Analysis II
Lecture notes

Christoph Thiele
(lectures 11,12 by Roland Donninger
lecture 22 by Diogo Oliveira e Silva) *
Summer term 2015
Universität Bonn

July 5, 2016

Contents

*Notes by Polona Durcik and Gennady Uraltsev

1 Analysis in several variables

1.1 Euclidean space $\mathbb{R}^d$

The Euclidean space $\mathbb{R}^d$ is the set of all functions $x : I_d \to \mathbb{R}$, where $I_d := \{0, 1, \dots, d-1\}$. We call the elements $x = (x(0), \dots, x(d-1))$ of $\mathbb{R}^d$ *ordered d-tuples of real numbers*. Many authors use a different convention, considering functions from $\{1, \dots, d\}$ to $\mathbb{R}$. The two concepts are easily seen equivalent. The operations

$$\begin{aligned}(x+y)(i) &= x(i) + y(i) \qquad x,\, y \in \mathbb{R}^d \\ (ax)(i) &= ax(i) \qquad x \in \mathbb{R}^d,\, a \in \mathbb{R}.\end{aligned}$$

yield on $\mathbb{R}^d$ the structure of a vector space over $\mathbb{R}$.
We equip $\mathbb{R}^d$ with the metric structure by defining the metric $\rho : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$$\rho(x, y) := \|x - y\|,$$

where $\|x\|$ is the *length* of x

$$\|x\| := \sqrt{\sum_{i=0}^{d-1} x_i^2}.$$

We call d the *canonical* or *Euclidean metric* or *distance*.

Note that if the dimension d equals to 1, we are on the real line $\mathbb{R}$. The length $\|x\|$ of $x \in \mathbb{R}$ is the usual absolute value $|x|$. If $d = 2, 3$, then the length coincides with the natural geometric length, as one can see from repeated use of the Pythagorean theorem applied to right angled triangles. This is illustrated in Figure 1.

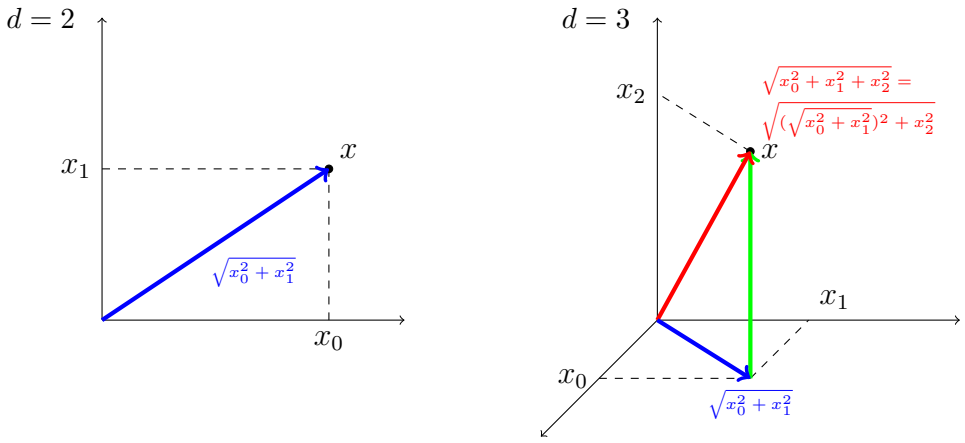


Figure 1: Pythagorean theorem.

We state some properties of the Euclidean metric.

1. For all $x \in \mathbb{R}^d$, $a \in \mathbb{R}$, $\|ax\| = |a|\|x\|$.
   Proof. We calculate

$$\sqrt{\sum_{i=0}^{d-1}(ax_i)^2} = \sqrt{a^2 \sum_{i=0}^{d-1} x_i^2} = \sqrt{a^2}\sqrt{\sum_{i=0}^{d-1} x_i^2} = |a|\|x\|. \quad \square$$

   Note that this in particular shows $\|x\| = \|-x\|$ and thus symmetry of the metric ρ. Also, by choosing $a = 0$ we obtain[1] $\|0\| = 0$.

2. For all $x \in \mathbb{R}^d$, $\|x\| = 0 \Rightarrow x = 0$.
   Proof.

$$\|x\| = 0 \Rightarrow \sum_{i=0}^{d-1} x_i^2 = 0 \Rightarrow \forall i : x_i^2 = 0 \Rightarrow \forall i : x_i = 0 \Rightarrow x = 0. \quad \square$$

[1] Here $\|0\|$ means the length of the null vector $0 = (0, \ldots, 0)$.

3. (Cauchy-Schwarz inequality) For all $x, y \in \mathbb{R}^d$,

$$\sum_{i=0}^{d-1} x_i y_i \leq \|x\|\|y\|.$$

Proof. Note that if $x = 0$ or $y = 0$, the inequality trivially holds. Thus it suffices to show that for all $x \neq 0, y \neq 0$,

$$\sum_{i=0}^{d-1} \frac{x_i}{\|x\|} \frac{y_i}{\|y\|} \leq 1.$$

In other words, since $\|\frac{x}{\|x\|}\| = 1$ it is enough to show that for x, y with $\|x\| = \|y\| = 1$,

$$\sum_{i=0}^{d-1} x_i y_i \leq 1. \tag{1}$$

For every i we have

$$0 \leq (x_i - y_i)^2 = x_i^2 + y_i^2 - 2x_i y_i$$

and hence

$$x_i y_i \leq \frac{1}{2} x_i^2 + \frac{1}{2} y_i^2.$$

Inserting this into (1) we obtain

$$\sum_{i=0}^{d-1} x_i y_i \leq \sum_{i=0}^{d-1} \frac{1}{2} x_i^2 + \frac{1}{2} y_i^2 = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 = 1. \quad \square$$

4. (Triangle inequality) For all $x, y \in \mathbb{R}^d$, $\|x + y\| \leq \|x\| + \|y\|$.
Proof. We compute

$$\begin{aligned} \|x + y\|^2 &= \sum_{i=0}^{d-1} (x_i + y_i)^2 = \sum_{i=0}^{d-1} (x_i^2 + y_i^2 + 2x_i y_i) \\ &= \|x\|^2 + \|y\|^2 + 2 \sum_{i=0}^{d-1} x_i y_i \end{aligned}$$

Using the Cauchy-Schwarz inequality we estimate this by

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

The claim follows by taking the square root on both sides. $\square$

Remark. Let X be a vector space over $\mathbb{R}$. A map $\|\cdot\| : X \to [0,\infty)$ satisfying (1), (2) and (4) is called a *norm*. A norm on X induces a metric by setting

$$\rho(x,y) := \|x-y\|.$$

Note that this is exactly what we have done in the Euclidean case.

1.2 Metric spaces

In all of the following X will be a metric space with the metric induced by a norm. Most of what we have to say easily extends to more general metric spaces.

Definition 1.1. The *open ball centered at $x \in X$ of radius ε* is the set

$$B_\varepsilon(x) := \{y \in X : \|x-y\| < \varepsilon\}.$$

Definition 1.2. A set $A \subseteq X$ is called *open* if for each $x \in A$ there is an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$.

Example. The interval $(0,1)$ is open in $X = \mathbb{R}$, while $[0,1)$ is not.

Definition 1.3. A set $A \subseteq X$ is called *closed* if $X \setminus A$ is open.

Example. For $x \in X$ and $\varepsilon > 0$, the set

$$\{y \in X : \|x-y\| \leq \varepsilon\} \tag{2}$$

is closed. To see this, let $z \in X \setminus A$. We need to find an open ball $B_\delta(z)$ with $B_\delta(z) \subseteq X \setminus A$. Since $\|z-x\| > \varepsilon$, we find a $\delta > 0$ such that $\|z-x\| > \varepsilon + \delta$. We claim that for this δ we have $B_\delta(z) \subseteq X \setminus A$. Indeed, let $a \in B_\delta(z)$. Then $\|a-z\| < \delta$ and hence

$$\|x-a\| \geq \|x-z\| - \|a-z\| > \varepsilon + \delta - \delta = \varepsilon.$$

The set (1.15) is also called *the closed ball at x of radius ε*.

The notions "open" and "closed" depend on the ambient space X, as can be seen from the following example.

Example. The interval $[0,1]$ is open in $X = [0,1]$, since $B_5(x) \subseteq X$ for all $x \in [0,1]$. The interval $(0,1)$ is closed in $X = (0,1)$. Indeed, note that the empty set is open.

Now we turn to the notion of convergence in X.

Definition 1.4. A sequence x_n in X is called *convergent*, if there exists an $x \in X$ with

$$\limsup_{n\to\infty} \|x_n - x\| = 0.$$

We also say that x_n *converges* to x. The element x is called the *limit* of x_n.

In a metric space, a sequence can have at most one limit, we leave this observation as an exercise.

Lemma 1.5. *A sequence x_n converges to $x \in X$ if and only if for every $\varepsilon > 0$ there is an n such that for all $m > n$ we have $x_m \in B_\varepsilon(x)$.*

We leave the proof as an exercise. In $\mathbb{R}^d$ we have an issue with double indices. We wrote x_i $i = 0, \dots d-1$ for the components of a vector, while we also wrote x_n for the members of a sequence $\mathbb{N} \to X$. The following text will have to be read carefully to avoid misunderstanding. We write $x_{n,i}$ for the i-th components of the n-th member of a sequence.

Lemma 1.6. *A sequence $x_n = (x_{n,0}, \dots, x_{n,d-1})$ in $\mathbb{R}^d$ converges to x if and only if for all i, $x_{n,i}$ converges to x_i, i.e.*

$$\forall i : \limsup_{n\to\infty} |x_{n,i} - x_i| = 0.$$

Proof. ($\Rightarrow$) We estimate

$$\limsup_{n\to\infty} |x_{n,i} - x_i| \le \limsup_{n\to\infty} \|x_n - x\| = 0.$$

($\Leftarrow$) We have

$$\begin{aligned}
\limsup_{n\to\infty} \|x_n - x\| &= \limsup_{n\to\infty} \sqrt{\sum_{i=0} (x_{n,i} - x_i)^2} \\
&\le d \limsup_{n\to\infty} \max_{0\le i\le d-1} |x_{n,i} - x_i| \\
&\le d \limsup_{n\to\infty} \sum_{i=0}^{d-1} |x_{n,i} - x_i| \\
&\le d \sum_{i=0}^{d-1} \limsup_{n\to\infty} |x_{n,i} - x_i| = d \cdot 0 = 0.
\end{aligned}$$

□

Theorem 1.7. *A set $A \subseteq X$ is closed if and only if for every convergent sequence in A the limit $x \in X$ lies in A.*

Proof. ($\Rightarrow$) Let $A \subseteq X$ be closed, i.e. $X \setminus A$ open. Let x_n be a convergent sequence with elements in A and let x be its limit. We need to show $x \in A$. Suppose the contrary that $x \in X \setminus A$. Thus there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq X \setminus A$. Since x_n converges to x, there exists an n such that for all $m > n$, $x_m \in B_\varepsilon(x)$. In particular, $x_{n+1} \in B_\varepsilon(x) \subseteq X \setminus A$. A contradiction to $x_{n+1} \in A$.

($\Leftarrow$) We need to show that $X \setminus A$ is open. Let $y \in X \setminus A$. We have to find an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq X \setminus A$. We again show the claim by contradiction. Suppose there is no such ball. Then for all $k \in \mathbb{N}_{>0}$ there is an $x_k \in A$ with $x_k \in B_{\frac{1}{k}}(y)$. Convince yourself that this implies that x_k converges to y, hence $y \in A$. Contradiction. □

Definition 1.8. The *closure* $\overline{A}$ of a set $A \subseteq X$ is the set of all $x \in X$ for which there is a sequence in A which converges to x.

As an exercise one can show that $\overline{A} \supseteq A$. Moreover, $A = \overline{A}$ if and only if A is closed. Observe also that $\overline{\overline{A}} = \overline{A}$ and that $X \setminus \overline{A}$ is open. The set $X \setminus \overline{A}$ is also called the *interior* of $X \setminus A$.

Definition 1.9. A metric space X is called *separable* if there is a countable set $\mathcal{A}$ of open balls with the following property: every open set in X is the union of the elements of a subset of $\mathcal{A}$.

End of lecture 1. April 9, 2015

This section is dedicated to discussing some relevant concepts relating to metric spaces. Previously we considered metrics induced by norms. In a more general context we introduce the following definition.

Definition 1.10. A metric space X is a set with a mapping $\rho : X \times X \to \mathbb{R}_{\geq 0}$that satisfies the following properties:

- $\forall x, y \in X \quad \rho(x; y) = \rho(y; x)$, this property is called symmetry;
- $\forall x, y \in X \quad \rho(x; y) = 0 \iff x = y$;
- $\forall x, y, z \in X \quad \rho(x; z) \leq \rho(x, y) + \rho(y, z)$, this property is called the triangle inequality.

Important properties that a metric space can possess is that of being *separable, complete,* and *compact.*

1.2.1 Separability

Definition 1.11. A metric space X is said to be *separable* if there exists a countable set of open balls $\mathcal{A}$ so that any open set of the space can be written as a union of a subset of balls from $\mathcal{A}$.

Definition 1.12. A subset $Y \subseteq X$ of a metric space X is said to be dense if all elements of X are limits of a sequence of elements of Y.

We are already familiar with a dense set of $\mathbb{R}_{\geq 0}$: the dyadic numbers $\mathbb{Y} = \left\{ \frac{n}{2^m} \text{ with } n \in \mathbb{N},\, m \in \mathbb{Z} \right\}$. The two notions we have just introduced are closely related by the following theorem.

Theorem 1.13. *A metric space X is separable if and only if it has a countable dense subset.*

Proof.

$\Rightarrow$ Let X be separable and let $\mathcal{A}$ be a countable set of open balls as in Definition 1.11. Let Y be the set of all the centers of the balls in $\mathcal{A}$; we must show that Y is dense in X. Let $x \in X$ be a given point; for each $n \in \mathbb{N}_{>0}$ consider the open ball $B_{\frac{1}{n}}(x)$. Being a non-empty open set (it contains at least x itself) it can be represented as a non-empty union of a subset of balls from $\mathcal{A}$. Let us call one of the balls used in this representation by $B_{\epsilon_n}(y_n)$ so that we have $B_{\epsilon_n}(y_n) \subseteq B_{\frac{1}{n}}(x)$ with $\epsilon_n \in \mathbb{R}_{>0}$ and $y_n \in Y$. It is now sufficient to show that the sequence y_n converges to x i.e. $\limsup_{n \to +\infty} \rho(y_n; x) = 0$. For any $\delta > 0$ choose $n \in \mathbb{N}$ so that $\frac{1}{n} < \delta$, then $\forall m > n$ we have that $\rho\,(y_m;\, x) < \frac{1}{n} \leq \delta$ as required.

$\Leftarrow$ Let Y be a countable dense subset of X. Set $\mathcal{A} = \left\{ B_{\frac{1}{n}}(y) \text{ with } y \in Y,\, n \in \mathbb{N}_{\geq 1} \right\}$; it is clear that $\mathcal{A}$ is countable. The rest is left as an exercise.

□

Theorem 1.14. *The Euclidean space $\mathbb{R}^d$ is separable.*

Proof. The proof consist of several steps. The proof for any $d \in \mathbb{N}$ can be done by induction and is left as an exercise. Here we limit ourselves to proving the statement for $d = 2$.

- $\mathbb{R}_{\geq 0}$is separable. This is true because the set of dyadic numbers $\mathbb{Y}$ is a countable dense subset of $\mathbb{R}_{\geq 0}$.
- $\mathbb{R}$ is separable since the set $Y = \mathbb{Y} \cup -\mathbb{Y}$ is countable and dense in $\mathbb{R}$.

- $\mathbb{R}^2$ is separable since the set $Y^2 = \{(y_0, y_1) \text{ with } y_0 \in Y\, y_1 \in Y\}$ is countable and dense in $\mathbb{R}^2$. To see this, for any $\delta > 0$ choose $n \in \mathbb{N}$ such that $\frac{1}{n} < \delta$ and then choose points $y_{n,i} \in Y$ so that $|y_{n,i} - x_i| < \frac{1}{2n}$ with $i \in \{0, 1\}$. We have that $\|y_n - x\| \leq |y_{n,0} - x_0| + |y_{n,1} - x_1| \leq \frac{1}{n}$ and this concludes the proof.

□

1.2.2 Completeness

Another very important property of metric spaces is completeness. To state this property we need to introduce the concept of closed balls.

Definition 1.15. A closed ball $\overline{B}_\epsilon(x) \subset X$ of a metric space X is the set $\overline{B}_\epsilon(x) = \{y \in X\ \rho(y; x) \leq \epsilon\}$.

Definition 1.16. A metric space X is *complete* if for any set of closed balls $\mathcal{A}$ with the properties

1. that for any two balls of $\mathcal{A}$, one is included in the other, and
2. that for every $\epsilon > 0$ there is ball of radius at most ϵ in $\mathcal{A}$,

the intersection of all the balls in $\mathcal{A}$ is non-empty.

The positive real line $\mathbb{R}_{\geq 0}$ is an example of a complete space. Here the completeness follows from the property that any set of points has a supremum and an infimum. As a matter of fact we know that any closed ball is a closed interval $\overline{B}_\epsilon(x) = [x - \epsilon, x + \epsilon]$. Given a collection of closed balls $\mathcal{A}$ let $a = \sup_{\overline{B}_\epsilon(x) \in \mathcal{A}}(x - \epsilon)$ and $b = \inf_{\overline{B}_\epsilon(x) \in \mathcal{A}}(x + \epsilon)$ where upper and lower bounds are taken over all the balls in $\mathcal{A}$. It is clear that the interval $[a, b]$ lies in the intersection of all the balls and since it is easy to check that $a \leq b$ the interval is non-empty. Notice, however, that the condition that the balls are closed is of crucial importance for this reasoning to hold.

