, not just null sets. We leave this generalization to the reader.

10.3.2 Examples and basic properties

Example 10.3.3: The set $\mathbb{Q}^n \subset \mathbb{R}^n$ of points with rational coordinates is a set of measure zero.

Proof: The set $\mathbb{Q}^n$ is countable and therefore let us write it as a sequence $q_1, q_2, \ldots$. For each q_j find an open rectangle R_j with $q_j \in R_j$ and $V(R_j) < \varepsilon 2^{-j}$. Then

$$\mathbb{Q}^n \subset \bigcup_{j=1}^{\infty} R_j \qquad \text{and} \qquad \sum_{j=1}^{\infty} V(R_j) < \sum_{j=1}^{\infty} \varepsilon 2^{-j} = \varepsilon.$$

The example points to a more general result.

Proposition 10.3.4. *A countable union of measure zero sets is of measure zero.*

Proof. Suppose

$$S = \bigcup_{j=1}^{\infty} S_j,$$

where S_j are all measure zero sets. Let $\varepsilon > 0$ be given. For each j there exists a sequence of open rectangles $\{R_{j,k}\}_{k=1}^{\infty}$ such that

$$S_j \subset \bigcup_{k=1}^{\infty} R_{j,k}$$

and

$$\sum_{k=1}^{\infty} V(R_{j,k}) < 2^{-j}\varepsilon.$$

Then

$$S \subset \bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} R_{j,k}.$$

As $V(R_{j,k})$ is always positive, the sum over all j and k can be done in any order. In particular, it can be done as

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} V(R_{j,k}) < \sum_{j=1}^{\infty} 2^{-j}\varepsilon = \varepsilon. \qquad \square$$

The next example is not just interesting, it will be useful later.

Example 10.3.5: Let $P := \{x \in \mathbb{R}^n : x_k = c\}$ for a fixed $k = 1, 2, \ldots, n$ and a fixed constant $c \in \mathbb{R}$. Then P is of measure zero.

Proof: First fix s and let us prove that

$$P_s := \{x \in \mathbb{R}^n : x_k = c, \left|x_j\right| \leq s \text{ for all } j \neq k\}$$

is of measure zero. Given any $\varepsilon > 0$ define the open rectangle

$$R := \{x \in \mathbb{R}^n : c - \varepsilon < x_k < c + \varepsilon, \left|x_j\right| < s + 1 \text{ for all } j \neq k\}.$$

It is clear that $P_s \subset R$. Furthermore

$$V(R) = 2\varepsilon\big(2(s+1)\big)^{n-1}.$$

As s is fixed, we can make $V(R)$ arbitrarily small by picking ε small enough.

Next we note that

$$P = \bigcup_{j=1}^{\infty} P_j$$

and a countable union of measure zero sets is measure zero.

Example 10.3.6: If $a < b$, then $m^*([a,b]) = b - a$.

Proof: In the case of $\mathbb{R}$, open rectangles are open intervals. Since $[a,b] \subset (a-\varepsilon, b+\varepsilon)$ for all $\varepsilon > 0$. Hence, $m^*([a,b]) \leq b - a$.

Let us prove the other inequality. Suppose $\{(a_j, b_j)\}$ are open intervals such that

$$[a,b] \subset \bigcup_{j=1}^{\infty} (a_j, b_j).$$

We wish to bound $\sum(b_j - a_j)$ from below. Since $[a,b]$ is compact, then there are only finitely many open intervals that still cover $[a,b]$. As throwing out some of the intervals only makes the sum smaller, we only need to take the finite number of intervals still covering $[a,b]$. If $(a_i, b_i) \subset (a_j, b_j)$, then we can throw out (a_i, b_i) as well. Therefore we have $[a,b] \subset \bigcup_{j=1}^{k} (a_j, b_j)$ for some k, and we assume that the intervals are sorted such that $a_1 < a_2 < \cdots < a_k$. Note that since (a_2, b_2) is not contained in (a_1, b_1) we have that $a_1 < a_2 < b_1 < b_2$. Similarly $a_j < a_{j+1} < b_j < b_{j+1}$. Furthermore, $a_1 < a$ and $b_k > b$. Thus,

$$m^*([a,b]) \geq \sum_{j=1}^{k} (b_j - a_j) \geq \sum_{j=1}^{k-1} (a_{j+1} - a_j) + (b_k - a_k) = b_k - a_1 > b - a.$$

Proposition 10.3.7. *Suppose $E \subset \mathbb{R}^n$ is a compact set of measure zero. Then for every $\varepsilon > 0$, there exist finitely many open rectangles $R_1, R_2, \ldots, R_k$ such that*

$$E \subset R_1 \cup R_2 \cup \cdots \cup R_k \qquad \textit{and} \qquad \sum_{j=1}^{k} V(R_j) < \varepsilon.$$

Also for any $\delta > 0$, there exist finitely many open balls $B_1, B_2, \ldots, B_k$ of radii $r_1, r_2, \ldots, r_k < \delta$ such that

$$E \subset B_1 \cup B_2 \cup \cdots \cup B_k \qquad \textit{and} \qquad \sum_{j=1}^{k} r_j^n < \varepsilon.$$

Proof. Find a sequence of open rectangles $\{R_j\}$ such that

$$E \subset \bigcup_{j=1}^{\infty} R_j \qquad \text{and} \qquad \sum_{j=1}^{\infty} V(R_j) < \varepsilon.$$

By compactness, finitely many of these rectangles still contain E. That is, there is some k such that $E \subset R_1 \cup R_2 \cup \cdots \cup R_k$. Hence

$$\sum_{j=1}^{k} V(R_j) \leq \sum_{j=1}^{\infty} V(R_j) < \varepsilon.$$

The proof that we can choose balls instead of rectangles is left as an exercise. □

10.3.3 Images of null sets

Before we look at images of measure zero sets, let us see what a continuously differentiable function does to a ball.