Theorem 1.17. *A metric space X is complete if and only if any Cauchy sequence has a limit in X.*

Proof. □

⇒ Let X be a complete metric space; a sequence x_n is a Cauchy sequence if $\forall m \in \mathbb{N}\ \exists n(m) \in \mathbb{N}$ such that $\forall n', n'' \geq n(m)$ we have that $\rho\left(x_{n'}; x_{n''}\right) \leq 2^{-m}$. Consider the sets $U_m = \overline{B}_{10\,2^{-m}}\left(x_{n(m)}\right)$; we need to show that

they are ordered by inclusion as required. Consider some $m' > m$; $\forall z \in \overline{B}_{10\,2^{-m'}}\left(x_{n(m')}\right)$ we have that $\rho\left(x_{n(m)};\, x_{n(m')}\right) \leq 2^{-m}$ by applying the Cauchy sequence hypothesis starting from index $\min\left(n(m);\, n(m')\right)$. Thus $\rho\left(x_{n(m)};\, z\right) \leq \rho\left(x_{n(m)};\, x_{n(m')}\right) + \rho\left(x_{n(m')};\, z\right) \leq 2^{-m} + 10\,2^{-m'} \leq 6\,2^{-m}$, so $y \in \overline{B}_{10\,2^{-m}}\left(x_{n(m)}\right)$ and we have shown that the $U_{m'} \subset U_m$ as required. Let $y \in \bigcap_{m \in N_{\geq 1}} U_m$, checking that $\lim_{n \to +\infty} x_n = y$ is left as an exercise.

$\Leftarrow$ The proof that if all Cauchy sequences in a metric space X have a limit in X then the space X is complete as per Definition 1.16 is left to the reader.

In the above statement the fact that we consider closed and not open balls is crucial. On the other hand, while it does not matter for the above statement, there is a difference between our Definition 1.15 of a closed ball and that of the closure of the open ball of the same radius. The closure of a set A written as $\overline{A}$ is the set of all limit points of sequences of elements in A. The interior of A indicated as A° is the union of all the open balls contained in A. For any metric space X we have the relation $\overline{A} = X \setminus ((X \setminus A)^\circ)$. Notice that for a general metric space X we have that the closure of an open ball $\overline{B_\epsilon(x)}$ is contained but may not coincide with the closed ball of the same radius $\overline{B}_\epsilon(x)$. As an example consider $X = \mathbb{R}_{\geq 0} \cup \{-1\}$ as a subset of $\mathbb{R}$ with the same distance: the closed ball is $\bar{B}_1(0) = \{-1\} \cup [0,\, 1]$ while for the closure of the open ball of the same radius we have $-1 \notin \overline{B_1(0)}$ since there is no sequence in $B_1(0) \subset X$ converging to -1.
Let us now return to the properties of complete metric spaces.

Theorem 1.18. *The Euclidean space $\mathbb{R}^d$ is a complete metric space.*

Proof. Once again we restrict ourselves to the proof for the case $d = 2$.

- $\mathbb{R}$ is a complete metric space. This is due to the fact that all Cauchy sequences in $\mathbb{R}$ have a limit (as seen in the course of Analysis 1).

- For $\mathbb{R}^2$ we need to show that all Cauchy sequences have a limit in $\mathbb{R}^2$. Consider such a sequence x_n and the first and second coordinates $x_{n,0},\, x_{n,1} \in \mathbb{R}$. Since $|x_{n,i} - x_{m,i}| \leq \|x_n - x_m\|$ for $i = 1,\, 2$ we have that $x_{n,i}$ are also Cauchy sequences but in $\mathbb{R}$ and as such converge to y_i, $i = 1,\, 2$ respectively. The proof of the fact that $(y_0,\, y_1) \in \mathbb{R}^2$ is the limit of x_n is left to the reader.

□

1.2.3 Compactness

The third important property that a metric space can have is *compactness*.

Definition 1.19. A subset $K \subset X$ of a metric space X is compact if for every set $\mathcal{A}$ of open sets of X such that $K \subset \bigcup_{A \in \mathcal{A}} A$ (such a set is called an open covering) one can find a *finite* subset $\mathcal{A}' \subset \mathcal{A}$ such that $K \subset \bigcup_{A \in \mathcal{A}'} A$.

Example. Any finite subset $K \subset X$ of a metric space X is compact. Given any covering $\mathcal{A}$ it is sufficient to select one open set $A \in \mathcal{A}$ for every element $x \in K$. Since there are only finitely many elements one needs to select only finitely many open sets A.

Similarly for the other properties we have encountered, the compactness of a subset, that is expressed in topological terms (open sets, coverings), has important implications on the behavior of sequences with elements in the subset. In particular we have the following sequential property that is equivalent to compactness in the case of metric spaces.

Theorem 1.20. *A subset $K \subset X$ of a metric space X is compact if and only if any sequence of elements in K has a subsequence that has a limit in K.*

First of all we will start with some basic properties: for a subset of a given metric space X to be compact it has to contain all its limit points and be bounded. The proof of this statement is left as an exercise.

Definition 1.21. A subset $A \subset X$ of a metric space X is bounded if it is contained in some ball of X i.e. $\exists B_R(x)$ such that $A \subset B_R(x)$.

If a subset $A \subset X$ is bounded then for any point $y \in X$ there is a ball centered in that point that contains A. As a matter of fact if $A \subset B_R(x)$ then $A \subset B_{R+\rho(x,y)}(y)$ by the triangle inequality.
Euclidean metric spaces have an explicit characterization of compact subsets. While being closed and bounded is necessary for a subset to be compact, in the case of Euclidean spaces it is also sufficient. This statement is known as the Heine-Borel Theorem.

Theorem 1.22. *All closed and bounded subsets of $\mathbb{R}^d$ are compact.*

Proof. As usual we will give the proof in the case $d = 2$. The more general case is based on the same argument. Let $K \subset \mathbb{R}^2$ be a closed and bounded set. We will use the characterization of compactness via sequences given by Theorem 1.20. Let x_n be a sequence of elements in K then the sequences

$x_{n,i}$ with $i = 1, 2$ are also bounded because $|x_{n,i}| \leq \|x_n\| \leq R$ for some $R > 0$. Then let $x_{n_k,0}$ be a monotone subsequence of $x_{n,1}$ and let us select a further subsequence so that $x_{n_{k_l},1}$is also monotone. Since for $i = 1, 2$ $x_{n_{k_l},i} \in \mathbb{R}$ are monotone and bounded they have finite limit points y_i and so $\lim_{l\to+\infty} x_{n_{k_l}} = (y_0, y_1)$. The fact that $(y_0, y_1) \in K$ is due K being closed. □

Theorem 1.23. *A compact subset $K \subset X$ of a metric space X is closed and bounded.*

Proof. Left as an exercise. □

End of lecture 2. April 13, 2015

1.3 Hilbert spaces

Definition 1.24. A *(real) normed space* is a (real) vector space, on which a norm is defined.

Recall from Lecture 1 that a *norm* on a real vector space V is a map $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ satisfying

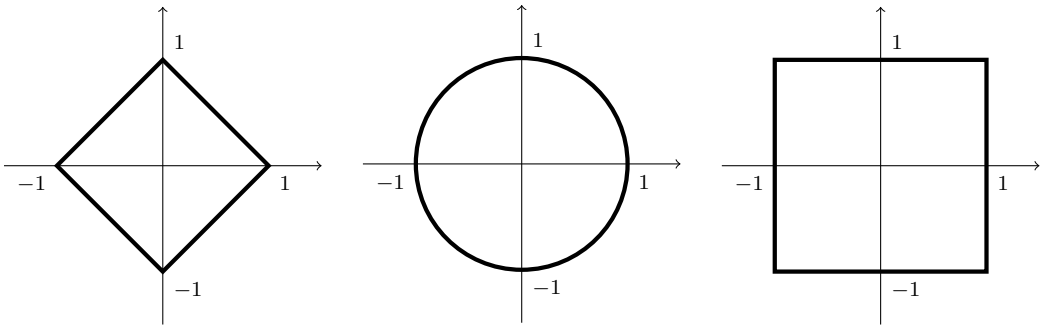
1. $\|x\| = 0 \Rightarrow x = 0$ for all $x \in V$
2. $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in V$, $\lambda \in \mathbb{R}$
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$

One can also consider more general normed spaces. For instance, if V is a vector space over $\mathbb{C}$, we talk about complex normed spaces. In this case one has to modify 2. in the definition of the norm to hold for all $\lambda \in \mathbb{C}$. However, for now we will only focus on real normed spaces.

Example. On $\mathbb{R}^2$ we may consider

- $\|x\| = \sqrt{|x_0| + |x_1|}$
- $\|x\| = \sqrt{x_0^2 + x_1^2}$
- $\|x\| = \max(|x_0|, |x_1|)$.

The second expression defines the Euclidean norm. It is easy to verify that the other two expressions also define a norm on $\mathbb{R}^2$.
The closed unit ball centered at 0 with respect to each of these norms can be seen in Figure 2.



$|x_0| + |x_1|$ $\qquad$ $(|x_0|^2 + |x_1|^2)^{1/2}$ $\qquad$ $\max(|x_0|, |x_1|)$

Figure 2: Unit balls in $\mathbb{R}^2$.

We have rewritten the Euclidean norm as $\sqrt{x_0^2 + x_1^2} = (|x_0|^2 + |x_1|^2)^{1/2}$. This way we see that replacing the exponent 2 by 1 yields the first norm. More generally, one can replace 2 by any exponent $1 \leq p < \infty$ and define the norm

$$\|x\| := (|x_0|^p + |x_1|^p)^{1/p}.$$

Definition 1.25. A normed vector space is called *Banach space*, if the induced metric is complete.

Recall that the induced metric is defined as $\rho(x, y) := \|x - y\|$.
Among the above examples, intuitively the case $p = 2$ produces the most "round" ball. The following is an algebraic condition which singles out the case $p = 2$ above, and thus can be viewed as a metric condition of "roundness" of the ball.

Definition 1.26. A real Banach space V is called *Hilbert space*, if for all $x, y \in V$,

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{3}$$

The identity (3) is called the *parallelogram law*. Namely, it can be interpreted as stating that the sum of squares of the lengths of the two diagonals of a parallelogram is equal to the sum of squares of the lengths of the four sides of a parallelogram.

We claim that on $\mathbb{R}^d$, the Euclidean norm $\|x\| = \left(\sum_{i=0}^{d-1} |x_i|^2\right)^{1/2}$ satisfies

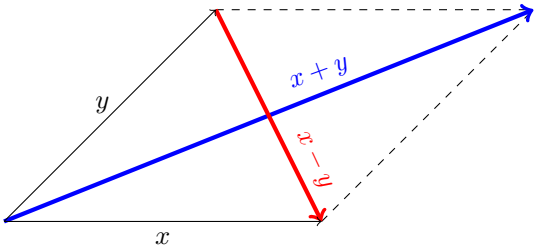


Figure 3: Parallelogram law.

the parallelogram law. Indeed, we have

$$\begin{aligned}\sum_{i=0}^{d-1}(x_i+y_i)^2+\sum_{i=0}^{d-1}(x_i-y_i)^2 =&\sum_{i=0}^{d-1} x_i^2+2x_iy_i+y_i^2+x_i^2-2x_iy_i+y_i^2\\ =&2\sum_{i=0}^{d-1}x_i^2+2\sum_{i=0}^{d-1}y_i^2.\end{aligned}$$

Our next goal is to show that in some sense the converse also holds. That is, if a norm on $\mathbb{R}^d$ satisfies the parallelogram law, it is up to a possible change of basis of the vector space equal to the Euclidean norm. First we turn our attention to some consequences of the parallelogram rule.

A closed subset A of a Hilbert space V is called *convex*, if for all $x, y \in A$ also $\frac{1}{2}(x+y) \in A$.

Theorem 1.27 (Projection theorem)**.** *Let V be a Hilbert space and $A \subset V$ closed and convex. Let $x \in V$. Then there exists $y \in A$ such that*

$$\|y-x\| = \inf_{z\in A}\|z-x\|.$$

Note that the condition of A being closed cannot be omitted. This can already be seen from the example $V = \mathbb{R}$ with norm being the absolute value, and $A = (0,1)$ and $x = 2$.

Proof. Set $r := \inf_{z\in A}\|z-x\|$. Let y_n be a sequence in A such that $r = \lim_{n\to\infty}\|y_n - x\|$. Let $\varepsilon > 0$. First we show that that y_n is Cauchy. Let n be such that for all $m \geq n$ we have

$$\|y_m - x\|^2 \leq r^2 + \frac{\varepsilon^2}{4}.$$

By (3) we have

$$\|y_m - y_n\|^2 + \|y_n + y_m - 2x\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2$$

and hence for all $m \geq n$

$$\begin{aligned}\|y_m - y_n\|^2 \leq & \left(2r^2 + \frac{\varepsilon^2}{2}\right) + \left(2r^2 + \frac{\varepsilon^2}{2}\right) - \|y_n + y_m - 2x\| \\ = & 4r^2 + \varepsilon^2 - 4\|\frac{y_n + y_m}{2} - x\|. \end{aligned} \tag{4}$$

By convexity $\frac{y_n+y_m}{2} \in A$. Thus we certainly have

$$\left\|\frac{y_n + y_m}{2} - x\right\| \geq r,$$

since r is the infimum of all $\|z - x\|$ for $z \in A$. Hence (4) is bounded by

$$\leq 4r^2 + \varepsilon^2 - 4r^2 = \varepsilon^2.$$

This shows that y_n is Cauchy.
Now let $y = \lim_{n\to\infty} y_n$, we know that such y exists since the Banach space is complete. Since A is closed, $y \in A$. It remains to show that $r = \|y - x\|$. By the triangle inequality

$$\|y - x\| - \|y_n - x\| \leq \|y_n - x\|.$$

Taking limits on both sides we obtain

$$\|y - x\| - r \leq 0 \Leftrightarrow \|y - x\| \leq r.$$

By definition of r we have $\|y - x\| \geq r$, and we conclude $\|y - x\| = r$ as desired. Note that the last steps in this proof could be done by referring to the continuity of the norm. □

Remark. The vector y from the previous theorem is unique. Indeed, suppose we have $y, y' \in V$ satisfying $\|y - x\| = r = \|y' - x\|$. Then

$$\|y - y'\|^2 + \|y + y' - 2x\|^2 = 2(\|y - x\|^2 + \|y' - x\|^2)$$

i.e.

$$\|y - y'\|^2 + 4\left\|\frac{y + y'}{2} - x\right\|^2 = 4r^2.$$

In the spirit of the previous proof we conclude that

$$\|y - y'\|^2 + 4r^2 \leq 4r^2$$

which implies $\|y - y'\|^2 \leq 0$ and thus $y = y'$.

We have the following generalization of the parallelogram rule.

Theorem 1.28. *Let V be a Hilbert space. Then for all $x, y \in V, \lambda \in \mathbb{R}$,*

$$\|x + \lambda y\|^2 - \|x^2\| - \|\lambda y\|^2 = \lambda(\|x + y\|^2 - \|x\|^2 - \|y\|^2). \tag{5}$$

Note that if $\lambda = -1$, this identity is exactly the parallelogram law (3). Note also that if $\lambda = 0$ or $\lambda = 1$, the identity trivially holds.

We can also interpret Theorem 1.28 in the following way. For a fixed x define the function

$$f(y) := \|x + y\|^2 - \|x\|^2 - \|y\|^2 \tag{6}$$

The theorem then says that

$$f(\lambda y) = \lambda f(y),$$

that is, the function f is homogeneous.

Proof. For $\lambda \in \mathbb{N}$ we show this by induction. As we said, for $\lambda = 0, 1$ the theorem holds. Assume now that we already know (5) for λ and we want to prove it for $\lambda + 1$. We use (3) on $x + \lambda y$ and y:

$$\|x + (\lambda + 1)y\|^2 + \|x + (\lambda - 1)y\|^2 = 2\|x + \lambda y\|^2 + 2\|y\|^2.$$

Now we add $-\|x\|^2 - \|(\lambda + 1)y\|^2$ on both sides, which gives

$$\begin{aligned} &\|x + (\lambda + 1)y\|^2 - \|x\|^2 - \|(\lambda + 1)y\|^2 \\ =& -(\lambda + 1)\|y\|^2 \\ &+ 2\|x + \lambda y\|^2 - \|x + (\lambda - 1)y\|^2 - \|x\|^2 - 2\|y\|^2 \end{aligned} \tag{7}$$

We apply the induction hypothesis on terms in (7) involving $x + \lambda y$ and $x + (\lambda - 1)y$, so as to express (7) as linear combination of the norms squared of $x + y$ and x and y. After a short calculation we see that the result equals

$$= (\lambda + 1)\|x + y\|^2 - (\lambda + 1)\|x\|^2 - (\lambda + 1)\|y\|^2,$$

which finishes the induction step and thus the proof for $\lambda \in \mathbb{N}$.
The claim for $\lambda \in \mathbb{Z}$ follows now by the observation that by the parallelogram law the quantity

$$\|x - \lambda y\|^2 - \|x^2\| - \|\lambda y\|^2$$

is the negative of the quantity

$$\|x + \lambda y\|^2 - \|x^2\| - \|\lambda y\|^2 \ .$$

Now we show (5) for $\lambda \in \mathbb{Q}$. Writing $\lambda = \frac{p}{q}, p, q \in \mathbb{Z}, q \neq 0$ we compute

$$\begin{aligned} &\|x + \frac{p}{q}y\|^2 - \|x\|^2 - \|\frac{p}{q}y\|^2 \\ =&\frac{1}{q^2}(\|qx + py\|^2 - \|qx\|^2 - \|py\|^2). \end{aligned}$$

Since we know the claim for $p \in \mathbb{Z}$, this equals

$$\frac{p}{q^2}(\|qx + y\|^2 - \|qx\|^2 - \|y\|^2).$$

Using the same argument for $q \in \mathbb{Z}$ and by symmetry in x, y we obtain

$$\frac{p}{q}(\|x + y\|^2 - \|x\|^2 - \|y\|^2).$$

It remains to prove the theorem for $\lambda \in \mathbb{R}$. For this we approximate λ with a sequence of rational numbers, for which we already know (5). The conclusion then follows by continuity of the norm. We leave details as an exercise. □

Theorem 1.29. *Let V be a Hilbert space and W a closed subspace of V. Let $x \in V$ be such that $\|x\| = \inf_{y \in W} \|x + y\|$. Then for all $y \in W$*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2. \tag{8}$$

By a subspace W we mean a vector subspace of V. The equality (8) can be seen as "half of the parallelogram identity": replacing y with $-y$ gives

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2. \tag{9}$$

Summing (8) and (9) we obtain the "full" identity (3). In particular, in a Hilbert space the identities (8) and (9) are equivalent and they are equivalent to

$$\|x - y\| = \|x + y\|$$

as well.
Figure 4 depicts the case when W is a subspace of $\mathbb{R}^2$ with the Euclidean norm. The angle between x and y is right-angled and (8) is the well-known Pythagorean theorem.

Proof. Let $y \in W$. By Theorem 1.28,

$$\|x + \lambda y\|^2 - \|x\|^2 - \lambda^2\|y\|^2 = \lambda(\|x + y\|^2 - \|x\|^2 - \|y\|^2)$$

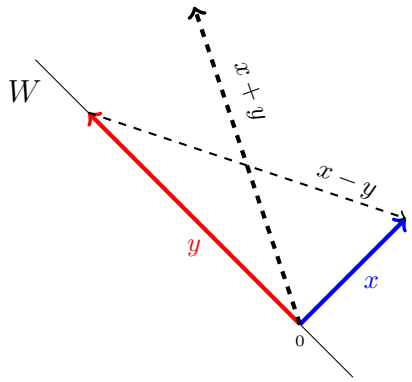


Figure 4: Orthogonality of x and y.

Since $\lambda y \in W$, by definition of x we have

$$\|x + \lambda y\|^2 - \|x\|^2 \geq 0.$$

Thus,

$$0 \leq \lambda^2 \|y\|^2 + \lambda(\|x + y\|^2 - \|x\|^2 - \|y\|^2).$$

The right hand side is a quadratic function in λ. Since $\lambda = 0$ is on of its zeroes and its leading term is positive, the right hand-side can be non-negative if and only if $\lambda = 0$ is a double zero. But this implies

$$\|x + y\|^2 - \|x\|^2 - \|y\|^2 = 0.$$

□

The preceding discussion motivates the following definition.

Definition 1.30. Let V be a Hilbert space. We say that $x, y \in V$ are *orthogonal*, if $\|x + y\| = \|x - y\|$.

End of lecture 3. April 16, 2015

The fact that x, y are orthogonal we shortly express by $x \perp y$. Note that orthogonality of x, y can be rephrased by saying $f(y) = 0$, where f is defined in (6). Homogeneity of f implies that for $\lambda \in \mathbb{R}$,

$$x \perp y \Rightarrow x \perp \lambda y.$$

Theorem 1.31 (Gram-Schmidt)**.** *Let V be a Hilbert space and let W be a $d-$dimensional subspace, $d \in \mathbb{N}_{\geq 1}$. Then there exists a basis $y_0, \ldots, y_{d-1}$ of W such that*

$$v = \sum_{i=0}^{d-1} \alpha_i y_i \Rightarrow \|v\| = \Big(\sum_{i=0}^{d-1} |\alpha_i|^2 \Big)^{\frac{1}{2}}.$$

Proof. We induct on d. Let $d = 1$. Pick $x_0 \neq 0$. It exists, since $d \neq 0$. Define

$$y_0 := \frac{x_0}{\|x_0\|}$$

and observe that $\|y_0\| = 1$. If $v = \alpha_0 y_0$, then

$$\|v\| = |\alpha_0| \|y_0\| = |\alpha_0| = \Big(\sum_{i=0}^{0} |\alpha_0|^2 \Big)^{1/2}.$$

Assume now that the theorem holds for $d \in \mathbb{N}_{\geq 1}$. Let W be a $(d+1)-$ dimensional subspace of V. Choose a basis $x_0, \ldots, x_d$ of W and set $W' = \operatorname{span}(x_0, \ldots, x_{d-1})$. We have $\dim W' = d$. By the induction hypothesis there exists a basis $y_0, \ldots, y_{d-1}$ of W' such that

$$v = \sum_{i=0}^{d-1} \alpha_i y_i \Rightarrow \|v\| = \Big(\sum_{i=0}^{d-1} |\alpha_i|^2 \Big)^{1/2}. \tag{10}$$

Now we would like to complete $y_0, \ldots, y_{d-1}$ to the desired basis of W. Consider the map from $\mathbb{R}^d$ to W' defined via

$$(\alpha_i)_i \mapsto \sum_{i=0}^{d-1} \alpha_i y_i, \tag{11}$$

which is a bijection. Moreover, by (10) it is an isometry, i.e.

$$\|(\alpha_i)_i\| = \left\| \sum_{i=0}^{d-1} \alpha_i y_i \right\|.$$

This implies that W' is complete, since $\mathbb{R}^d$ is complete. Then, W' is closed in V. Since W' is a subspace of V, it is also convex. By Theorem 1.27 there exists $y \in W'$ such that

$$\|x_d - y\| = \inf_{z \in W'} \|x_d - z\|.$$

Moreover, since for $z \in W'$ also $y + z \in W'$, we have

$$\|x_d - y\| = \inf_{z \in W'} \|x_d - y - z\|.$$

By Theorem 1.29 know $x_d - y \perp z$ for all $z \in W'$, and thus $\lambda(x_d - y) \perp z$ for all $z \in W', \lambda \in \mathbb{R}$. Define

$$y_d := \frac{x_d - y}{\|x_d - y\|}$$

Note that this is possible since from $x_d \notin W'$ it follows $x_d - y \neq 0$. By the above discussion also $y_d \perp z$. In particular, $y_d \perp y_l$ for all $l < d$.
Let now v be an arbitrary vector in W. Then we can write it as a linear combination

$$v = \sum_{i=0}^{d} \alpha_i y_i$$

where $\alpha_i \in \mathbb{R}$. By orthogonality of y_d to vectors in W' we have

$$\|v\|^2 = \left\| \sum_{i=0}^{d-1} \alpha_i y_i \right\|^2 + \|\alpha_d y_d\|^2$$

Since $\sum_{i=0}^{d-1} \alpha_i y_i \in W'$, by the induction hypothesis the last display equals

$$\Big(\sum_{i=0}^{d-1} |\alpha_i|^2 \Big)^{1/2} + |\alpha_d|^2 \|y_d\|^2 = \Big(\sum_{i=0}^{d} |\alpha_i|^2 \Big)^{1/2}$$

□

The bijection (11) is linear and thus an isomorphism. By this theorem, every finite dimensional Hilbert space is isometrically isomorphic to the Euclidean space $\mathbb{R}^d$.

The procedure described in the proof is also called *Gram-Schmidt orthogonalization*. A consequence of the proof is the following: If V is a Hilbert space and W a $d-$dimensional subspace, there exists a basis $y_0, \ldots, y_{d-1}$ of W with

$$\begin{aligned} &\|y_k\| = 1 \text{ for } k = 0, \ldots, d-1 \\ &y_k \perp y_l \text{ if } k \neq l. \end{aligned}$$

That is, all vectors in this basis are of unit length and they are pairwise orthogonal. We call such a basis an *orthonormal basis*.

The following identity may be called a parallelepiped identity, as it involves the eight corners of a parallelepiped if one adds $0 = \|0\|^2$ on the right hand side.

Theorem 1.32. *Let V be a Hilbert space. Then for all $x, y, z \in V$,*

$$\|x+y+z\|^2 + \|x\|^2 + \|y\|^2 + \|z\|^2 = \|x+y\|^2 + \|x+z\|^2 + \|y+z\|^2.$$

Proof. First we prove the theorem for $V = \mathbb{R}^d$, for which a simple computation shows that both hand-sides equal

$$\sum_{i=0}^{d-1} 2x_i^2 + 2y_i^2 + 2z_i^2 + 2x_iy_i + 2y_iz_i + 2z_ix_i.$$

For a general V consider the subspace W spanned by x, y, z, for which $1 \leq \dim(W) \leq 3$ unless in the trivial case $x = y = z = 0$. Since W is isometrically isomorphic to $\mathbb{R}^d$, the claim follows. □

The formula from the previous theorem has the following consequence.

Theorem 1.33. *Let V be a Hilbert space and $x \in V$. Then the function*

$$f(y) := \|x+y\|^2 - \|x\|^2 - \|y\|^2$$

is linear in $y \in V$.

Note that the defining expression for f is symmetric in x, y, so that a symmetric statement to the theorem holds as well.

Proof. We have already mentioned the homogeneity $f(\lambda y) = \lambda f(y)$, which follows from Theorem 1.28. To show additivity we calculate

$$f(y+z) = \|x+y+z\|^2 - \|x\|^2 - \|y+z\|^2,$$

while

$$f(y) + f(z) = \|x+y\|^2 - \|x\|^2 - \|y\|^2 + \|x+z\|^2 - \|x\|^2 - \|z\|^2.$$

Hence, using Theorem 1.32, $f(y+z) = f(y) + f(z)$. □

Definition 1.34. For x, y in a (real) Hilbert space V we define their *scalar product*

$$\langle x, y\rangle := \frac{1}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2).$$

We observe the following properties of the scalar product.

1. For a fixed x it is linear in y, i.e. for $\lambda, \mu \in \mathbb{R}$,

$$\langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle.$$

2. It is symmetric, i.e. $\langle x, y \rangle = \langle y, x \rangle$.

3. For a fixed y it is linear in x, i.e. for $\lambda, \mu \in \mathbb{R}$,

$$\langle \lambda x + \mu z, y \rangle = \lambda \langle x, y \rangle + \mu \langle z, y \rangle.$$

   Properties 1. and 3. are usually stated saying the scalar product is *bilinear*.

4. $\langle x, x \rangle = \|x\|^2$
   This holds since both hand-sides equal $\frac{1}{2}(\|2x\|^2 - \|x\|^2 - \|x\|^2)$.

5. $\langle x, y \rangle = 0 \Leftrightarrow x \perp y$,
   This holds since both statements are equivalent to $\|x + y\|^2 - \|x\|^2 - \|y\|^2 = 0$.

6. (Cauchy-Schwarz inequality) $\langle x, y \rangle \leq \|x\| \|y\|$.
   To derive this we first square the triangle inequality $\|x+y\| \leq \|x\|+\|y\|$ which yields

$$\begin{aligned} & \|x + y\|^2 \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \\ \Leftrightarrow\ & 2\langle x, y \rangle \leq 2\|x\|\|y\| \\ \Leftrightarrow\ & \langle x, y \rangle \leq \|x\|\|y\|. \end{aligned}$$

1.3.1 Complex Hilbert spaces

Let V be a vector space over $\mathbb{C}$. A norm on the complex vector space $\| \cdot \| : V \to \mathbb{R}_{\geq 0}$ satisfies

1. $\|x\| = 0 \Rightarrow x = 0 \ \forall x \in V$
2. $\|\lambda x\| = |\lambda| \|x\| \ \forall x \in V, \lambda \in \mathbb{C}$
3. $\|x + y\| \leq \|x\| + \|y\| \ \forall x, y \in V$

The difference to the real case is Property 2., which needs to hold for more general λ. But it still holds for all $\lambda \in \mathbb{R}$, so the normed space remains a real

normed space as well. As such it is still called Hilbert space, if the induced metric is complete and for all $x, y \in V$ the parallelogram law

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

holds.
In a complex Hilbert space, we call the property $\|x-y\| = \|x+y\|$ of two vectors real orthogonality, to distinguish it from complex orthogonality which will be defined below.
We observe that

$$\|x+ix\| = |1+i|\|x\| = \sqrt{2}\|x\| = |1-i|\|x\| = \|x-ix\|,$$

which means that x and ix are real orthogonal. In the complex setting this is somewhat undesirable, as ix is a complex multiple of x and thus x and ix are linearly dependent in the complex vector space.
In view of this, the real Gram Schmidt procedure does not produce a basis of the complex vector space. To rectify this, we modify the real Gram Schmidt as follows. We first choose a vector y_0 with $\|y_0\| = 1$. Then we define $y_i = iy_0$. Inductively if we have chosen $y_0, \ldots, y_{2d-1}$, we choose a unit vector y_{2d} orthogonal to the previously chosen vectors and then we choose $y_{2d+1} = iy_{2d}$. This gives a collection of $2d+2$ pairwise real orthogonal vectors such that they pair up into pairs of linearly dependent vectors over the complex numbers. If the real dimension of the vector space is $2d$, then taking the even numbered vectors $y_0, y_2, \ldots, y_{2d-2}$ provide a basis of the d-complex-dimensional space. It would be desirable to have a notion of complex orthogonality that deems these $2d$ vectors a maximal set of orthonormal vectors.
Thus we define a different scalar product.

Definition 1.35. Let V be a complex Hilbert space. The *complex scalar product* of two vectors $x, y \in V$ is defined as

$$\langle x, y\rangle := \frac{1}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2) + \frac{i}{2}(\|x+iy\|^2 - \|x\|^2 - \|iy\|^2).$$

Note that this equals the real scalar product plus an additional imaginary component. By the parallelogram law, $\langle x, y\rangle$ can also be written as

$$\frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) + \frac{i}{4}(\|x+iy\|^2 - \|x-iy\|^2).$$

First we show that we have the property

$$\langle x, iy\rangle = -i\langle x, y\rangle,$$

which in particular implies $\langle x, ix\rangle = -i\|x\|^2$ and thus $\langle x, ix\rangle \neq 0$ for $x \neq 0$. To see that, we write by definition

$$\langle x, iy\rangle = \frac{1}{2}(\|x+iy\|^2 - \|x\|^2 - \|iy\|^2 + i\|x-y\|^2 - i\|x\|^2 - i\|y\|^2),$$

which by the parallelogram law equals

$$\begin{aligned}
&\frac{1}{2}(\|x+iy\|^2 - \|x\|^2 - \|iy\|^2 - i\|x+y\|^2 + i\|x\|^2 + i\|y\|^2)\\
=&-\frac{i}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2 + i(\|x+iy\|^2 - \|x\|^2 - \|iy\|^2))\\
=&-i\langle x, y\rangle.
\end{aligned}$$

More generally, the complex scalar product satisfies the following.

1. For a fixed y it is linear in x.
2. It is *conjugate symmetric*, i.e. $\langle x, y\rangle = \overline{\langle y, x\rangle}$.
3. For a fixed x it is *conjugate linear* in y, i.e. for $\lambda, \mu \in \mathbb{C}$,

$$\langle x, \lambda y + \mu z\rangle = \bar{\lambda}\langle x, y\rangle + \bar{\mu}\langle x, y\rangle.$$

We leave the proof of these properties as an exercise. Linearity in x and conjugate linearity in y together are also called *sesquilinearity.*
In analogy with the real case we say that x, y are *complex orthogonal* if

$$\langle x, y\rangle = 0.$$

In particular, x and ix are not complex orthogonal if $x \neq 0$. Complex orthogonality is stronger than real orthogonality: it is tantamount to

$$\|x+y\|^2 - \|x\|^2 - \|y\|^2 = 0 \wedge \|x+iy\|^2 - \|x\|^2 - \|iy\|^2 = 0$$

As discussed above one can construct for a finite dimensional complex Hilbert space an orthonormal basis by the mentioned variant of the Gram Schmidt procedure.

1.3.2 Infinite dimensional Hilbert spaces

We are set to construct an infinite dimensional analog of $\mathbb{R}^d$. For a set M we define

$$\ell^2(M) := \{\alpha : M \to \mathbb{R} : \sum_{m\in M} |\alpha(m)|^2 < \infty\}$$

where

$$\sum_{m\in M} |\alpha(m)|^2 := \sup_{\substack{M'\subseteq M\\ M' \text{finite}}} \sum_{m\in M'} |\alpha(m)|^2.$$

One could extend the definition of $\ell^2(M)$ to $\mathbb{C}$ by considering all square summable functions mapping M to $\mathbb{C}$. Note that if $M = \{0, \dots, d-1\}$, we obtain $\mathbb{R}^d$.

Remark. If $\alpha \in \ell^2(M)$, for every $\varepsilon > 0$ only finitely many $\alpha(m)$ satisfy $|\alpha(m)| > \varepsilon$. Since

$$\{|\alpha(m)| > 0\} = \bigcup_{n\in\mathbb{N}_{\geq 1}} \{|\alpha(m)| > \frac{1}{n}\},$$

only countably many $\alpha(m)$ can be different from 0. This is true even if M is uncountable.

The space $\ell^2(M)$ is a real vector space with the operations

$$\begin{aligned}(\alpha+\beta)(m) &= \alpha(m) + \beta(m)\\ (\lambda\alpha)(m) &= \lambda(\alpha(m)),\ \lambda\in\mathbb{R}.\end{aligned}$$

Definition 1.36. For $\alpha \in \ell^2(M)$ set

$$\|\alpha\| := \sqrt{\sum_{m\in M} |\alpha(m)|^2}. \tag{12}$$

This expression defines a norm, so $\ell^2(M)$ is a normed space. In fact, even more is true.