Lemma 10.3.8. *Suppose $U \subset \mathbb{R}^n$ is an open set, $B \subset U$ is an open or closed ball of radius at most r, $f\colon B \to \mathbb{R}^n$ is continuously differentiable and suppose $\|f'(x)\| \leq M$ for all $x \in B$. Then $f(B) \subset B'$, where B' is a ball of radius at most Mr.*

Proof. Without loss of generality assume B is a closed ball. The ball B is convex, and hence via Proposition 8.4.2, that $\|f(x) - f(y)\| \leq M\|x-y\|$ for all x, y in B. In particular, suppose $B = C(y,r)$, then $f(B) \subset C\bigl(f(y), Mr\bigr)$. □

The image of a measure zero set using a continuous map is not necessarily a measure zero set. However if we assume the mapping is continuously differentiable, then the mapping cannot "stretch" the set too much. The proposition does not require compactness, and this is left as an exercise.

Proposition 10.3.9. *Suppose $U \subset \mathbb{R}^n$ is an open set and $f\colon U \to \mathbb{R}^n$ is a continuously differentiable mapping. If $E \subset U$ is a compact measure zero set, then $f(E)$ is measure zero.*

Proof. We must first handle a couple of techicalities. First let us replace U by a smaller open set to make $\|f'(x)\|$ bounded. At each point $x \in E$ pick an open ball $B(x, r_x)$ such that the closed ball $C(x, r_x) \subset U$. By compactness we only need to take finitely many points $x_1, x_2, \ldots, x_q$ to still conver E. Define

$$U' := \bigcup_{j=1}^{q} B(x_j, r_{x_j}), \qquad K := \bigcup_{j=1}^{q} C(x_j, r_{x_j}). \tag{10.3}$$

We have $E \subset U' \subset K \subset U$. The set K is compact. The function that takes x to $\|f'(x)\|$ is continuous, and therefore there exists an $M > 0$ such that $\|f'(x)\| \leq M$ for all $x \in K$.

So without loss of generality we may replace U by U' and from now on suppose that $\|f'(x)\| \leq M$ for all $x \in U$.

At each point $x \in E$ pick a ball $B(x, \delta_x)$ of maximum radius so that $B(x, \delta_x) \subset U$. Let $\delta = \inf_{x \in E} \delta_x$. Take a sequence $\{x_j\} \subset E$ so that $\delta_{x_j} \to \delta$. As E is compact, we can pick the sequence to be convergent to some $y \in E$. Once $\|x_j - y\| < \frac{\delta_y}{2}$, then $\delta_{x_j} > \frac{\delta_y}{2}$ by the triangle inequality. Therefore $\delta > 0$.

Given $\varepsilon > 0$, there exist balls $B_1, B_2, \ldots, B_k$ of radii $r_1, r_2, \ldots, r_k < \delta$ such that

$$E \subset B_1 \cup B_2 \cup \cdots \cup B_k \qquad \text{and} \qquad \sum_{j=1}^{k} r_j^n < \varepsilon.$$

Suppose $B'_1, B'_2, \ldots, B'_k$ are the balls of radius $Mr_1, Mr_2, \ldots, Mr_k$ from Lemma 10.3.8.

$$f(E) \subset f(B_1) \cup f(B_2) \cup \cdots \cup f(B_k) \subset B'_1 \cup B'_2 \cup \cdots \cup B'_k \qquad \text{and} \qquad \sum_{j=1}^{k} Mr_j^n < M\varepsilon. \qquad \square$$

10.3.4 Exercises

Exercise* 10.3.1: Finish the proof of Proposition 10.3.7, that is, show that you can use balls instead of rectangles.*

Exercise* 10.3.2: If $A \subset B$ then $m^*(A) \leq m^*(B)$.*

Exercise* 10.3.3: Show that if $R \subset \mathbb{R}^n$ is a closed rectangle then $m^*(R) = V(R)$.*

Exercise* 10.3.4: Prove a version of Proposition 10.3.9 without using compactness:*
a) Mimic the proof to first prove that the proposition holds only if E is relatively compact*; a set $E \subset U$ is relatively compact if the closure of E in the subspace topology on U is compact, or in other words if there exists a compact set K with $K \subset U$ and $E \subset K$.*
Hint: The bound on the size of the derivative still holds, but you may need to use countably many balls. Be careful as the closure of E need no longer be measure zero.
b) Now prove it for any null set E.
Hint: First show that $\{x \in U : d(x,y) \geq 1/m$ for all $y \notin U$ and $d(0,x) \leq m\}$ is a compact set for any $m > 0$.