Theorem 1.37. *$\ell^2(M)$ is a Hilbert space.*

Proof. First we show that (12) defines a norm. That $\|\alpha\| = 0 \Rightarrow \alpha(m) = 0$ for all $m \in M$ and that $\|\lambda\alpha\| = |\lambda|\|\alpha\|$ is clear. For the triangle inequality we expand

$$\|\alpha+\beta\| = \sqrt{\sup_{\substack{M'\subseteq M\\ M \text{ finite}}} \sum_{m\in M'} |\alpha(m)+\beta(m)|^2} \leq \sup_{\substack{M'\subseteq M\\ M \text{ finite}}} \sqrt{\sum_{m\in M'} |\alpha(m)+\beta(m)|^2}$$

the last inequality following by continuity of the square root. Using the triangle inequality in the finite dimensional space $\mathbb{R}^{|M'|}$, this is bounded by

$$\begin{aligned}
&\sup_{\substack{M'\subseteq M\\ M \text{ finite}}} \sqrt{\sum_{m\in M'} |\alpha(m)|^2} + \sqrt{\sum_{m\in M'} |\beta(m)|^2} \\
\leq &\sup_{\substack{M'\subseteq M\\ M \text{ finite}}} \sqrt{\sum_{m\in M'} |\alpha(m)|^2} + \sup_{\substack{M'\subseteq M\\ M \text{ finite}}} \sqrt{\sum_{m\in M'} |\beta(m)|^2} \\
= &\|\alpha\| + \|\beta\|.
\end{aligned}$$

The next thing to show is completeness. This can be shown analogously as completeness of $\ell^1(M)$, which was discussed in Analysis 1 and will be addressed in the next lecture. The validity of the parallelogram law can deduced from its validity in the finite dimensional Hilbert space $\mathbb{R}^{|M'|}$. The details of this last statement are left as an exercise. □

End of lecture 4. April 20, 2015

We now turn to the question of completeness of the space $l^2(M)$ for some, possibly infinite set of indices M.

Theorem 1.38. *Given a set of indices M, the normed vector space $l^2(M)$ is complete.*

Proof. Let α_n be a Cauchy sequence in $l^2(M)$. Then $\forall m \in M$ the sequence $\alpha_n(m)$ is a Cauchy sequence in $\mathbb{R}$ since $|\alpha_n(m) - \alpha_{n'}(m)| \leq \|\alpha_n - \alpha_{n'}\|$. We define

$$\alpha(m) = \lim_{n\to\infty} \alpha_n(m) \ .$$

This is the candidate α being the $l^2(M)$ limit of the sequence of functions α_n. We need to verify that α is an element of $l^2(M)$. First note that for the Cauchy sequence α_n there exists a $C > 0$ so that $\|\alpha_n\| < C$ for all $n \in \mathbb{N}$. We will show that $\|\alpha\| = \left(\sum_{m\in M} |\alpha(m)|^2\right)^{1/2} \leq C + 1$. Given any finite subset $M' \subset M$ one must show that $(\sum_{m\in M} |\alpha(m)|^2)^{1/2} \leq C + 1$. Choose $n' \in \mathbb{N}$ large enough with $n' > n$ so that $|\alpha_{n'}(m) - \alpha(m)| < \epsilon$ for all $m \in M'$ and for a suitably chosen $\epsilon > 0$. Then, by the triangle inequality in $\mathbb{R}^d$ and d the cardinality of M',

$$(\sum_{m\in M'} |\alpha(m)|^2)^{1/2} \leq (\sum_{m\in M'} (|\alpha_n(m)| + \epsilon)^2)^{1/2}$$

$$\leq (\sum_{m\in M'} |\alpha_n(m)|^2)^{1/2} + (M')^{1/2}\epsilon \leq C + 1 \ ,$$

where the last line follows by suitable choice of ϵ.
We now need to show that α is, as a matter of fact the l^2 limit of α_n. For any $\epsilon > 0$ choose $n \in \mathbb{N}$ such that for every n', $n'' \geq n$ we have that $\|\alpha_{n'} - \alpha_{n''}\|^2 \leq \epsilon$. Let $M' \subset M$ be a finite set such that

$$\sum_{m\in M'} |\alpha_n(m)|^2 \geq \|\alpha_n\|^2 - \epsilon$$
$$\sum_{m\in M'} |\alpha(m)|^2 \geq \|\alpha\|^2 - \epsilon$$

This also gives us that for all $n' > n$

$$\|\alpha_{n'}\|^2 - \sum_{m\in M'} |\alpha_{n'}(m)|^2 \leq \|\alpha_n\|^2 - \sum_{m\in M'} |\alpha_n(m)|^2 - 2\left|\|\alpha_{n'}\|^2 - \|\alpha_n\|^2\right| \leq 5\epsilon.$$

Let us choose $N > n$ sufficiently large so that for all $n' \geq N$ and for all $m \in M'$ we have $|\alpha_{n'}(m) - \alpha(m)| < \frac{\epsilon}{|M'|}$. This is possible since M' is a finite set and $\lim_{n'\to\infty} \alpha_{n'}(m) = \alpha(m)$. It follows that for $n' > N$

$$\|\alpha_{n'} - \alpha\|^2 = \sum_{m\in M} |\alpha_{n'}(m) - \alpha(m)|^2 \leq \sum_{m\in M'} |\alpha_{n'}(m) - \alpha(m)|^2 +$$

$$\sum_{m\in M\setminus M'} |\alpha_{n'}(m) - \alpha(m)|^2 \leq \sum_{m\in M'} \frac{\epsilon}{|M'|} + 2\left(\sum_{m\in M\setminus M'} |\alpha_{n'}(m)|^2\right.$$

$$\left. + \sum_{m\in M\setminus M'} |\alpha(m)|^2\right) \leq \epsilon + 2\left(\|\alpha_{n'}\|^2 - \sum_{m\in M'} |\alpha_{n'}(m)|^2 +\right.$$

$$\left. \|\alpha\|^2 - \sum_{m\in M'} |\alpha(m)|^2\right) \leq \epsilon + 10\epsilon + \epsilon \leq 12\epsilon.$$

This proves the claim that $\alpha_n \to \alpha$ in the metric space $l^2(M)$. We have used the fact that for positive quantities $a \leq b + c$ we have that $a^2 \leq 2(b^2 + c^2)$.
□

The parallelogram identity holds for $l^2(M)$. The proof of this is left as an exercise. As a consequence the scalar product on $l^2(M)$ is defined via the expression

$$\langle \alpha, \beta \rangle = \sum_{m\in M} \alpha(m)\beta(m).$$

If we consider complex-valued functions the scalar product is given by

$$\langle \alpha,\, \beta \rangle = \sum_{m \in M} \alpha(m)\overline{\beta(m)}.$$

As usual the complex scalar product is linear in the first term and anti-linear in the second term and, as expected, we have $\langle \alpha,\, \alpha \rangle = \sum_{m \in M} \alpha(m)\overline{\alpha(m)} = \sum_{m \in M} |\alpha(m)|^2 = \|\alpha\|_{l^2(M)}$.
To see that all the above expressions are well defined for elements $\alpha,\, \beta \in l^2(M)$ we show that the sequences are absolutely summable. In particular, given that for any $m \in M$ we have $|\alpha(m)\beta(m)| \leq \frac{1}{2}\left(\alpha(m)^2 + \beta(m)^2\right)$, for any finite subset M' we have that $\sum_{m \in M'} |\alpha(m)\beta(m)| \leq \frac{1}{2}\left(\|\alpha\|^2 + \|\beta\|^2\right)$.

1.3.3 Orthonormal systems and bases

The Hilbert space $l^2(M)$ is infinite dimensional when the set M is infinite. While it would be tempting to assume that the set of elements

$$b_{m'}(m) = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}$$

for $m' \in M$ are a basis of $l^2(M)$ this is false according to the definition given in linear algebra. The elements $b_{m'}$ are linearly independent but it is not true that any element of $l^2(M)$ can be represented as a *finite* linear combination of elements $b_{m'}$. This is what would be required for $b_{m'}$ to be a Hamel basis i.e. a basis in the sense one uses this term in linear algebra.
However since $l^2(M)$ is a Hilbert space and thus has the additional property of being a metric space and of having a scalar product we can introduce a new set of notions that will allow a basis of whose elements we will take infinite, albeit convergent in the Hilbert space metric, linear combinations.

Definition 1.39. Let V be a Hilbert space (real or complex). We say that a set of vectors $B \subset V$ is an orthonormal system if the following properties are satisfied

- $\forall b \in B$ the elements are *normalized* i.e. $\|b\| = 1$;
- $\forall b,\, b' \in B$ the elements coincide or are (real or complex, as the case may be) orthogonal: $b = b' \wedge \langle b,\, b' \rangle = 0$.

We next argue that for an orthonormal set B there exists an linear map $f : l^2(B) \to V$ such that $f(\alpha) = \sum_{b \in B} \alpha(b) b$ for any given sequence $\alpha \in l^2(B)$. The infinite sum is to be interpreted as convergent in the following sense:

there exists a unique element $v \in V$ so that for every $\epsilon > 0$ there exists a finite subset $M' \subset B$ so that for all $M'' \supset B'$ we have

$$\|v - \sum_{b \in M''} \alpha(b)b\|_V < \epsilon \ .$$

We say that a linear map f from a normed vector spaces X to a normed vector space Y is an isometry if $\|f(x)\|_Y = \|x\|_X$ for all vectors $x \in X$. Notice that such an isometric map is injective.

Proposition 1.40. *Given an orthonormal system $B \subset V$ of a Hilbert space V. The map f defined by setting $f(\alpha) = \sum_{b \in B} \alpha(b)b$ for all $\alpha \in l^2(B)$ is a well defined isometry.*

Proof. We must show that the sum $\sum_{b \in B} \alpha(b)b$ converges to an element in V for any $\alpha \in l^2(B)$.
Let $n \in N$ and let us choose a finite subset $M_n \subset B$ so that $\sum_{m \in M_n} |\alpha_m|^2 \geq \|\alpha\|^2_{l^2(B)} - 2^{-n}$. For any finite $M'' \supset M_n$ we have that

$$\| \sum_{b \in M''} \alpha(b)b - \sum_{b \in M_n} \alpha(b)b\|^2 = \| \sum_{b \in M'' \setminus M_n} \alpha(b)b\|^2$$

since the sets M_n, M'', $M'' \setminus M_n$ are all finite we can expand the above norm using the scalar product on V to obtain

$$\left\| \sum_{b \in M'' \setminus M_n} \alpha(b)b \right\|^2 = \sum_{b,\, b' \in M'' \setminus M_n} \alpha(b)\alpha(b')\langle b,\, b' \rangle = \sum_{b \in M'' \setminus M_n} |\alpha(b)|^2 \leq 2^{-n}.$$

Here we used that B is an orthonormal system.
Now consider the closed balls $B_{10\,2^{-n}}\left(\sum_{b \in M_n} \alpha(b)b\right)$ and let us call the centers of these balls $v_n := \sum_{b \in M_n} \alpha(b)b$. Since M_n can be chosen to be contained in one another as n increases the above balls are also contained in one another. Since V is a complete space there exists an element $v \in \bigcap_{n \in \mathbb{N}} B_{10 2^{-n}}(v_n)$ and we claim that $v = \sum_{b \in B} \alpha(b)b$. To verify this is left as an exercise.
Linearity of f is also easy to show and is left as an exercise. □

Similarly to how a set of linearly independent vectors must have the additional property of generating a finite-dimensional vector space, orthonormal systems also require a notion of generating a Hilbert space before for them to be an adequate substitute for the notion of a basis of Euclidean spaces.

Definition 1.41. An orthonormal system B of a Hilbert space V is said to be an orthonormal basis if the map $f : l^2(B) \to V$ given by $f(\alpha) = \sum_{b \in B} \alpha(b)b$ is surjective.

Theorem 1.42. *An orthonormal system B of a Hilbert space V is an orthonormal basis if and only if for any vector $v \in V$ we have that $v = 0$ or $\exists b \in B \ \langle v, b\rangle \neq 0$.*

Proof. $\Leftarrow$ Suppose that $\forall v \in V$ we have either $v = 0$ or $\exists b \in B \ \langle v, b\rangle \neq 0$. We must show that f is surjective. The image $W = f\left(l^2(B)\right)$ is a subspace of V because f is linear and thus W is closed with respect to the vector space operations (linear combinations of vectors). W is also a closed subspace of V since f is an isometry. Let $v \in V$ and let $w \in W$ be a vector such that $\|v - w\| = \inf_{z \in W} \|v - z\|$. Then we have that $\|v - w\| \perp z$ for all $z \in W$. Thus $\langle v - w; b\rangle = 0$ for all vectors $b \in B$. This means that $v - w = 0$ and thus $v \in W$ as required

$\Rightarrow$ We leave this implication as an exercise.

□

1.4 The Hilbert spaces $L^2(\mathbb{R})$ and $L^2([0, 1])$

In the last section we studied Hilbert spaces including the space $l^2(\mathbb{Z})$, that is the space of sequences parameterized by $\mathbb{Z}$ which are square summable with norm $\|\alpha\| = \left(\sum_{m \in \mathbb{Z}} |\alpha(m)|^2\right)^{1/2}$. It is natural to ask whether one could define a space of functions f on $\mathbb{R}$ which are square integrable in some sense and whose norm is defined as an integral with norm $\left(\int_{\mathbb{R}} |f(x)|^2 dx\right)^{1/2}$. However, it turns out that in order to obtain a Hilbert space in this fashion, that is in particular to obtain a complete metric space, one needs a rather general class of "functions" f and a powerful notion of the integral, that we do not have introduced yet. Here we put the word "functions" into quotation marks, since the exact development of the theory requires objects other than naive functions $f : \mathbb{R} \to \mathbb{R}$ in the literal sense. There are several approaches to this task, one involving Lebesgue integration theory, which we do not delve into now. Another approach is via martingale theory, which we will develop here.

Definition 1.43. Let $\mathcal{I}$ be the set of dyadic subintervals of $\mathbb{R}$ of the form $[n2^k, (n + 1)2^k)$ with $k, n \in \mathbb{Z}$. A martingale on $\mathbb{R}$ is a function $F : \mathcal{I} \to \mathbb{R}$ with the property that

$$\forall I \in \mathcal{I} \quad F(I) = \frac{1}{2}\left(F(I_l) + F(I_r)\right)$$

where I_r and I_l are the left and right dyadic children of I i.e. $I_l, I_r \in \mathcal{I}$ $I = I_l \cup I_r$ and $|I_l| = |I_r| = \frac{|I|}{2}$. If $I = [n2^k, (n + 1)2^k]$ then $I_l = \left[2n2^{k-1}, (2n + 1)^{k-1}\right)$ and $I_r = \left[(2n + 1)2^{k-1}, (2n + 2)2^{k-1}\right)$.

Similarly we define martingales on $[0,1)$ by considering functions on the collection $\mathcal{I}$ of dyadic intervals contained in $[0,1)$.
Locally integrable functions on the real line or on $[0,1)$ naturally define a martingale on the respective sets. We restrict attention to functions f of bounded variation. For each such f, the function $F(I) = \frac{1}{|I|}\int_I f(x)dx$ is a martingale. To see this, note that $I = I_l \cup I_r$ and $|I_r| = |I_l| = \frac{1}{2}|I|$ and hence

$$F(I) = \frac{1}{|I|}\int_I f(x)dx = \frac{1}{2|I_l|}\int_{I_l} f(x)dx + \frac{1}{2|I_l|}\int_{I_l} f(x)dx = \frac{1}{2}\left(F(I_l) + F(I_r)\right).$$

Definition 1.44. The space $L^2(\mathbb{R})$ is the vector space

$$L^2(\mathbb{R}) = \left\{ F \text{ martingale on } \mathbb{R} : \sup_{k\in\mathbb{Z}} \sum_{\substack{I\in\mathcal{I}\\ |I|=2^k}} |F(I)|^2|I| < \infty \right\}.$$

The space $L^2\left([0,\,1)\right)$ is the vector space

$$L^2\left([0,\,1)\right) = \left\{ F \text{ martingale on } [0,\,1) : \sup_{k\in\mathbb{Z}} \sum_{\substack{I\in\mathcal{I}\\ |I|=2^k}} |F(I)|^2|I| < \infty \right\}.$$

Note that in the first case the sums over intervals of fixed size are infinite sum, while in the second case they are finite sums.
We define on these spaces given by

$$\|F\|_{L^2} = \sup_{k\in\mathbb{Z}} \sum_{\substack{I\in\mathcal{I}\\ |I|=2^k}} |F(I)|^2|I|$$

where by $\mathcal{I}$ we intend the set of dyadic subintervals of $\mathbb{R}$ or $[0,\,1)$ respectively.

These quantities can be verified to be norms and make these spaces into actual Hilbert spaces. This will be done in more detail in the next lecture.
The expression $\sum_{|I|=2^k} |F(I)|^2|I|$ is monotone as $k \to -\infty$ since $a^2 + b^2 \geq 2\left(\frac{a+b}{2}\right)^2$. As a matter of fact notice that

$$\sum_{|I|=2^k} |F(I)|^2|I| = \sum_{|I|=2^k} \frac{1}{4}\left|F(I_l) + F(I_r)\right|^2|I| \leq$$
$$\sum_{I=2^k} \frac{1}{2}\left(|F(I_l)|^2 + |F(I_r)|^2\right)|I| = \sum_{|J|=2^{k-1}} F(J)^2|J|$$

Thus we can actually write that $\|F\| = \lim_{k\to-\infty} \sum_{|I|=2^k} |F(I)|^2|I|$.