Exercise* 10.3.5: Let $U \subset \mathbb{R}^n$ be an open set and let $f \colon U \to \mathbb{R}$ be a continuously differentiable function. Let $G := \{(x,y) \in U \times \mathbb{R} : y = f(x)\}$ be the graph of f. Show that f is of measure zero.*

Exercise* 10.3.6: Given a closed rectangle $R \subset \mathbb{R}^n$, show that for any $\varepsilon > 0$ there exists a number $s > 0$ and finitely many open cubes $C_1, C_2, \ldots, C_k$ of side s such that $R \subset C_1 \cup C_2 \cup \cdots \cup C_k$ and*

$$\sum_{j=1}^{k} V(C_j) \leq V(R) + \varepsilon.$$

Exercise* 10.3.7: Show that there exists a number $k = k(n,r,\delta)$ depending only on n, r and δ such the following holds. Given $B(x,r) \subset \mathbb{R}^n$ and $\delta > 0$, there exist k open balls $B_1, B_2, \ldots, B_k$ of radius at most δ such that $B(x,r) \subset B_1 \cup B_2 \cup \cdots \cup B_k$. Note that you can find k that really only depends on n and the ratio δ/r.*

10.4 The set of Riemann integrable functions

Note: ??? lectures

10.4.1 Oscillation and continuity

Let $S \subset \mathbb{R}^n$ be a set and $f\colon S \to \mathbb{R}$ a function. Instead of just saying that f is or is not continuous at a point $x \in S$, we we need to be able to quantify how discontinuous f is at a function is at x. For any $\delta > 0$ define the oscillation of f on the δ-ball in subset topology that is $B_S(x,\delta) = B_{\mathbb{R}^n}(x,\delta) \cap S$ as

$$o(f,x,\delta) := \sup_{y \in B_S(x,\delta)} f(y) - \inf_{y \in B_S(x,\delta)} f(y) = \sup_{y_1,y_2 \in B_S(x,\delta)} \bigl(f(y_1) - f(y_2)\bigr).$$

That is, $o(f,x,\delta)$ is the length of the smallest interval that contains the image $f\bigl(B_S(x,\delta)\bigr)$. Clearly $o(f,x,\delta) \geq 0$ and notice $o(f,x,\delta) \leq o(f,x,\delta')$ whenever $\delta < \delta'$. Therefore, the limit as $\delta \to 0$ from the right exists and we define the *oscillation* of a function f at x as

$$o(f,x) := \lim_{\delta \to 0^+} o(f,x,\delta) = \inf_{\delta > 0} o(f,x,\delta).$$

Proposition 10.4.1. *$f\colon S \to \mathbb{R}$ is continuous at $x \in S$ if and only if $o(f,x) = 0$.*

Proof. First suppose that f is continuous at $x \in S$. Then given any $\varepsilon > 0$, there exists a $\delta > 0$ such that for $y \in B_S(x,\delta)$ we have $|f(x) - f(y)| < \varepsilon$. Therefore if $y_1, y_2 \in B_S(x,\delta)$ then

$$f(y_1) - f(y_2) = f(y_1) - f(x) - \bigl(f(y_2) - f(x)\bigr) < \varepsilon + \varepsilon = 2\varepsilon.$$

We take the supremum over y_1 and y_2

$$o(f,x,\delta) = \sup_{y_1,y_2 \in B_S(x,\delta)} \bigl(f(y_1) - f(y_2)\bigr) \leq 2\varepsilon.$$

Hence, $o(x,f) = 0$.

On the other hand suppose that $o(x,f) = 0$. Given any $\varepsilon > 0$, find a $\delta > 0$ such that $o(f,x,\delta) < \varepsilon$. If $y \in B_S(x,\delta)$ then

$$|f(x) - f(y)| \leq \sup_{y_1,y_2 \in B_S(x,\delta)} \bigl(f(y_1) - f(y_2)\bigr) = o(f,x,\delta) < \varepsilon. \qquad \square$$

Proposition 10.4.2. *Let $S \subset \mathbb{R}^n$ be closed, $f\colon S \to \mathbb{R}$, and $\varepsilon > 0$. The set $\{x \in S : o(f,x) \geq \varepsilon\}$ is closed.*

Proof. Equivalently we want to show that $G = \{x \in S : o(f,x) < \varepsilon\}$ is open in the subset topology. As $\inf_{\delta > 0} o(f,x,\delta) < \varepsilon$, find a $\delta > 0$ such that

$$o(f,x,\delta) < \varepsilon$$

Take any $\xi \in B_S(x, \delta/2)$. Notice that $B_S(\xi, \delta/2) \subset B_S(x, \delta)$. Therefore,

$$o(f, \xi, \delta/2) = \sup_{y_1, y_2 \in B_S(\xi, \delta/2)} \big(f(y_1) - f(y_2)\big) \leq \sup_{y_1, y_2 \in B_S(x, \delta)} \big(f(y_1) - f(y_2)\big) = o(f, x, \delta) < \varepsilon.$$

So $o(f, \xi) < \varepsilon$ as well. As this is true for all $\xi \in B_S(x, \delta/2)$ we get that G is open in the subset topology and $S \setminus G$ is closed as is claimed. □

10.4.2 The set of Riemann integrable functions

We have seen that continuous functions are Riemann integrable, but we also know that certain kinds of discontinuities are allowed. It turns out that as long as the discontinuities happen on a set of measure zero, the function is integrable and vice versa.