End of lecture 5. April 23, 2015

We have defined the spaces $L^2(\mathbb{R})$ and $L^2([0,1))$ in terms of martingales. We now verify that the expression $\|F\|_{L^2} = \sup_k \sum_{|I|=2^k} |F(I)|^2|I| = \lim_{k\to-\infty} \sum_{|I|=2^k} |F(I)|^2|I|$ is actually a Hilbert space norm. We check these properties for $L^2([0,1))$, the slightly more involved case $L^2(\mathbb{R})$, which requires to handle infinite sums, is left as an exercise.

- $\|F\|_{L^2} = 0 \Rightarrow F = 0$ follows from the fact that for all $I \in \mathcal{I}$ we find k with $|I| = 2^k$ and we have that $|F(I)|^2|I| \le \sum_{|I|=2^k} |F(I)|^2|I| \le \|F\|_{L^2} = 0$ so $F(I) = 0$.

- To see $\|\lambda F\| = |\lambda|\|F\|$ we note that for each k

$$\sum_{|I|=2^k} |\lambda F(I)|^2|I| = |\lambda|^2 \sum_{|I|=2^k} |F(I)|^2|I| \ .$$

  Applying the supremum over k gives the desired result.

- The triangle inequality: $\|F+G\| \le \|F\| + \|G\|$. To show this we first consider fixed length 2^k. For $F, G \in L^2([0,1))$ and for any k we have

$$\left(\sum_{|I|=2^k} |F(I)+G(I)|^2|I|\right)^{1/2}$$

$$\le 2^{k/2} \left((\sum_{|I|=2^k} |F(I)|^2)^{1/2} + (\sum_{|I|=2^k} |G(I)|^2)^{1/2} \right)$$

$$= \left(\sum_{|I|=2^k} |F(I)|^2|I|\right)^{1/2} + \left(\sum_{|I|=2^k} |G(I)|^2|I|\right)^{1/2} .$$

  Here we used the triangle inequality on the space $l^2\left(\{I \in \mathcal{I},\, |I| = 2^k\}\right)$. Comparing with the supremum on the right hand side we obtain

$$\left(\sum_{|I|=2^k} |F(I)+G(I)|^2|I|\right)^{1/2} \le \|F\| + \|G\|$$

  Since this holds for all k, we obtain

$$\|F+G\|^2 \le \|F\| + \|G\|$$

  as required.

- The procedure to prove the parallelogram identity is very similar. We use that it holds in Euclidean spaces to conclude the corresponding identity for square sums over intervals of fixed length. This time it is crucial that the norm $\|\cdot\|_{L^2([0,1))}$ is given not only by supremum over all k of the expression $\left(\sum_{|I|=2^k} |F(I)|^2|I|\right)^{1/2}$ but by its limit as $k \to -\infty$. This allows to show equality in the limit and thus the parallelogram identity.

- Finally we need to check completeness of $L^2([0,1))$. This result is given by the following theorem.

Theorem 1.45. *The space of martingales $L^2([0,1))$ is complete.*

The proof will show some similarities to the proof that $l^2(M)$ is complete.

Proof. Let F_n be a Cauchy sequence of martingales in $L^2([0,1))$. We have for all $n, n' \in \mathbb{N}$ that $|F_n(I) - F_{n'}(I)| \leq \|F_n - F_{n'}\|$ so for every dyadic interval we have that $F_n(I)$ is a Cauchy sequence. Set $F(I) = \lim_n F_n(I)$ for every interval $I \in \mathcal{I}$.
We must check that F thus defined is an element of $L^2([0,1))$.

- F is a martingale because

$$F(I) = \lim_{n\to\infty} F_n(I) = \lim_{n\to\infty} \frac{1}{2}\left(F_n(I_l) + F_n(I_r)\right) = \frac{1}{2}\left(F(I_l) + F(I_r)\right).$$

  In words, since the martingale relationship is finite linear expression, one may pass to the limit in this expression.

- $\|F\|_{L^2} < \infty$. As a matter of fact for every k we have that the number of addends in the sum $\sum_{|I|=2^k} |F(I)|^2|I|$ is finite, thus $\sum_{|I|=2^k} |F(I)|^2|I| = \lim_{n\to\infty} \sum_{|I|=2^k} |F_n(I)|^2|I| \leq \sup_n \|F_n\| < \infty$ where the last inequality holds since F_n is a Cauchy sequence and thus is bounded. This yields that $\sup_k \sum_{|I|=2^k} |F(I)|^2|I| \leq \sup_n \|F_n\| < \infty$.

Now we need to show that $\lim_{n\to\infty} F_n = F$ in the sense of $L^2([0,1))$ i.e. that $\limsup_{n\to\infty} \|F_n - F\| = 0$. The procedure is somewhat similar to the proof of the completeness of $l^2(M)$. Let us define

$$\mathbb{E}_k F(I) = \begin{cases} F(I) & \text{if } |I| \geq 2^k \\ F(J) \text{ with } J \supset I \text{ and } J = 2^k & \text{if } |I| < 2^k \end{cases}.$$

This corresponds to truncating or stopping the martingale at length 2^k. We leave as an exercise to show that $\lim_{k\to-\infty} \|\mathbb{E}_k F - F\| = 0$. Let $\epsilon > 0$ and $n \in \mathbb{N}$ so that $\forall n' > n$ we have that $\|F_{n'} - F_n\| \leq \epsilon$. Let $m \in \mathbb{N}$, so that for all $k > m$ $\|\mathbb{E}_k F - F\| < \epsilon$ and $\|\mathbb{E}_k F_n - F_n\| < \epsilon$. Now select $n'' > n$ so that for all $n''' > n''$, $\|\mathbb{E}_k F_{n'''} - \mathbb{E}_k F\| < \epsilon$ holds. This is possible since when comparing $\mathbb{E}_k F$ with the stopped martingales $\mathbb{E}_k F_n$ we are only considering the norm at interval length 2^k since then it stabilizes. This implies that for all $n''' > n''$ we have $\|F_{n'''} - F\| \leq 10\epsilon$. □

So we have shown that $L^2([0,\,1))$ is a Hilbert space. The scalar product on $L^2([0,\,1))$ can be deduced using the polarization formula and is given by

$$\langle F,\, G\rangle = \lim_{k\to-\infty} \sum_{\substack{I\in\mathcal{I}\\ |I|=2^k}} F(I)G(I)|I|.$$

Let us provide an example of how martingales on intervals are related to functions.
Let $f : [0,\,1) \to \mathbb{R}$ be a function of bounded variation. As seen in Analysis 1 this means that f can be represented as a difference of two positive monotone functions. Then, as we have already mentioned, the expression $F(I) = \frac{1}{|I|}\int_I f(x)dx$ defines a martingale.
We have already checked the martingale property previously. Let us check that $F \in L^2([0,\,1))$. First suppose that f and g are positive, and monotone increasing. For any k we have

$$\sum_{|I|=2^k} F(I)G(I)|I| \leq \sum_{|I|=2^k} f(r(I))g(r(I))|I| \\ \sum_{|I|=2^k} f(l(I))g(l(I))|I|$$

Here we denoted by $l(I)$ and $r(I)$ the left and right endpoints of I. Note that on the left and on the right we see lower and upper Riemann sums of the function fg on the interval $[0, 1)$. Passing to the limit $k \to -\infty$ we see that

$$\lim_{k\to-\infty} \sum_{|I|=2^k} F(I)G(I)|I| = \int_0^1 f(x)g(x)\,dx$$

The latter identity first extends to g of bounded variation since such functions are difference of two monotone increasing functions of bounded variation, and then it extends to f of bounded variation as well. Finally, specializing $f = g$ we see that

$$\lim_{k\to-\infty} \sum_{|I|=2^k} |F(I)|^2|I| = \int_0^1 |f(x)|^2\,dx$$

for any function f of bounded variation.
The natural question is whether the converse holds in some sense. Does a martingale in $L^2([0,1))$ define a function. It is natural to assume that if a martingale is defined by a function, then the function can be mostly recovered by $f(x) := \lim_{\substack{|I|\to 0 \\ I\ni x}} F(I)$. However, beginning with a function f bounded variation, then defining the martingale F by averaging over dyadic intervals, and then returning to a function by passing to the above limit, one does not necessarily recover $f(x)$ at all points x as the limit need not exists at all points. Only if F comes from a function that is both continuous and of bounded variation then one can check that $\lim_{\substack{|I|\to 0 \\ I\ni x}} F(I) = f(x)$. at every point (exercise). For arbitrary martingale (not necessarily in $L^2([0,1))$), this limit does not need to exists for any $x \in [0,1)$. Using rather elaborate arguments one can see that for martingales in $L^2([0,1))$ this limit exists at many points, making such statement more precise requires the notion of Lebesgue measure.

1.4.1 Orthonormal basis for $L^2([0,1))$

We will now pass to describing important orthonormal systems and bases for $L^2([0,1))$.

Definition 1.46. A Haar function of a dyadic interval $I \in \mathcal{I}$ is a function of bounded variation on I given by

$$h_I(x) = \sqrt{\frac{1}{|I|}}\left(1_{I_l}(x) - 1_{I_r}(x)\right).$$

The associated martingales are indicated by H_I.

We can explicitly express H_I:

$$H_I(J) = \begin{cases} 0 & \text{if } J \supseteq I \\ \left(\frac{1}{|I|}\right)^{1/2} & \text{if } J \subseteq I_l \\ -\left(\frac{1}{|I|}\right)^{1/2} & \text{if } J \subseteq I_r \end{cases}$$

The Haar function together with the constant function $1_{[0,1)}$ on $[0,1)$ (more precisely the associated martingale) form an orthonormal basis of $L^2([0,1))$. We prove this in the following two propositions. We will refer to the constant function as 1 so $1 = 1_{[0,1)}$.

Proposition 1.47. *The Haar martingales H_I of all the dyadic subintervals $I \in \mathcal{I}$ and the martingale* 1 *form an orthonormal system.*

Proof. Checking that $\|H_I\| = \|1\| = 1$ is left as an exercise. First we show that $\langle H_I, 1\rangle = 0$ for all $I \in \mathcal{I}$. This follows from easily from the mean 0 property of Haar functions: $\langle H_I, 1\rangle = \int_0^1 h_I(x)dx = 0$. Now suppose $I, J \in \mathcal{I}$, $I \neq J$; we must show that $\langle H_I, H_J\rangle = 0$. As usual we use that $\langle H_I, H_J\rangle = \int h_I(x)h_J(x)dx$.

- If $I \cap J = \emptyset$ then h_I and h_J have disjoint supports and the statement follows immediately.
- If $I \subset J$ and $I \neq J$ then $I \subset J_l$ or $I \subset J_r$. However h_I has integral 0 and h_J is constant on J_l and J_r respectively so $\int_0^1 h_I(x)h_J(x)dx = H_J(I)\int_I h(x)dx = 0$.
- If $I \supset J$ the reasoning is symmetric to the previous case.

□

Theorem 1.48. *The set $\{1\}\cup\{H_I\}_{I\in\mathcal{I}}$ forms an orthonormal basis of $L^2([0, 1))$.*

Proof. The set $\{1\}\cup\{H_I\}_{I\in\mathcal{I}}$ is an orthonormal system as checked above. We only need to check surjectivity of the associated isometry. Using the criterion of orthogonality we must show that if $V \in L^2([0, 1))$ such that $V \perp H_I$ for all dyadic intervals $I \in \mathcal{I}$ and $V \perp 1_{[0,1)}$ then $V = 0$ i.e. $V(I) = 0$ for all dyadic intervals $I \in \mathcal{I}$.
First of all $0 = \langle V, 1\rangle = \lim_{k\to-\infty}\sum_{J=2^k} V(J)|J| =_{k=0} V([0, 1))$.
We proceed by induction on k, the length of the dyadic intervals. Suppose that $V(I) = 0$ for all $|I| = 2^k$. Let $|I| = 2^{k-1}$, then there exists $J \in \mathcal{I}$ with $|J| = 2^k$ such that $I = J_l$ or $I = J_r$. Using the properties of the martingales we have that $V(J_l)+V(J_r) = 2V(J)$ while $V(J_l)-V(J_r) = \langle V, H_J\rangle|J|^{-1/2} = 0$ since $V \perp H_J$. This implies both that $V(J_l) = 0$ and $V(J_r) = 0$ and thus that $V(I) = 0$. This concludes the proof. □

Finally we turn attention to complex martingales, that is function $F : \mathcal{I} \to \mathbb{C}$ with the martingale condition

$$2F(I) = F(I_l) + F(I_r)$$

for all $I \in \mathcal{I}$. The complex martingales satisfying

$$\sup_k \sum_{I\in\mathcal{I},|I|=2^k} |F(I)|^2|I|$$

form a complex Hilbert space $L^2([0,1))$. This notation does not distinguish real and complex Hilbert space, it will have to be clear from the context which space is meant. One might write $L^2([0,1),\mathbb{R})$ $L^2([0,1),\mathbb{C}))$ if one wants to distinguish the spaces explicitly.
The behavior of the complex Hilbert space is completely analogous to that of the real Hilbert space. The scalar product is given by

$$\langle F,\, G\rangle = \lim_{k\to-\infty} \sum_{\substack{I\in\mathcal{I}\\ |I|=2^{-k}}} F(I)\overline{G(I)}|I|.$$

The Haar function together with 1 once again are an orthonormal basis. However we have another important orthonormal system of functions. Let $f_n(x) = e^{2\pi inx}$. These functions are continuous and of bounded variation therefore the associated martingales F_n are elements of $L^2\left([0,\,1)\right)$. The martingales F_n are orthogonal because

$$\int_0^1 f_n(x)\overline{f_m(x)}dx = \int_0^1 e^{2\pi inx}e^{-2\pi imx}dx = \begin{cases} 1 & n=m \\ 0 & n\neq m \end{cases}.$$

The scalar product is 0 when $n \neq m$ because the exponential function is $e^{2\pi i(n-m)x}$ has integral 0.

Theorem 1.49. *The functions $e^{2\pi i(n-m)x}$ are an orthonormal basis of the space of complex martingales $L^2\left([0,\,1),\,\mathbb{C}\right)$*

Proof. Once again we must only show surjectivity and to do so we use the orthogonality criterion. Let $V \perp F_n$ for all $n \in \mathbb{Z}$. We must show that $V(I) = 0$ for all $I \in \mathcal{I}$. This follows from the fact that we can approximate characteristic functions of dyadic intervals. Let $I \in \mathcal{I}$ and $\epsilon > 0$; set $g_{I,\epsilon}(x) = \sup_{z\in[0,1)} \max\{0, 1_I(z) - \epsilon^{-1}|z-x|\}$. The functions $g_{I,\epsilon}(x)$ are of bounded variation, are bounded by 1 and are supported on $\{z \in [0,\,1) \,|\, B_\epsilon(z)\cap I \neq \emptyset\}$. By the Theorem of Stone-Weirstrasse for any $\epsilon > 0$ there exists a trigonometric polynomial p i.e. a finite linear combination of the functions f_n such that $|p(x) - g_{I,\epsilon}(x)| < \epsilon$ for all $x \in [0,\,1)$. Since f_n are of bounded variation and p is a trigonometric polynomial we have that $\|P - G_{I,\epsilon}\|^2 = \int_0^1 |p(x) - g_{I,\epsilon}(x)|^2 dx \leq \epsilon^2$ and for a similar reason $\|G_{I,\epsilon} - 1_I\|^2 < 2\epsilon$. Notice also that we have $V(I) = \langle V, 1_I\rangle$ where in this case we intend by 1_I the martingale associated to the characteristic function of the interval I. But this means that $|V(I)| = |\langle V, 1_I\rangle| \leq |\langle V, P\rangle| + |\langle V, P - G_{I,\epsilon}\rangle| + |\langle V, 1_I - G_{I,\epsilon}\rangle|$. Applying Cauchy-Schwarz we obtain that $|\langle V, P - G_{I,\epsilon}\rangle| \leq \|V\|\|P - G_{I,\epsilon}\| \leq \epsilon\|V\|$ and $|\langle V, 1_I - G_{I,\epsilon}\rangle| \leq \|V\|\|1_I - G_{I,\epsilon}\| \leq (2\epsilon)^{1/2}\|V\|$. By linearity

$\langle V, P\rangle = 0$ so we have that $|V(I)| < \|V\|\left(\epsilon + (2\epsilon)^{1/2}\right)$. Since $\epsilon > 0$ can be chosen arbitrarily small this implies that $V(I) = 0$ as required.

□

End of lecture 6. April 27, 2015

We have the following representation, analogous to the case of $\mathbb{R}^d$.