Theorem 10.4.3 (Riemann-Lebesgue)**.** *Let $R \subset \mathbb{R}^n$ be a closed rectangle and $f \colon R \to \mathbb{R}$ a bounded function. Then f is Riemann integrable if and only if the set of discontinuities of f is of measure zero (a null set).*

Proof. Let $S \subset R$ be the set of discontinuities of f. That is $S = \{x \in R : o(f, x) > 0\}$. The trick to this proof is to isolate the bad set into a small set of subrectangles of a partition. There are only finitely many subrectangles of a partition, so we will wish to use compactness. If S is closed, then it would be compact and we could cover it by small rectangles as it is of measure zero. Unfortunately, in general S is not closed so we need to work a little harder.

For every $\varepsilon > 0$, define

$$S_\varepsilon := \{x \in R : o(f, x) \geq \varepsilon\}.$$

By Proposition 10.4.2 S_ε is closed and as it is a subset of R which is bounded, S_ε is compact. Furthermore, $S_\varepsilon \subset S$ and S is of measure zero. Via Proposition 10.3.7 there are finitely many open rectangles $S_1, S_2, \ldots, S_k$ that cover S_ε and $\sum V(S_j) < \varepsilon$.

The set $T = R \setminus (S_1 \cup \cdots \cup S_k)$ is closed, bounded, and therefore compact. Furthermore for $x \in T$, we have $o(f, x) < \varepsilon$. Hence for each $x \in T$, there exists a small closed rectangle T_x with x in the interior of T_x, such that

$$\sup_{y \in T_x} f(y) - \inf_{y \in T_x} f(y) < 2\varepsilon.$$

The interiors of the rectangles T_x cover T. As T is compact there exist finitely many such rectangles $T_1, T_2, \ldots, T_m$ that covers T.

Now take all the rectangles $T_1, T_2, \ldots, T_m$ and $S_1, S_2, \ldots, S_k$ and construct a partition out of their endpoints. That is construct a partition P with subrectangles $R_1, R_2, \ldots, R_p$ such that every R_j is contained in T_ℓ for some ℓ or the closure of S_ℓ for some ℓ. Suppose we order the rectangles so that $R_1, R_2, \ldots, R_q$ are those that are contained in some T_ℓ, and $R_{q+1}, R_{q+2}, \ldots, R_p$ are the rest. In particular, we have

$$\sum_{j=1}^{q} V(R_j) \leq V(R) \qquad \text{and} \qquad \sum_{j=q+1}^{p} V(R_j) \leq \varepsilon.$$

Let m_j and M_j be the inf and sup over R_j as before. If $R_j \subset T_\ell$ for some ℓ, then $(M_j - m_j) < 2\varepsilon$. Let $B \in \mathbb{R}$ be such that $|f(x)| \leq B$ for all $x \in R$, so $(M_j - m_j) < 2B$ over all rectangles. Then

$$\begin{aligned} U(P,f) - L(P,f) &= \sum_{j=1}^{p} (M_j - m_j) V(R_j) \\ &= \left(\sum_{j=1}^{q} (M_j - m_j) V(R_j) \right) + \left(\sum_{j=q+1}^{p} (M_j - m_j) V(R_j) \right) \\ &\leq \left(\sum_{j=1}^{q} 2\varepsilon V(R_j) \right) + \left(\sum_{j=q+1}^{p} 2B V(R_j) \right) \\ &\leq 2\varepsilon V(R) + 2B\varepsilon = \varepsilon(2V(R) + 2B). \end{aligned}$$

Clearly, we can make the right hand side as small as we want and hence f is integrable.

For the other direction, suppose that f is Riemann integrable over R. Let S be the set of discontinuities again and now let

$$S_k := \{x \in R : o(f,x) \geq 1/k\}.$$

Fix a $k \in \mathbb{N}$. Given an $\varepsilon > 0$, find a partition P with subrectangles $R_1, R_2, \ldots, R_p$ such that

$$U(P,f) - L(P,f) = \sum_{j=1}^{p} (M_j - m_j) V(R_j) < \varepsilon$$

Suppose $R_1, R_2, \ldots, R_p$ arr ordered so that the interiors of $R_1, R_2, \ldots, R_q$ intersect S_k, while the interiors of $R_{q+1}, R_{q+2}, \ldots, R_p$ are disjoint from S_k. If $x \in R_j \cap S_k$ and x is in the interior of R_j so sufficiently small balls are completely inside R_j, then by definition of S_k we have $M_j - m_j \geq 1/k$. Then

$$\varepsilon > \sum_{j=1}^{p} (M_j - m_j) V(R_j) \geq \sum_{j=1}^{q} (M_j - m_j) V(R_j) \geq \frac{1}{k} \sum_{j=1}^{q} V(R_j)$$

In other words $\sum_{j=1}^{q} V(R_j) < k\varepsilon$. Let G be the set of all boundaries of all the subrectangles of P. The set G is of measure zero (see Example 10.3.5). Let R_j° denote the interior of R_j, then

$$S_k \subset R_1^\circ \cup R_2^\circ \cup \cdots \cup R_q^\circ \cup G.$$

As G can be covered by open rectangles arbitrarily small volume, S_k must be of measure zero. As

$$S = \bigcup_{k=1}^{\infty} S_k$$

and a countable union of measure zero sets is of measure zero, S is of measure zero. $\square$

10.4.3 Exercises

Exercise* 10.4.1: Suppose $f\colon (a,b)\times(c,d)\to\mathbb{R}$ is a bounded continuous function. Show that the integral of f over $R=[a,b]\times[c,d]$ makes sense and is uniquely defined. That is, set f to be anything on the boundary of R and compute the integral.*

Exercise* 10.4.2: Suppose $R\subset\mathbb{R}^n$ is a closed rectangle. Show that $\mathscr{R}(R)$, the set of Riemann integrable functions, is an algebra. That is, show that if $f,g\in\mathscr{R}(R)$ and $a\in\mathbb{R}$, then $af\in\mathscr{R}(R)$, $f+g\in\mathscr{R}(R)$ and $fg\in\mathscr{R}(R)$.*

Exercise* 10.4.3: Suppose $R\subset\mathbb{R}^n$ is a closed rectangle and $f\colon R\to\mathbb{R}$ is a bounded function which is zero except on a closed set $E\subset R$ of measure zero. Show that $\int_R f$ exists and compute it.*

Exercise* 10.4.4: Suppose $R\subset\mathbb{R}^n$ is a closed rectangle and $f\colon R\to\mathbb{R}$ and $g\colon R\to\mathbb{R}$ are two Riemann integrable functions. Suppose $f=g$ except for a closed set $E\subset R$ of measure zero. Show that $\int_R f=\int_R g$.*

Exercise* 10.4.5: Suppose $R\subset\mathbb{R}^n$ is a closed rectangle and $f\colon R\to\mathbb{R}$ is a bounded function.*
a) Suppose there exists a closed set $E\subset R$ of measure zero such that $f|_{R\setminus E}$ is continuous. Then $f\in\mathscr{R}(R)$.
b) Find am example where $E\subset R$ is a set of measure zero (but not closed) such that $f|_{R\setminus E}$ is continuous and $f\notin\mathscr{R}(R)$.

10.5 Jordan measurable sets

Note: ??? lectures

10.5.1 Volume and Jordan measurable sets

Given a bounded set $S \subset \mathbb{R}^n$ its *characteristic function* or *indicator function* is

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

A bounded set S is *Jordan measurable* if for some closed rectangle R such that $S \subset R$, the function χ_S is in $\mathscr{R}(R)$. Take two closed rectangles R and R' with $S \subset R$ and $S \subset R'$, then $R \cap R'$ is a closed rectangle also containing S. By Proposition 10.1.13 and Exercise 10.1.7, $\chi_S \in \mathscr{R}(R \cap R')$ and so $\chi_S \in \mathscr{R}(R')$. Thus

$$\int_R \chi_S = \int_{R'} \chi_S = \int_{R \cap R'} \chi_S.$$

We define the *n-dimensional volume* of the bounded Jordan measurable set S as

$$V(S) := \int_R \chi_S,$$

where R is any closed rectangle containing S.

Proposition 10.5.1. *A bounded set $S \subset \mathbb{R}^n$ is Jordan measurable if and only if the boundary ∂S is a measure zero set.*

Proof. Suppose R is a closed rectangle such that S is contained in the interior of R. If $x \in \partial S$, then for every $\delta > 0$, the sets $S \cap B(x, \delta)$ (where χ_S is 1) and the sets $(R \setminus S) \cap B(x, \delta)$ (where χ_S is 0) are both nonempty. So χ_S is not continuous at x. If x is either in the interior of S or in the complement of the closure $\overline{S}$, then χ_S is either identically 1 or identically 0 in a whole neighbourhood of x and hence χ_S is continuous at x. Therefore, the set of discontinuities of χ_S is precisely the boundary ∂S. The proposition then follows. □

Proposition 10.5.2. *Suppose S and T are bounded Jordan measurable sets. Then*

(i) The closure $\overline{S}$ is Jordan measurable.

(ii) The interior S° is Jordan measurable.

(iii) $S \cup T$ is Jordan measurable.

(iv) $S \cap T$ is Jordan measurable.

(v) $S \setminus T$ is Jordan measurable.

The proof of the proposition is left as an exercise. Next, we find that the volume that we defined above coincides with the outer measure we defined above.

Proposition 10.5.3. *If $S \subset \mathbb{R}^n$ is Jordan measurable then $V(S) = m^*(S)$.*

Proof. Given $\varepsilon > 0$, let R be a closed rectangle that contains S. Let P be a partition of R such that

$$U(P,\chi_S) \leq \int_R \chi_S + \varepsilon = V(S) + \varepsilon \qquad \text{and} \qquad L(P,\chi_S) \geq \int_R \chi_S - \varepsilon = V(S) - \varepsilon.$$

Let $R_1, \ldots, R_k$ be all the subrectangles of P such that χ_S is not identically zero on each R_j. That is, there is some point $x \in R_j$ such that $x \in S$. Let O_j be an open rectangle such that $R_j \subset O_j$ and $V(O_j) < V(R_j) + \varepsilon/k$. Notice that $S \subset \bigcup_j O_j$. Then

$$U(P,\chi_S) = \sum_{j=1}^{k} V(R_k) > \left(\sum_{j=1}^{k} V(O_k) \right) - \varepsilon \geq m^*(S) - \varepsilon.$$

As $U(P,\chi_S) \leq V(S) + \varepsilon$, then $m^*(S) - \varepsilon \leq V(S) + \varepsilon$, or in other words $m^*(S) \leq V(S)$.