Theorem 1.50. *Let B be an orthonormal basis of a Hilbert space V. For $v \in V$ we have*

$$\sum_{b\in B}\langle v, b\rangle b = v.$$

The above sum should be interpreted in the following sense: for every $\varepsilon > 0$ there exists an $M \subset B$ such that for all finite M' with $B \supset M' \supset M$,

$$\|v - \sum_{b\in M'}\langle v, b\rangle b\| \leq \varepsilon.$$

Proof. We already know that there exists a surjective map $\ell^2(B) \to V$ given by $\alpha \mapsto \sum_{b\in B}\alpha(b)b$, such that

$$v = \sum_{b\in B}\alpha(b)b$$

in the above mentioned sense. Let now $\varepsilon > 0$ and M be as above. Let $b_0 \in B$ and set $M' := M \cup \{b_0\}$. Then

$$|\langle v, b_0\rangle - \alpha(b_0)| = |\langle v, b_0\rangle - \sum_{b\in M'}\alpha(b)\langle b, b_0\rangle|.$$

where we have used that the basis B is orthonormal and thus $\langle b, b_0\rangle = 0$ for $b \neq b_0$ and $\langle b_0, b_0\rangle = 1$. The last display can be further rewritten as

$$|\langle v - \sum_{b\in M'}\alpha(b), b_0\rangle|,$$

which is by the Cauchy-Schwarz inequality estimated by

$$\|v - \sum_{b\in M'}\alpha(b)\|\|b_0\| = \|v - \sum_{b\in M'}\alpha(b)\| \leq \varepsilon.$$

Since ε was arbitrary, it follows $\langle v, b_0\rangle = \alpha(b_0)$. □

We discuss an important application of this theorem. We already know that $b_n(x) = e^{2\pi inx}$ form an orthonormal basis of $L^2([0,1),\mathbb{C})$. For all f of bounded variation, for which we have learned how to write the inner product as an integral, we define

$$\langle f, b_n\rangle = \int_0^1 f(x)e^{-2\pi inx}dx =: \widehat{f}(n)$$

The numbers $\widehat{f}(n)$ are called the Fourier coefficients of f. By the last theorem we have

$$f = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{2\pi inx}. \tag{13}$$

The series on the right hand-side is called the *Fourier series* of f. We stress once more that the sum in (13) should be interpreted in the sense described above. In the present case this means that $\forall\varepsilon > 0\,\exists N : \forall n' > N$:

$$\Big(\int_0^1 |\sum_{n=-n'}^{n'} \widehat{f}(n)e^{2\pi inx} - f(x)|^2 dx\Big)^{1/2} \leq \varepsilon.$$

For general functions of bounded variation, the equality (13) need not hold pointwise at every point in the domain of f, in fact the series need not converge for a given point x.
Since the mentioned surjective map $\ell^2(B) \to L^2([0,1],\mathbb{C})$ is an isometry, for an f of bounded variation we have

$$\int_0^1 |f(x)|^2 dx = \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2.$$

2 Differentiation in $\mathbb{R}^n$

In this chapter we shall examine differentiation of functions $f : \mathbb{R}^n \to \mathbb{R}^m$.
In Linear Algebra I we have already met linear maps. An $m \times n$ matrix

$$A = (a_{ij})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$$

defines a linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ by setting

$$f(x) = Ax.$$

For $1 \leq i \leq n$, the i-th component of $f(x)$ is given by $(f(x))_i = \sum_{j=1}^n a_{ij}x_j$. Conversely, if we fix respective bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, every linear map f :

$\mathbb{R}^n \to \mathbb{R}^m$ can be represented by an $m \times n$ matrix. Recall that if $n = 2$ and $m = 1$, the graph of f is a plane through the origin in $\mathbb{R}^3$.
More generally, a map of the form

$$f(x) = Ax + b$$

where $x \to Ax$ is linear and b is a constant vector in $\mathbb{R}^m$ is called an *affine linear map*. Note that if $n = 2$ and $m = 1$, the graph of f is a plane through the the point $(0, 0, b)$ in $\mathbb{R}^3$.

In all of the following Ω will be an open set of $\mathbb{R}^n$. The next definition generalizes the notion of differentiability from Analysis I to functions $\mathbb{R}^n \to \mathbb{R}^m$.

Definition 2.1. A function $f : \Omega \to \mathbb{R}^m$ is said to be *totally differentiable* at $x_0 \in \Omega$, if there exists an affine linear map $x \mapsto Ax + b$ such that

$$\forall \varepsilon > 0 \, \exists \delta > 0 : \forall x \in B_\delta(x_0) \cap \Omega : \|f(x) - (Ax + b)\| < \varepsilon \|x - x_0\|.$$

The role of the vector b in the above is less dominant, we can rephrase the definition as follows:

Theorem 2.2. *A function $f : \Omega \to \mathbb{R}^m$ is totally differentiable at $x_0 \in \Omega$ if and only if there is a linear map A such that*

$$\forall \varepsilon > 0 \, \exists \delta > 0 : \forall \|h\| < \delta, x_0 + h \in \Omega : \|f(x_0 + h) - f(x_0) - Ah\| < \varepsilon \|h\|. \tag{14}$$

Proof. ($\Rightarrow$) The claim follows with A being the linear part of the affine map from definition of total differentiability. We elaborate on details. Pick an $\varepsilon > 0$. Then there is a $\delta > 0$ and an affine map $Ax + b$ such that

$$\|f(x_0) - (Ax_0 + b)\| \leq \varepsilon \|x_0 - x_0\| = 0.$$

Thus, $f(x_0) = Ax_0 + b$. Then for every $\|h\| < \delta$ with $x_0 + h \in \Omega$,

$$\begin{aligned} \|f(x_0 + h) - f(x_0) - Ah\| &= \|f(x_0 + h) - Ax_0 - b - Ah\| \\ &= \|f(x_0 + h) - (A(x_0 + h) + b)\| \\ &\leq \varepsilon \|x - x_0\| = \varepsilon \|h\|. \end{aligned}$$

The last inequality again follows from Definition 2.1, since $x_0 + h \in B_\delta(x_0)$.

The direction ($\Leftarrow$) is left as an exercise. $\square$

We would like to call the map A from the definition of differentiability or the subsequent theorem the *total derivative* of f at x_0. To do so, we need to know uniqueness of this map, which is proved in the following theorem. This is analogous to the known case $n = m = 1$.

Theorem 2.3. *Let $f : \Omega \to \mathbb{R}^m$ and $x_0 \in \Omega$. There exists at most one linear map A such that* (14) *holds.*

Proof. Assume there are two linear maps A_1, A_2 such that (14) is true. Let $\varepsilon > 0$. Then

$$\exists \delta_1 > 0 : \forall \|h\| < \delta_1, x_0 + h \in \Omega : \|f(x_0 + h) - f(x_0) - A_1 h\| < \varepsilon \|h\|$$

and

$$\exists \delta_2 > 0 : \forall \|h\| < \delta_2, x_0 + h \in \Omega : \|f(x_0 + h) - f(x_0) - A_2 h\| < \varepsilon \|h\|.$$

Define $\delta := \min(\delta_1, \delta_2)$ and let $x_0 + h \in \Omega$, $\|h\| < \delta$. Then

$$\begin{aligned} \|A_1 h - A_2 h\| =& \| - (f(x_0 + h) - f(x_0) - A_1 h) + f(x_0 + h) - f(x_0) - A_2 h\| \\ \leq& \|f(x_0 + h) - f(x_0) - A_1 h\| + \|f(x_0 + h) - f(x_0) - A_2 h\| \\ \leq& 2\varepsilon \|h\|. \end{aligned}$$

Thus, $\|(A_1 - A_2)h\| \leq 2\varepsilon \|h\|$. We would like to show that the same holds if we replace h with a general $y \in \mathbb{R}^n$, which follows from the following scaling argument. Let $y \in \mathbb{R}^n$. Since Ω is open, we can choose $\lambda > 0$ such that $x_0 + \lambda y \in \Omega$, $\|\lambda y\| < \delta$. Set $h := \lambda y$. Then by the above discussion

$$\|(A_1 - A_2)\lambda y\| \leq \varepsilon \|\lambda y\|$$

By linearity of $A_1 - A_2$ and of the norm, this is equivalent to

$$\begin{aligned} &\lambda \|(A_1 - A_2)y\| \leq \varepsilon \lambda \|y\| \\ \Leftrightarrow &\|(A_1 - A_2)y\| \leq \varepsilon \|y\| \end{aligned}$$

Since ε was arbitrary, for all $y \in \mathbb{R}^n$ we have $(A_1 - A_2)y = 0$, hence $A_1 - A_2 = 0$, i.e. $A_1 = A_2$. □

Note that if f is linear, then it is its own total derivative at every $x_0 \in \Omega$.

Since $\mathbb{R}^n$ is an $n-$fold product of $\mathbb{R}$, it is a natural question whether we can characterize total differentiability of functions $\mathbb{R}^n \to \mathbb{R}^m$ with total differentiability of functions mapping $\mathbb{R}^n \to \mathbb{R}$ or $\mathbb{R} \to \mathbb{R}^m$.

For an $f : \Omega \to \mathbb{R}^m$ we write

$$f(x) = (f_1(x), \ldots, f_m(x)),$$

where for $1 \leq i \leq m$, $f_i : \Omega \to \mathbb{R}$ are called *component functions.*

Theorem 2.4. *A function $f : \Omega \to \mathbb{R}^m$ is totally differentiable at x_0 if and only for all $1 \leq i \leq m$, the component functions f_i are totally differentiable at x_0.*

Proof. ($\Leftarrow$) Assume that all component functions are totally differentiable at $x_0 \in \Omega$. Then for every $1 \leq i \leq m$ there exists a vector $(a_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n$ such that $\forall \varepsilon > 0 \exists \delta_i > 0 : \forall \|h\| < \delta_i, x_0 + h \in \Omega$,

$$|f_i(x_0 + h) - f_i(x_0) - \sum_{j=1}^{n} a_{ij} x_j| \leq \frac{\varepsilon}{\sqrt{m}} \|h\|.$$

Set $\delta := \min_{1 \leq i \leq m} \delta_i$. Let $\|h\| < \delta, x_0 + h \in \Omega$. Define A by the matrix

$$(a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}.$$

Then we have

$$\begin{aligned} \|f(x_0 + h) - f(x_0) - Ah\| &= \Big(\sum_{i=1}^{m} |f_i(x_0 + h) - f_i(x) - \sum_{j=1}^{n} a_{ij} x_j|^2 \Big)^{1/2} \\ &\leq \sqrt{m} \frac{\varepsilon}{\sqrt{m}} \|h\| = \varepsilon \|h\|. \end{aligned}$$

Thus, f is totally differentiable at x_0.
($\Rightarrow$) Exercise. □

This theorem gives an affirmative answer to the first question, that is, total differentiability of functions $\mathbb{R}^n \to \mathbb{R}^m$ is equivalent to total differentiability of all of its component functions, which map $\mathbb{R}^n \to \mathbb{R}$. To examine the second question we introduce the following definition.

Definition 2.5. Let $v \neq 0$ be a vector in $\mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}^m$ is said to be *differentiable in the direction of v* at x_0 if $f \circ g : I \to \mathbb{R}^m$ is totally differentiable at 0, where $I := \{t \in \mathbb{R} : x_0 + tv \in \Omega\}$ and $g : I \to \mathbb{R}^n$ is defined by $g(t) := x_0 + tv$. The total derivative of $f \circ g$ at 0 is called the *directional derivative along v.*

Note that the set I is open. One often normalizes v to satisfy $\|v\| = 1$.

Theorem 2.6. *If $f : \Omega \to \mathbb{R}^m$ is totally differentiable at $x_0 \in \Omega$, then it is differentiable in the direction of all $v \neq 0$ at x_0 and the directional derivative equals Av.*

Proof. Let $\varepsilon > 0$. We know that $\exists \delta > 0 : \forall \|h\| < \delta, x_0 + h \in \Omega$: $\|f(x_0 + h) - f(x_0) - Ah\| \leq (\varepsilon/\|v\|)\|h\|$ for some linear map $A : \mathbb{R}^n \to \mathbb{R}^m$. Set $\delta' := \delta/\|v\|$, so that $|t| < \delta' \Rightarrow \|tv\| < \delta$. Then for every $|t| < \delta', x_0 + tv \in \Omega$:

$$\|f(x_0 + tv) - f(x_0) - A(tv)\| \leq (\varepsilon/\|v\|)\|tv\| = (\varepsilon/\|v\|)|t|\|v\| = \varepsilon|t|.$$

Since $A(tv) = tA(v)$ and $t \mapsto tAv$ is linear in t, this means that $f \circ g$ is differentiable at 0 with the derivative Av. □

However, the converse of this theorem does not hold in general.

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

$$f(x_1, x_2) = \begin{cases} \frac{x_1^3}{x_1^2 + x_2^2} & ; (x_1, x_2) \neq 0 \\ 0 & ; (x_1, x_2) = 0 \end{cases}$$

Then f is clearly differentiable at $(x_1, x_2) \neq 0$. At 0, f is continuous. We claim that f has directional derivatives at 0 in the direction of all $v \neq 0$, but f is not totally differentiable at 0.
Directional differentiability: Let $v \neq 0 \in \mathbb{R}^2$. Then

$$f \circ g(t) = f(0 + tv) = f(tv_1, tv_2) = t\Big(\frac{v_1^3}{v_1^2 + v_2^2}\Big),$$

i.e.

$$f(0 + tv) - f(0) - t\Big(\frac{v_1^3}{v_1^2 + v_2^2}\Big) = 0.$$

Since the map $t \mapsto t\frac{v_1^3}{v_1^2+v_2^2}$ is linear in t, this means that f is differentiable along each $v \neq 0$.
Total differentiability: If f were totally differentiable at 0, by the previous theorem its total derivative would have to be

$$A(v_1, v_2) = \frac{v_1^3}{v_1^2 + v_2^2},$$

where A is given by $A = (a_1, a_2)$. Thus, for all $v \neq 0$ we would have

$$a_1 v_1 + a_2 v_2 = \frac{v_1^3}{v_1^2 + v_2^2}.$$

Inserting $v = (1, 0)$ gives the condition $a_1 = 1$. Inserting $v = (0, 1)$ yields $a_2 = 0$. Thus, $a_1 + a_2 = 1$. Now, plugging in $v = (1, 1)$ gives $a_1 + a_2 = 1/2$, which is a contradiction. Hence f cannot be totally differentiable at 0.

End of lecture 7. April 30, 2015

Let us recall the definition of total differentiability from the previous lecture. Let $\Omega \subset \mathbb{R}^n$ be an open set and let $f : \Omega \to \mathbb{R}^m$; fix a point $x \in \Omega$. The function f is said to be totally differentiable in x if there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that $\forall \epsilon > 0 \, \exists \delta > 0$ such that for all vectors h such that $\|h\| < \delta$ and $x + h \in \Omega$ we have $\|f(x + h) - f(x) - Ah\| \leq \epsilon \|h\|$. If the above holds we say that $A = Df(x)$ i.e. A is the differential of f in x. Since $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map we know that it is represented by an $m \times n$ matrix $a_{i,j}$. In particular we have that $(Ah)_i = \sum_{j=1}^n a_{i,j} h_j$. Let us recall the Theorem from the last lecture that relates the total differentiability of the components f_i of an $\mathbb{R}^m$-valued function with the total differentiability of the function f itself.
Recall that a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ defined on an open set $\Omega \subset \mathbb{R}^n$ with $f = (f_i)_{i=1,\dots,m}$ with $f_i : \Omega \to \mathbb{R}$ is totally differentiable in x is an only if all the functions f_i are totally differentiable in x for all $i \in \{1, \dots, m\}$.
It is easy to relate the total differentials $D(f_i)$ of the single components to the total differential of Df. The proof from the previous lesson yields that $D(f_i)h = (Dfh)_i = \sum_{j=1}^n Df_{i,j} h_j$ so the differentials of f_i are just the $1 \times n$ matrixes that are the row vectors of Df.
A different and more complicated situation occurs when we try to study the differentiability of a functions f when we look at the domain $\mathbb{R}^m$ as a product space. One can already imagine that having directional derivatives is not the same as having a total derivative as we will see happens in the already mentioned counterexample.
However we will see that under appropriate continuity assumptions we can deduce total differentiability from the differentiability along the single components of $\mathbb{R}^n$. We begin with an important definition. For ease of notation we will suppose that the functions we are dealing with are $\mathbb{R}$-valued. We have seen that functions with values in $\mathbb{R}^m$ can be dealt simply by studying their single components.

Definition 2.7. Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. f is partially differentiable in x if for every $j \in \{1, \dots, n\}$ there exists $a_j \in \mathbb{R}$ such that for every $\epsilon > 0$ there exists a $\delta > 0$ such that for any $|t| < \delta$ one has $|f(x + te_j) - f(x) - a_j t| \leq \epsilon |t|$ where e_j is the j^{th} basis vector $e_j = (0, \dots, 0, \underbrace{1}_{j^{\text{th}}}, 0, \dots, 0)$. In this case we will write $D_j f(x) = a_j$ to indicate the partial derivative of the function f in direction j. In mathematical literature these quantities are sometimes written as $\partial_j f$ or $\frac{\partial f}{x_j}$.

Notice that similarly to standard derivatives, total derivatives and partial derivatives if they exists are respectively unique.
Obviously total differentiability is a stronger notion than partial differentiability.

Proposition 2.8. *If f is totally differentiable in x then it is also partially differentiable in x and one has $D_j f(x) = Df(x)e_j$. This is a special case of directional differentiability along e_j.*

The interesting fact is that while it is generally not true that if all partial derivatives exists in a point then the function is totally differentiable, one can recover total differentiability if one has local continuity of the partial derivatives.

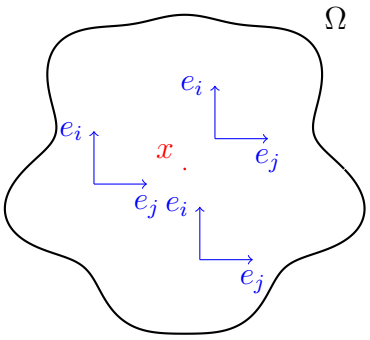


Figure 5: Partial derivatives must exist in an open set and be continuous in x

Theorem 2.9. *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ and let $x \in \Omega$. If all the partial derivatives $D_j f$ exist in all points of Ω and are continuous functions on Ω then f is totally differentiable in x and $(Df)_{1,j} = a_{1,j} = \partial_j f(x)$*

Proof. Let $\epsilon > 0$ and let us choose a $\delta > 0$ so that for any $h \in \mathbb{R}^n$, $\|h\| < \delta$ and for all $j \in \{1, \dots, n\}$ one has if $x + h \in \Omega$ then

$$|D_j f(x + h) - D_j f(x)| \leq \frac{\epsilon}{n}$$

This can be done using the continuity assumption on the partial derivatives we require continuity only in the point x.
We introduce the notation $h^{(j)} = (h_1, \dots, h_{j-1}, h_j, 0, \dots, 0) = \sum_{l=1}^{j} \langle h;\, e_j \rangle e_j$ to indicate the projection of h on the subspace spanned by the first j basis vectors. Using a telescoping sum we can write

$$f(x + h) - f(x) = f(x + h^{(n)}) - f(x + h^{(0)}) = \sum_{j=1}^{n} f(x + h^{(j)}) - f(x + h^{(j-1)})$$

But for each addend we have the relation

$$f(x+h^{(j)})-f(x+h^{(j-1)})=f(x+h^{(j-1)}+h_je_j)-f(x+h^{(j-1)})$$

so let us set $g(t)=f(x+h^{(j-1)}+te_j)-f(x+h^{(j-1)})$. Notice that $g'(t)=D_jf(x+h^{(j-1)}+te_j)$ for all t so that $x+h^{(j-1)}+te_j\in\Omega$.
Applying the Lagrange Theorem on the derivative in an intermediate point we have that $g(h_j)-g(0)=g'(t_j)h_j$ for some $t_j\in(0,h_j)$. This holds since g is differentiable for all for all $t\in(0,\,h_j)$ and continuous for $t\in[0,\,h_j]$. Substituting these relations into the original identity for f we have that

$$f(x+h)-f(x)=\sum_{j=1}^{n}D_jf(x+h^{(j-1)}+t_je_j)h_j.$$

We also have the following estimate on the norm of the displacement vector that will allow us to use the continuity assumption on the partial derivatives of f:

$$\begin{aligned}&\|h^{(j-1)}+t_je_j\|\leq\|h\|<\delta\quad\text{ so that}\\&\left|D_jf(x+h^{(j-1)}+t_je_j)-D_jf(x))\right|\leq\frac{\epsilon}{n}.\end{aligned}$$

The candidate for the full differential of f in x is the linear map $h\mapsto\sum_{j=1}^{n}D_jf(x)h_j$ so we estimate

$$\begin{aligned}&\left|f(x+h)-f(x)-\sum_{j=1}^{n}D_jf(x)h_j\right|\leq\\&\underbrace{\left|f(x+h)-f(x)-\sum_{j=1}^{n}D_jf(x+h^{(j-1)}+t_jh_j)\right|}_{=0\text{ via telescoping}}+\\&+\sum_{j=1}^{n}\left|D_jf(x+h^{(j-1)})-D_jf(x)\right|\,\|h\|\leq\epsilon\|h\|.\end{aligned}$$

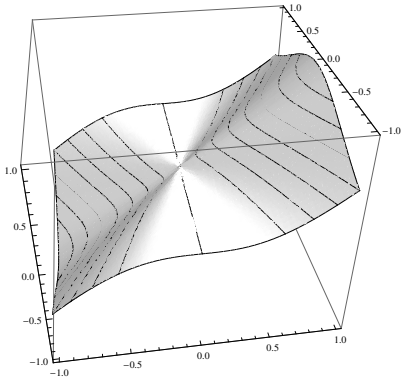
This concludes our proof. □

The Theorem above gives us an insight that continuous differentiability plays a crucial role in analysis in several variables.

Definition 2.10. A function $f:\Omega\to\mathbb{R}^n$ is called *continuously differentiable* if all partial derivatives D_jf_i exist and are continuous on Ω.

Applying the above propositions it is clear that the full differential $A(x) = Df(x) = a_{i,j}(x)$ is given by the expression $a_{i,j}(x) = D_j f_i(x)$ so the total differential of f also exists and is continuous on Ω when intended as a map $Df : \Omega \to \mathbb{R}^n \times \mathbb{R}^m = M_{n\times m}(\mathbb{R})$ where $M_{n\times m}(\mathbb{R})$ are the $n \times m$ matrixes with real entries.
Let us turn back to our example:

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



Let us calculate the partial derivatives:

$$\begin{aligned} D_1 f(x,y) &= \frac{3x^2(x^2+y^2) - 2x^4}{(x^2+y^2)^2} && \text{when } (x,y) \neq 0 \\ D_1 f(0,0) &= 1 && \text{since } f(x,0) = x \\ D_2 f(x,y) &= \frac{-2x^3 y}{(x^2+y^2)^2} && \text{when } (x,y) \neq 0 \\ D_2 f(0,0) &= 0 && \text{since } f(0,y) = 0; \end{aligned}$$

so $D_1 f$ and $D_2 f$ are not continuous in 0. For example we can check that $D_2 f(t,t) = \frac{-2t^4}{4t^4} = -\frac{1}{2} \neq 0 = D_2 f(0,0)$ and this contradicts continuity when $t \to 0$.

Theorem 2.11 (Schwarz Theorem / Clairant Theorem)**.** *Let $f : \Omega \to \mathbb{R}$ be a twice continuously differentiable (i.e. $D_j f$ are all continuously differentiable) then the partial derivatives commute i.e. $D_i D_j f(x) = D_j D_i f(x)$ for all $i, j \in \{1, \ldots, n\}$.*

Proof. Let $\epsilon > 0$ and choose $\delta > 0$ using the continuity assumptions on the second derivatives so that for all vectors $\|h\| < \delta$ with $x+h \in \Omega$ we have that both $|D_iD_jf(x+h) - D_iD_jf(x)| \leq \frac{\epsilon}{10}$ and $|D_jD_if(x+h) - D_jD_if(x)| \leq \frac{\epsilon}{10}$.

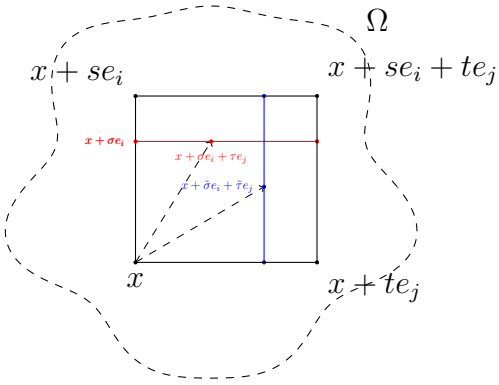


Figure 6: $D_jD_if(x + \sigma e_i + \tau e_j) = D_iD_jf(x + \tilde{\sigma} e_i + \tilde{\tau} e_j)$

Consider the expression

$$f(x + se_i + te_j) - f(x + se_i) - f(x + te_j) + f(x)$$

that approximates the double partial derivatives in the sense that it represents the increment along e_j of the increment along e_i and vice-versa. It is precisely this symmetry that is the base idea of this proof.
Set $g(s) = f(x+se_i+te_j) - f(x+se_i)$. We have that $f(x+se_i+te_j) - f(x+se_i) - f(x+te_j) + f(x) = g(s) - g(0) = sg'(\sigma)$ for some $\sigma \in (0, s)$ given by the Lagrange Theorem about the derivative in the intermediate point. Here we are using the condition on the differentiability of f. Since deriving g is equivalent to taking the partial derivative in direction e_i of f we have that

$$g(s) - g(0) = (D_if(x + \sigma e_i + te_j) - D_if(x + \sigma e_i))\, s = D_jD_if(x + \sigma e_i + \tau e_j)st$$

by once again applying the Lagrange Theorem now in the direction e_j. Using the same argument but inverting the order of the directions we get that

$$f(x + se_i + te_j) - f(x + se_i) - f(x + te_j) + f(x) = D_iD_jf(x + \tilde{\sigma} e_i + \tilde{\tau} e_j)st.$$

But this means that $D_jD_if(x+\sigma e_i+\tau e_j) = D_iD_jf(x+\tilde{\sigma} e_i+\tilde{\tau} e_j)$. Using the continuity assumptions on both the second partial derivatives we get that

$$\begin{aligned} |D_iD_jf(x) - D_jD_if(x)| \leq |D_iD_jf(x) - D_iD_jf(x + \tilde{\sigma} e_i + \tilde{\tau} e_i)| + \\ |D_jD_if(x + \sigma e_i + \tau e_j) - D_jD_if(x)| \leq 2\frac{\epsilon}{10} \end{aligned}$$

by choosing s, t small enough so that $\|se_i + te_j\| < \delta$ and thus both $\|\sigma e_i + \tau e_j\| < \delta$ and $\|\tilde{\sigma} e_i + \tilde{\tau} e_j\| < \delta$.
Since the choice of $\epsilon > 0$ was arbitrary we can conclude that $D_i D_j f(x) = D_j D_i f(x)$ as required. □

Going back once again to our example

$$f(x,y) = \begin{cases} \dfrac{x^3}{x^2+y^2} & \text{if } (x,y) \neq 0 \\ 0 & \text{if } (x,y) = 0 \end{cases}$$

we have that for $(x,y) \neq (0,0)$

$$D_2 D_1 f(x,y) = \frac{6x^2y(x^2+y^2)^2 - 4(x^4+3x^2y^2)(x^2+y^2)y}{(x^2+y^2)^4}$$
$$D_1 D_2 f(x,y) = \frac{-6x^2y(x^2+y^2)^2 + 8x^4y(x^2+y^2)}{(x^2+y^2)^4}$$

By explicit computation one can see that the double partial derivatives coincide. However this is true by the theorem as long as we are away from the point where the partial derivatives fail to be continuous i.e. $(x,y) = (0,0)$.

End of lecture 8. May 4, 2015

To deal with higher order partial derivatives we set up some convenient notation. A *(n-dimensional) multiindex* is a n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n_{\geq 0}$. We write

$$D^\alpha f := D_1^{\alpha_1} D_2^{\alpha_2} \ldots D_n^{\alpha_n} = \underbrace{D_1(\ldots D_1}_{\alpha_1}(\underbrace{D_2(\ldots D_2}_{\alpha_2}(\ldots(\underbrace{D_n(\ldots D_n}_{\alpha_n} f))\ldots))))$$

and note $D^0 f = f$. One also writes

$$D^\alpha f := \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \ldots \partial_{x_n}^{\alpha_n}} f.$$

For instance,

$$D_1 D_2 f = \frac{\partial^2}{\partial_{x_1} \partial_{x_2}} f.$$

We denote the sum of the components of a multi-index α by $|\alpha| := \sum_{i=1}^n \alpha_i$. The following theorem is a generalization of Theorem 2.11 by Schwarz.

Theorem 2.12. *Let $\alpha, \beta \in N^n_{\geq 0}$ be multi-indices Let $f : \Omega \to \mathbb{R}$ be $(|\alpha|+|\beta|)$-times continuously differentiable. Then*

$$D^\alpha(D^\beta f) = D^\beta(D^\alpha f) = D^{\alpha+\beta} f.$$

Proof. We start with induction on $|\beta|$. If $|\beta| = 0$,

$$D^\alpha(D^\beta f) = D^\alpha f = D^\beta(D^\alpha f) = D^{\alpha+\beta} f.$$

Now consider $|\beta| = 1$. Then $\beta = (b_1, \ldots, b_n)$ with $b_j = 1$ for some $1 \leq j \leq n$, and $b_k = 0$ for all $k \neq j$. Thus $D^\beta = D_1^0 D_2^0 \ldots D_{j-1}^0 D_j^1 D_{j+1}^0 \ldots D_n^0 = D_j$ for some $1 \leq j \leq n$, and so $D^\alpha(D^\beta f) = D^\alpha(D_j f)$. Now we induct once more, this time on $|\alpha|$. The case $|\alpha| = 0$ is clear as above. Consider $|\alpha| = 1$, and so $D^\alpha = D_i$ for some $1 \leq i \leq n$. By Theorem 2.11 we have the desired identity

$$D^\alpha(D^\beta f) = D_i D_j f = D_j D_i f = D^\beta(D^\alpha f). \tag{15}$$

It remains to compute $D^{\alpha+\beta}$. If $i \neq j$, the multi-index $\alpha + \beta$ has i-th and j-th component equal to 1, while the other entries equal 0. Then, if $i < j$, we have $D^{\alpha+\beta} = D_i D_j$. If $i > j$, $D^{\alpha+\beta} = D_j D_i$. If $i = j$, $\alpha + \beta$ has i-th component equal to 2 and $D^{\alpha+\beta} = D_i^2$. Using (15), the claim follows.
We remain in the case $|\beta| = 1$ proceed with the induction step for α. Assume that we already know the theorem for $|\alpha| = n$ and let $|\alpha| = n + 1$. Let k be the highest index with $\alpha_k \neq 0$. We write $D^\alpha f = D^{\alpha'} D_k f$. Thus,

$$D^\alpha D^\beta f = D^{\alpha'}(D_k(D_j f))$$

Since $|\alpha'| + 1 = n$, by the induction hypothesis this equals

$$D_j(D^{\alpha'}(D_k f) = D_j D^\alpha f = D^\beta D^\alpha f$$

which is the needed identity. To finish the proof it remains to perform the induction step for $|\beta|$. This proceeds in a very similar way as the induction step for $|\alpha|$ and we leave it as an exercise. □

Recall that for $v \in \mathbb{R}^n$ and for a totally differentiable function $f : \Omega \to \mathbb{R}$, the derivative at $x \in \Omega$ in the direction v is defined as $D_v f(x) := g'(0)$, where $g : I \to \mathbb{R}$, $I = \{t \in \mathbb{R} : x + tv \in \Omega\}$, is given by $g(t) := f(x + tv)$. From the chain rule (which will be discussed in the following lecture) it follows

$$g'(t) = \sum_{j=1}^{n} D_j f(x + tv) v_j.$$

The right hand-side is the product of the $1 \times n$ matrix $Df(x)$ with the $n \times 1$ vector v. In particular we have $D_v f(x) = \sum_{j=1}^n D_j f(x) v_j$. Computing $D_v \widetilde{f}(x)$ where $\widetilde{f}(x) := f(x+tv)$ we obtain[2] $D_v f(x+tv) = g'(t) = \sum_{j=1}^n D_j f(x+tv) v_j$. Now we consider higher order directional derivatives. For $v \in \mathbb{R}^n$ we use the multi-index notation $v^\alpha := v_1^{\alpha_1} \dots v_n^{\alpha_n}$. We denote $\alpha! := \prod_{j=1}^n (\alpha_j)!$

Theorem 2.13. *For a k-times continuously differentiable $f : \Omega \to \mathbb{R}$,*

$$D_v^k f(x+tv) \overset{(1)}{=} g^{(k)}(t) \overset{(2)}{=} \sum_{\alpha : |\alpha| = k} \frac{k!}{\alpha!} D^\alpha f(x+tv) v^\alpha.$$

Proof. (1): By induction. Case $k = 1$ is discussed above. Assume that we know (1) for some k and consider

$$D_v^{k+1} f(x+tv) = D_v (D_v^k f)(x+tv)$$

Denote $\widetilde{f} := D_v^k f$ and $\widetilde{g}(t) := \widetilde{f}(x+tv)$. Using the result for $k = 1$ we have

$$D_v (D_v^k f)(x+tv) = D_v (\widetilde{f})(x+tv) = \widetilde{g}'(t)$$

By the induction hypothesis $\widetilde{g} = D_v^k f = g^{(k)}$, so

$$\widetilde{g}'(t) = (g^{(k)})'(t) = g^{(k+1)}(t).$$

Now we prove (2) by induction. Case $k = 1$ is the exercise above. Assuming the claim for k we consider $g^{(k+1)}(t)$. By the induction hypothesis and part (1) this equals

$$D_v \Big(\sum_{|\alpha|=k} \frac{k!}{\alpha!} (D^\alpha f)(x+tv) v^\alpha \Big) = \sum_{j=1}^n D_j \Big(\sum_{|\alpha|=k} \frac{k!}{\alpha!} (D^\alpha f)(x+tv) v^\alpha \Big) v_j$$

For each $1 \le j \le n$ we move D_j under the sum, which we can do by linearity of derivative. We can also move v_j inside the bracket. Then

$$D_j (D^\alpha f)(x+tv) v^\alpha v_j = (D^\beta f)(x+tv) v^\beta$$

where β is the multi-index with $\beta_l = \alpha_l$ for all $l \neq j$, and the $j-$th component is $\beta_j = \alpha_j + 1$. Summing in j and reshuffling the sum we obtain

$$\sum_{|\beta|=k+1} \Big(\sum_{j=1}^n \sum_{\substack{|\alpha|=k \\ \alpha_l = \beta_l, l \neq j \\ \alpha_j + 1 = \beta_j}} \frac{k!}{\alpha!} \Big) (D^\beta f)(x+tv) v^\beta$$

[2]Be careful not to confuse $D_v f(x+tv)$ with $D_v \widetilde{f}(x)$ where $\widetilde{f}(x) = f(x+tv)$. By $D_v f(x+tv)$ we denote the derivative $D_v f$ evaluated at $x+tv$ and more precisely it should be written as $(D_v f)(x+tv)$.

For a fixed β it remains to sum up the double sum in brackets. Note that there is only one α such that $|\alpha| = k, \alpha_l = \beta_l, l \neq j$ and $\alpha_j + 1 = \beta_j$. Since $\beta_j = \alpha_j + 1 = \frac{\beta!}{\alpha!}$ the expression in brackets equals

$$\sum_{j=1}^{n} \sum_{\substack{|\alpha|=k \\ \alpha_l=\beta_l, l\neq j \\ \alpha_j+1=\beta_j}} \frac{k!\beta_j}{\beta!} = \sum_{j=1}^{n} \frac{k!\beta_j}{\beta!} = \frac{k!|\beta|}{\beta!} = \frac{(k+1)!}{\beta!}.$$

This finishes the proof. □

2.1 Taylor's theorem in $\mathbb{R}^n$

In Analysis I we showed the following. If $f : I \to \mathbb{R}$ is $(N+1)-$times continuously differentiable, where I is an interval in $\mathbb{R}$ containing 0, for every $x \in I$ there exists $\vartheta \in I$, $0 < |\vartheta| < |x|$, such that

$$f(x) = \sum_{k=0}^{N} \frac{1}{k!} f^{(k)}(0)x^k + \frac{1}{(N+1)!} f^{(N+1)}(\vartheta)x^{N+1} \tag{16}$$

This statement is called *Taylor's theorem*. It gives an approximation of f with the N-th order polynomial $\sum_{k=0}^{N} \frac{1}{k!} f^{(k)}(0)x^k$, which is called the *Taylor polynomial of* f. Note that for $N = 0$, (16) is exactly the mean value theorem for f on $[0, x]$: it yields existence of $\vartheta \in (0, x)$ such that

$$f'(\vartheta) = \frac{f(x) - f(0)}{x}.$$

The following generalizes Taylor's theorem to functions $f : \Omega \to \mathbb{R}$.