Now let $R'_1, \ldots, R'_\ell$ be all the subrectangles of P such that χ_S is identically one on each R'_j. In other words, these are the subrectangles contained in S. The interiors of the subrectangles R'°_j are disjoint and $V(R'^{\circ}_j) = V(R'_j)$. It is easy to see from definition that

$$m^*\Big(\bigcup_{j=1}^{\ell} R'^{\circ}_j \Big) = \sum_{j=1}^{\ell} V(R'^{\circ}_j).$$

Hence

$$m^*(S) \geq m^*\Big(\bigcup_{j=1}^{\ell} R'_j \Big) \geq m^*\Big(\bigcup_{j=1}^{\ell} R'^{\circ}_j \Big) = \sum_{j=1}^{\ell} V(R'^{\circ}_j) = \sum_{j=1}^{\ell} V(R'_j) = L(P,f) \geq V(S) - \varepsilon.$$

Therefore $m^*(S) \geq V(S)$ as well. □

10.5.2 Integration over Jordan measurable sets

In one variable there is really only one type of reasonable set to integrate over: an interval. In several variables we have many very simple sets we might want to integrate over and these cannot be described so easily.

Definition 10.5.4. Let $S \subset \mathbb{R}^n$ be a bounded Jordan measurable set. A bounded function $f \colon S \to \mathbb{R}$ is said to be Riemann integrable on S, or $f \in \mathscr{R}(S)$, if for a closed rectangle R such that $S \subset R$, the function $\widetilde{f} \colon R \to \mathbb{R}$ defined by

$$\widetilde{f}(x) = \begin{cases} f(x) & \text{if } x \in S, \\ 0 & \text{otherwise,} \end{cases}$$

is in $\mathscr{R}(R)$. In this case we write

$$\int_S f := \int_R \widetilde{f}.$$

When f is defined on a larger set and we wish to integrate over S, then we apply the definition to the restriction $f|_S$. In particular note that if $f\colon R \to \mathbb{R}$ for a closed rectangle R, and $S \subset R$ is a Jordan measurable subset then

$$\int_S f = \int_R f\chi_S.$$

Proposition 10.5.5. *If $S \subset \mathbb{R}^n$ is a Jordan measurable set and $f\colon S \to \mathbb{R}$ is a bounded continuous function, then f is integrable on S.*

Proof. Define the function $\widetilde{f}$ as above for some closed rectangle R with $S \subset R$. If $x \in R \setminus \overline{S}$, then $\widetilde{f}$ is identically zero in a neighbourhood of x. Similarly if x is in the interior of S, then $\widetilde{f} = f$ on a neighbourhood of x and f is continuous at x. Therefore, $\widetilde{f}$ is only ever possibly discontinuous at ∂S and we are finished □

10.5.3 Images of Jordan measurable subsets

Finally, images of Jordan measurable sets are Jordan measurable under nice enough mappings. The following version will be enough.

Proposition 10.5.6. *Suppose $S \subset \mathbb{R}^n$ is a closed bounded Jordan measurable set, and $S \subset U$ for an open set $U \subset \mathbb{R}^n$. Suppose $g\colon U \to \mathbb{R}^n$ is a one-to-one continuously differentiable mapping such that J_g is never zero on S. Then $g(S)$ is Jordan measurable.*

Proof. Let $T = g(S)$. We claim that the boundary ∂T is contained in the set $g(\partial S)$. Suppose the claim is proved. As S is Jordan measurable, then ∂S is measure zero. Then $g(\partial S)$ is measure zero by Proposition 10.3.9. As $\partial T \subset g(\partial S)$, then T is Jordan measurable.

It is therefore left to prove the claim. First, S is closed and bounded and hence compact. By Lemma **??**, $T = g(S)$ is also compact and therefore closed. In particular $\partial T \subset T$. Suppose $y \in \partial T$, then there must exist an $x \in S$ such that $g(x) = y$. The Jacobian of g is nonzero at x.

We now use the inverse function theorem Theorem 8.5.1. We find a neighbourhood $V \subset U$ of x and an open set W such that the restriction $f|_V$ is a one-to-one and onto function from V to W with a continuously differentiable inverse. In particular $g(x) = y \in W$. As $y \in \partial T$, there exists a sequence $\{y_k\}$ in W with $\lim y_k = y$ and $y_k \notin T$. As $g|_V$ is invertible and in particular has a continuous inverse, there exists a sequence $\{x_k\}$ in V such that $g(x_k) = y_k$ and $\lim x_k = x$. Since $y_k \notin T = g(S)$, clearly $x_k \notin S$. Since $x \in S$, we conclude that $x \in \partial S$. The claim is proved, $\partial T \subset g(\partial S)$. □

10.5.4 Exercises

Exercise* 10.5.1: Prove Proposition 10.5.2.*

Exercise* 10.5.2*: *Prove that a bounded convex set is Jordan measurable. Hint: induction on dimension.*

Exercise* 10.5.3*: *Let $f\colon [a,b] \to \mathbb{R}$ and $g\colon [a,b] \to \mathbb{R}$ be continuous functions and such that for all $x \in (a,b)$, $f(x) < g(x)$. Let*

$$U := \{(x,y) \in \mathbb{R}^2 : a < x < b \text{ and } f(x) < y < g(x)\}.$$

a) Show that U is Jordan measurable.
b) If $f\colon U \to \mathbb{R}$ is Riemann integrable on U, then

$$\int_U f = \int_a^b \int_{g(x)}^{f(x)} f(x,y)\,dy\,dx.$$

10.6 Green's theorem

Note: 1 lecture

One of the most important theorems of analysis in several variables is the so-called generalized Stokes' theorem, which is a generalization of the fundamental theorem of calculus. Perhaps the most often used version is the version in two dimensions which is called Green's theorem, which we will prove here.

Definition 10.6.1. Let $U \subset \mathbb{R}^2$ be a bounded connected open set. Suppose the boundary ∂U is a finite union of (the images of) simple piecewise smooth paths such that near each point $p \in \partial U$ every neighbourhood V of p contains points of $\mathbb{R}^2 \setminus \overline{U}$. Then U is called a *bounded domain with piecewise smooth boundary* in $\mathbb{R}^2$.

The condition about points outside the closure means that locally ∂U separates $\mathbb{R}^2$ into "inside" and "outside". The condition prevents ∂U from being just a "cut" inside U. Therefore as we travel along the path in a certain orientation, there is a well defined left and a right, and either it is U on the left and the complement of U on the right, or vice-versa. Thus by orientation on U we mean the direction along which we travel along the paths. It is easy to switch orientation if needed by reparametrizing the path.

Definition 10.6.2. If $U \subset \mathbb{R}^2$ is a bounded domain with piecewise smooth boundary, then ∂U is *positively oriented* if as we travel along the path U is on the left.

If for example U is a bounded region with "no holes", that is ∂U is connected, then the positive orientation means we are travelling counterclockwise around ∂U. If we do have "holes", then we travel around them clockwise.

Proposition 10.6.3. *Let $U \subset \mathbb{R}^2$ be a bounded domain with piecewise smooth boundary, then U is Jordan measurable.*

Proof. We need that ∂U is of measure zero. As ∂U is a finite union of simple piecewise smooth paths, which themselves are finite unions of smooth paths we need only show that a smooth path is of measure zero in $\mathbb{R}^2$.

Let $\gamma \colon [a,b] \to \mathbb{R}^2$ be a smooth path. As γ is continuously differentiable, the derivative is bounded as $[a,b]$ is compact so there is some $M \in \mathbb{R}$ such that $\|\gamma'(t)\| \leq M$ for all $t \in [a,b]$. Further if we show that the image $\gamma\big((a,b)\big)$ is measure zero, then adding two points to a measure zero set still gets a measure zero set.

We will use Proposition 10.3.9, although we then need a mapping from an open set in $\mathbb{R}^2$. Define

$$f \colon (a,b) \times (-1,1) \to \mathbb{R}^2, \qquad \text{as} \qquad f(x,y) := \gamma(x).$$

As $(a,b) \times \{0\}$ is measure zero in $\mathbb{R}^2$ and $\gamma\big((a,b)\big) = f\big((a,b) \times \{0\}\big)$. Hence by Proposition 10.3.9, $\gamma\big((a,b)\big)$ is measure zero in $\mathbb{R}^2$ and so $\gamma\big([a,b]\big)$ is also measure zero, and so finally ∂U is also measure zero. $\square$

Theorem 10.6.4 (Green)**.** *Suppose $U \subset \mathbb{R}^2$ is a bounded domain with piecewise smooth boundary with the boundary positively oriented. Suppose P and Q are continuously differentiable functions defined on some open set that contains the closure $\overline{U}$. Then*

$$\int_{\partial U} P\,dx + Q\,dy = \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right).$$

We stated Green's theorem in general, although we will only prove a special version of it. That is, we will only prove it for a special kind of domain. The general version follows from the special case by application of further geometry, and cutting up the general domain into smaller domains on which to apply the special case. We will not prove the general case.

Let $U \subset \mathbb{R}$ be a domain with piecewise smooth boundary. We say U is of *type I* if there exist numbers $a < b$, and continuous functions $f\colon [a,b] \to \mathbb{R}$ and $g\colon [a,b] \to \mathbb{R}$, such that

$$U := \{(x,y) \in \mathbb{R}^2 : a < x < b \text{ and } f(x) < y < g(x)\}.$$

Similarly U is of *type II* if there exist numbers $c < d$, and continuous functions $h\colon [c,d] \to \mathbb{R}$ and $k\colon [c,d] \to \mathbb{R}$, such that

$$U := \{(x,y) \in \mathbb{R}^2 : c < y < d \text{ and } h(y) < x < k(y)\}.$$

Finally $U \subset \mathbb{R}$ is of *type III* if it is both of type I and type II.

If we have a type I domain, then the paths given by the graphs of f and g together with the straight line segments between $g(a)$ and $f(a)$, and also between $g(b)$ and $f(b)$ give the boundary as a piecewise smooth path. Similarly for type II, and hence for type II. Therefore such domains are domains with piecewise smooth boundary.

Let us state the version which we will prove.

Theorem 10.6.5 (Green)**.** *Let $U \subset \mathbb{R}^2$ be a domain of type III. Suppose P and Q are continuously differentiable functions defined on some open set that contains the closure $\overline{U}$ and suppose that ∂U is positively oriented. Then*

$$\int_{\partial U} P\,dx + Q\,dy = \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right).$$

Proof. By Exercise 10.5.3, U is Jordan measurable and as U is of type I, then

$$\begin{aligned}
\int_U \left(-\frac{\partial P}{\partial y} \right) &= \int_a^b \int_{g(x)}^{f(x)} \left(-\frac{\partial P}{\partial y}(x,y) \right) dy\,dx \\
&= \int_a^b \Big(-P\big(x, f(x)\big) + P\big(x, g(x)\big) \Big)\, dx \\
&= \int_a^b P\big(x, g(x)\big)\, dx - \int_a^b P\big(x, f(x)\big)\, dx.
\end{aligned}$$

The one form $P\,dx$ when integrated integrates to nothing when integrating along the straight vertical lines in the boundary. Therefore it only is integrated along the top and along the bottom. As a parameter, x runs from left to right. If we use the parametrization that takes x to $\big(x, f(x)\big)$ and $\big(x, g(x)\big)$ we recognize path integrals above. However the second path integral is in the wrong direction, the top should be going right to left, and so we must switch orientation.

$$\int_U \left(-\frac{\partial P}{\partial y}\right) = \int_a^b P\big(x, g(x)\big)\,dx + \int_b^a P\big(x, f(x)\big)\,dx = \int_{\partial U} P\,dx.$$

Simiarly U is also of type II. The form $Q\,dy$ integrates to zero along horizontal lines. So

$$\int_U \frac{\partial Q}{\partial x} = \int_c^d \int_{k(y)}^{h(y)} \frac{\partial Q}{\partial x}(x,y)\,dx\,dy = \int_a^b \Big(Q\big(y, h(y)\big) - Q\big(y, k(y)\big)\Big)\,dx = \int_{\partial U} Q\,dy.$$

Putting the two together we obtain

$$\int_{\partial U} P\,dx + Q\,dy = \int_{\partial U} P\,dx + \int_{\partial U} Q\,dy = \int_U \Big(-\frac{\partial P}{\partial y}\Big) + \int_U \frac{\partial Q}{\partial x} = \int_U \Big(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\Big). \qquad \square$$

Let us illustrate the usefulness of Green's theorem on a fundamental result about harmonic functions.

Example 10.6.6: Suppose $U \subset \mathbb{R}^2$ is an open set and $f\colon U \to \mathbb{R}$ is harmonic, that is, f is twice continuously differentiable and $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$. We will prove one of the most fundamental properties of Harmonic functions.

Let $D_r = C(p,r)$ be a closed disc such that $C(p,r) \subset U$. Write $p = (x_0, y_0)$. We orient ∂D_r positively Then

$$\begin{aligned}
0 &= \frac{1}{2\pi r}\int_{D_r} \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) \\
&= \frac{1}{2\pi r}\int_{\partial D_r} -\frac{\partial f}{\partial y}\,dx + \frac{\partial f}{\partial x}\,dy \\
&= \frac{1}{2\pi r}\int_0^{2\pi} \Bigg(-\frac{\partial f}{\partial y}\big(x_0 + r\cos(t), y_0 + r\sin(t)\big)\big(-r\sin(t)\big) \\
&\qquad\qquad + \frac{\partial f}{\partial x}\big(x_0 + r\cos(t), y_0 + r\sin(t)\big) r\cos(t)\Bigg)\,dt \\
&= \frac{d}{dr}\left[\frac{1}{2\pi}\int_0^{2\pi} f\big(x_0 + r\cos(t), y_0 + r\sin(t)\big)\,dt\right].
\end{aligned}$$

Therefore as a function of r, we let $g(r) := \frac{1}{2\pi}\int_0^{2\pi} f\big(x_0 + r\cos(t), y_0 + r\sin(t)\big)\,dt$, and then $g'(r) = 0$ for all $r > 0$. As

$$g(0) = \frac{1}{2\pi}\int_0^{2\pi} f\big(x_0 + 0\cos(t), y_0 + 0\sin(t)\big)\,dt = f(x_0, y_0).$$

The function is constant for $r > 0$ and continuous at $r = 0$ (exercise). Therefore $g(0) = g(r)$ for all $r > 0$. We have proved the *mean value property* of harmonic functions:

$$f(x_0, y_0) = \frac{1}{2\pi} \int_0^{2\pi} f\big(x_0 + r\cos(t), y_0 + r\sin(t)\big)\, dt = \frac{1}{2\pi r} \int_{\partial D_r} f\, ds.$$

That is, the value at $p = (x_0, y_0)$ is the average over a circle of any radius r centered at (x_0, y_0).

10.6.1 Exercises

Exercise* 10.6.1: Prove that a disc $B(p,r) \subset \mathbb{R}^2$ is a type III domain.*

Exercise* 10.6.2: For a disc $B(p,r) \subset \mathbb{R}^2$, orient the boundary $\partial B(p,r)$ positively:*
a) Compute $\int_{\partial B(p,r)} -y\, dx$.
b) Compute $\int_{\partial B(p,r)} x\, dy$.
c) Compute $\int_{\partial B(p,r)} \frac{-y}{2}\, dy + \frac{x}{2}\, dy$.

Exercise* 10.6.3: Using Green's theorem show that the area of a triangle with vertices (x_1, y_1), (x_2, y_2), (x_3, y_3) is $\frac{1}{2}\left|x_1y_2 + x_2y_3 + x_3y_1 - y_1x_2 - y_2x_3 - y_3x_1\right|$. Hint: see previous exercise.*

Exercise* 10.6.4: Using the mean value property prove the maximum value for harmonic functions: Suppose $U \subset \mathbb{R}^2$ is an connected open set and $f : U \to \mathbb{R}$ is harmonic. Suppose f attains a maximum at $p \in U$, then show that f is constant.*