Theorem 2.14. *Let $f : \Omega \to \mathbb{R}$ be $(N+1)$-times continuously differentiable, $x \in \Omega$ and $x + v \in B_\varepsilon(x)$ for some ball $B_\varepsilon(x) \subset \Omega$. Then there is $\vartheta \in [0, 1]$ such that*

$$f(x+v) = \sum_{|\alpha|\leq N} \frac{1}{\alpha!} D^\alpha f(x) v^\alpha + \sum_{|\alpha|=N+1} \frac{1}{\alpha!} D^\alpha f(x + \vartheta v) v^\alpha$$

Proof. We apply Taylor's theorem in one dimension (16) to the function $g(t) = f(x + tv)$ at $t = 1$, which gives

$$f(x+v) = g(1) = \sum_{k=0}^{N} \frac{1}{k!} g^{(k)}(0) + \frac{1}{(N+1)!} g^{(N+1)}(\vartheta).$$

Then we use Theorem 2.13 (2) for $g^{(k)}(0)$ and $g^{(N+1)}(\vartheta)$ to obtain the desired identity. □

Now we estimate the error in the approximation of f with its Taylor polynomial.[3]

Theorem 2.15. *Let $f : \Omega \to \mathbb{R}$ be N-times continuously differentiable and $x \in \Omega$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ and $B_\delta(x) \subset \Omega$, such that for $x + v \in B_\delta(x)$ we have*

$$\Big| f(x+v) - \sum_{|\alpha| \le N} \frac{1}{\alpha!} D^\alpha f(x) v^\alpha \Big| \le \varepsilon \|v\|^N$$

Proof. If $N = 1$, f is totally differentiable and the statement follows from the definition of total differentiability:

$$\Big| f(x+v) - \Big(f(x) + \sum_{j=1}^{n} D_j f(x) v_j \Big) | \le \varepsilon \|v\|^N.$$

For $N > 1$ we can apply Theorem 2.14 in the case $N - 1$: there exists $\vartheta \in [0, 1]$ such that

$$f(x+v) = \sum_{|\alpha| < N} \frac{1}{\alpha!} D^\alpha f(x) v^\alpha + \sum_{|\alpha| = N} \frac{1}{\alpha!} D^\alpha f(x + \vartheta v) v^\alpha$$

which can be by adding and subtracting $\sum_{|\alpha|=N} \frac{1}{\alpha!} D^\alpha f(x) v^\alpha$ rewritten as

$$f(x+v) - \sum_{|\alpha| \le N} \frac{1}{\alpha!} D^\alpha f(x) v^\alpha = \sum_{|\alpha| = N} \frac{1}{\alpha!} (D^\alpha f(x) - D^\alpha f(x + \vartheta v)) v^\alpha$$

By continuity of $D^\alpha f$ there is a $\delta > 0$ such that for all $x + v \in B_\delta(x)$ (in particular, $x + \vartheta v \in B_\delta(x)$) we have

$$\Big| \sum_{|\alpha| = N} \frac{1}{\alpha!} (D^\alpha f(x) - D^\alpha f(x + \vartheta v)) v^\alpha \Big| \le \varepsilon \|v\|^N$$

where we have estimated $|v^\alpha| = |v_1^{\alpha_1}| \cdots |v_n^{\alpha_n}| \le \|v\|^N$. □

2.2 Chain rule

The following theorem tells us how to express the derivative of the composition of two functions f, g in terms of the derivative of f and the derivative of g.

[3]Compare this with the definition of derivative, which gives an approximation of a function with a linear map. Also note that polynomials of first order (corresponding to case $N = 1$) are linear.

Theorem 2.16. *Let $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^n$ be open subsets. Let $g : U \to V$, $f : V \to \mathbb{R}^k$ be two maps such that g is differentiable at $x \in U$ and f is differentiable at $g(x) \in V$. Then $f \circ g : U \to \mathbb{R}^k$ is differentiable at x and*

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x)$$

Recall that (in the chosen bases) the linear map is represented by a matrix. The composition of these maps on the right hand-side corresponds to the matrix product of the respective matrices.

Proof. Exercise sheet 6, exercise 2. □

End of lecture 9. May 7, 2015

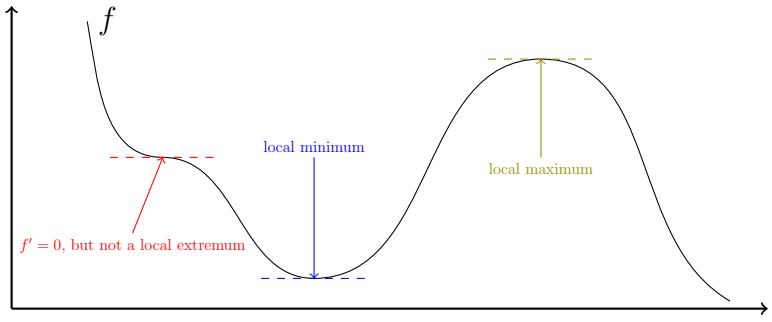


Figure 7: Local extrema and points with vanishing derivatives.

Theorem 2.17. *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable in x and let x be a point of local maximum (or minimum) of f. Then $Df = 0$*

Proof. We reason by contradiction: let $Df(x) \neq 0$ with $Df(x) : \mathbb{R}^n \to \mathbb{R}$. Then there exists a vector $v \in \mathbb{R}^n$ such that $Df(x)v \neq 0$. Let us consider such a vector with $\|v\| = 1$ (it suffices to consider $\frac{v}{\|v\|}$) and by linearity (up to multiplication by -1) we have that $Df(x)v > 0$. Let $\epsilon = \frac{Df(x)v}{2}$ and choose $\delta > 0$ so that

1. $\forall h$, $\|h\| < \delta$ we have that $x + h \in \Omega$ (Ω is open);
2. $\forall h$, $\|h\| < \delta$ we have that $f(x + h) \leq f(x)$ (x is a local maximum);
3. $\forall h$, $\|h\| < \delta$ we have that $\|f(x + h) - f(x) - Df(x)h\| \leq \epsilon\|h\|$.

Now choose $t \in (0, \delta)$ so that we have $t = \|tv\| < \delta$. We have that $\|f(x + tv) - f(x) - Df(x)tv\| < \epsilon t$ and thus $Df(x)vt - (f(x + tv) - f(x)) \leq \epsilon t$ so $(Df(x)v - \epsilon)\, t \leq f(x + tv) - f(x)$ but this is a contradiction because $0 < \frac{1}{2} Df(x)tv \leq (Df(x)v - \epsilon)\, t \leq f(x + vt) - f(x) \leq 0$. □

Now let us study what can happen in a critical point of a function. A critical point is a point $x \in \Omega$ where $Df(x) = 0$.
Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be twice differentiable i.e. let f have second derivatives $D_i D_j f(x)$ in a point x. The expression $D_i D_j f(x)$, $1 \leq i, j \leq n$ defines a symmetric $n \times n$ matrix. This matrix is symmetric by the Schwartz Theorem as long as f has second partial derivatives in an open neighborhood of x that are continuous in x.

Definition 2.18. A matrix $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$ is called symmetric if for all $i, j \in \{1, \ldots, n\}$ one has $a_{i,j} = a_{j,i}$. A symmetric matrix $(a_{i,j})$ is called positive definite if for all $v \in \mathbb{R}^n$ with $v \neq 0$ one has $\sum_{i,j=1}^n a_{i,j} v_i v_j > 0$. If we allow equality even for non-zero vectors, i.e. $\sum_{i,j=1}^n a_{i,j} v_i v_j \geq 0$ the matrix $(a_{i,j})$ is said to be positive semi-definite.

An example of a positive definite matrix is $\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$. This can be checked simply: $\sum_{i,j=1}^n \delta_{i,j} v_i v_j = \sum_{i=1}^n v_i^2 > 0$ if $v \neq 0$.

Theorem 2.19. *A matrix $(a_{i,j}) \in \mathbb{R}^{n \times n}$ is positive definite if and only if there exists an $\epsilon > 0$ such that $\forall v \in \mathbb{R}^n, v \neq 0$ we have that $\sum_{i,j=1}^n a_{i,j} v_i v_j \geq \epsilon \|v\|^2 > 0$.*

Proof. Set $S = \{v \in \mathbb{R}^n \mid \|v\| = 1\}$; S is a bounded and closed subset of $\mathbb{R}^n$ so it is compact.

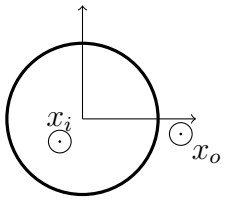


Figure 8: The unit circle S.

Now choose a sequence v^k in S such that $\lim_{k \to \infty} \sum_{i,j=1}^n a_{i,j} v_i^k v_j^k = \inf_{w \in S} \sum_{i,j=1}^n a_{i,j} w_i w_j$. This sequence admits a subsequence converging to some v since S is compact and we have that $v \in S$ since S is closed. Passing

to the limit in the previous equality we get that

$$\lim_{k\to\infty}\sum_{i,j=1}^{n} a_{i,j}v_i^k v_j^k = \underbrace{\sum_{i,j=1}^{n} a_{i,j}v_i v_j}_{=\epsilon} > 0$$

where the last inequality holds because the matrix is positive definite. But $\epsilon = \lim_{k\to\infty}\sum_{i,j=1}^{n} a_{i,j}v_i^k v_j^k = \inf_{w\in S}\sum_{i,j=1}^{n} a_{i,j}w_i w_j$ so $\sum_{i,j=1}^{n} a_{i,j}w_i w_j \geq \epsilon$ for all $w \in S$. If $w \notin S$ consider the vector $w' = \frac{w}{\|w\|} \in S$. Applying the obtained bound to w' we get that $\frac{1}{\|w\|^2}\sum_{i,j=1}^{n} a_{i,j}w_i w_j \geq \epsilon$ as required. □

Theorem 2.20. *Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable in a point $x \in \Omega$ i.e. f has second partial derivatives in an open neighborhood containing x and they are continuous in the point x. If $Df(x) = 0$ and the matrix $(D_iD_jf(x))$, called the Hessian, is positive definite then f has a (strict) local minimum in the point x.*

Proof. Let $\epsilon > 0$ so that $10\epsilon \leq \inf_{w\in S}\sum_{i,j=1}^{n} D_iD_jf(x)w_iw_j$. Such an ϵ exists due to the previous theorem. Using the Taylor expansion we get that there exists a $\delta > 0$ such that $\forall v$ with $\|v\| < \delta$ we have that $x + v \in \Omega$ and

$$\left| f(x+v) - f(x) - \underbrace{Df(x)v}_{=0} - \underbrace{\sum_{|\alpha|=2}\frac{1}{\alpha!}D^\alpha f(x)v^\alpha}_{\frac{1}{2}\sum_{i,j=1}^{n} D_iD_jf(x)v_iv_j} \right| \leq \epsilon\|v\|^2.$$

However, since the second differential is positive definite we have that

$$\begin{aligned} \frac{1}{2}\sum_{i,j=1}^{n} D_iD_jf(x)v_iv_j &> 5\epsilon\|v\|^2 \\ f(x+v) - f(x) - \frac{1}{2}\sum_{i,j=1}^{n} D_iD_jf(x)v_iv_j &\geq -\epsilon\|v\|^2 \\ \Rightarrow \qquad f(x+v) - f(x) &\geq 4\epsilon\|v\|^2. \end{aligned}$$

Moreover we have proven that the maximum is strict in the sense that $\forall v$ with $\|v\| < \delta$ and $v \neq 0$ then the inequality is strict: $f(x + v) > f(x)$. □

A combinatorial comment is due about the above application of the multivariable Taylor expansion. Let us consider the expressions that appears in

the Taylor expansion:

$$\sum_{|\alpha|=2}\frac{1}{\alpha!}D^\alpha f v^\alpha = \overbrace{\sum_{i<j}\frac{1}{1!1!}D_iD_jf v_iv_j}^{\alpha=(0,\dots,0,\overbrace{1}^{i},0,\dots,0,\overbrace{1}^{j},0,\dots,0)} + \overbrace{\sum_{i=j}\frac{1}{2!}D_iD_if v_iv_i}^{\alpha=(0,\dots,0,\overbrace{2}^{i},0\dots,0)} = \frac{1}{2}\sum_{i,j=1}^{n} D_iD_jf\, v_iv_j.$$

Further generalizations are just combinatorial induction arguments. For example

$$\sum_{|\alpha|=3}\frac{1}{\alpha!}D^\alpha f v^\alpha = \frac{1}{3!}\sum_{i,j,k=1}^{n} D_iD_jD_kf\, v_iv_jv_k.$$

Theorem 2.21 (Sylvester criterion)**.**
Let $A = \left[\begin{array}{c|c} A_{n-1} & v \\ \hline v^t & a \end{array}\right]$ *be an* $n\times n$ *symmetric matrix. Then* A *is positive definite if and only if* A_{n-1} *is positive definite and* $\det(A)>0$.

Proof. $\Leftarrow$ First of all one can check that $A = S^t\left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]S$ where the matrix S is given by $S = \left[\begin{array}{c|c} \mathrm{I}_{n-1} & A_{n-1}^{-1}v \\ \hline 0 & 1 \end{array}\right]$ and $b = a - v^tA_{n-1}^{-1}v$. Notice that a positive definite matrix is necessarily invertible so A_{n-1}^{-1} is well defined. The matrix S^t is the transposed matrix of S: $S^t_{i,j} = S_{j,i}$. As a matter of fact

$$\overbrace{\underbrace{\left[\begin{array}{c|c} \mathrm{I}_{n-1} & 0 \\ \hline v^tA_{n-1}^{-1} & 1 \end{array}\right]}_{\left[\begin{array}{c|c} A_{n-1} & v \\ \hline 0 & b \end{array}\right]}\left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]\left[\begin{array}{c|c} \mathrm{I}_{n-1} & A_{n-1}^{-1}v \\ \hline 0 & 1 \end{array}\right]}^{\left[\begin{array}{c|c} A_{n-1} & v \\ \hline v^t & v^tA_{n-1}^{-1}v+b \end{array}\right]} = A.$$

We thus have that $\sum_{i,j=1}^n A_{i,j}v_iv_j = v^tAv = v^tS^t\left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]Sv = (Sv)^t\left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]Sv$. Notice that S is invertible so A is positive definite if and only if $\left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]$ is positive definite. Let us show this

last fact via computation. Choose any $w \in \mathbb{R}^n$ with $w \neq 0$:

$$\left[\begin{array}{c|c} w_{n-1}^t & w_n \end{array}\right] \left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right] \left[\begin{array}{c} w_{n-1} \\ \hline w_n \end{array}\right] = w_{n-1}^t A_{n-1} w_{n-1} + w_n^2 b.$$

We require that the left hand side be positive. Since A_{n-1} is positive definite the term $w_{n-1}^t A_{n-1} w_{n-1}$ is always non-negative. Taking in account that $\det A = \det S^t \det \left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right] \det S$ but $\det S = \det S^t = 1$ so we have that $\det A = \det \left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]$. The determinant of a block matrix is given by $\det \left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right] = b \det A_{n-1}$. So the condition $\det A > 0$ implies $b > 0$ and thus $\det \left[\begin{array}{c|c} A_{n-1} & 0 \\ \hline 0 & b \end{array}\right]$ is positive definite.

$\Rightarrow$ The inverse implication can be shown reasoning by contradiction. Testing with vectors $w \in \mathbb{R}^n$ with w_n yields that A_{n-1} has to be positive definite and thus invertible. Then one applies the same decomposition as before. □

An induction argument suggests to introduce the following notions. A principle minor $[A]_{I,I}$ of an $n \times n$ matrix A is a matrix determined by a subset of indexes $I \subset \{1, \ldots, n\}$ such that $([A]_{I,I})_{i,j} = A_{k_i,k_j}$ where k_i and k_j are respectively the i^{th} and j^{th} elements in order of I. A leading principle minor (of order m) is the principle minor $[A]_{\{1,\ldots,m\},\{1,\ldots,m\}}$

Theorem 2.22. *A symmetric matrix A is positive definite if and only if all the principle determinants are positive. Furthermore the same is true if one considers only the leading principle determinants.*

An example of the application of the above theorems is the function on $\mathbb{R}^2$ given by $f(x, y) = x^2 + y^2$. We have that $Df(0) = 0$ while $(D_i D_j f(x)) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ for all $x \in \mathbb{R}^n$. This matrix is definite. We can thus deduce that f has a local minimum in $(0, 0)$. Something completely different happens for the function $f(x, y) = x^2 - y^2$. We have that $D_i D_j f(x) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ Differently from the one-dimensional case even if a critical point x is such that $Df(x) = 0$ and $D^2 f(x)$ is non-degenerate this still is not sufficient to conclude the presence of a local extremum. The behavior of the function f can be different in different directions.

End of lectures 10. May 11, 2015

Lectures 11 and 12 were held by Roland Donninger.

2.3 Banach fixed point theorem

Let X be a metric space. Throughout this section we shall assume that the metric is induced by a norm, although all of the following holds if we replace $\|\cdot\|$ by a general metric.
A map $\phi : X \to X$ is called a *contraction*, if there exists $\alpha \in (0,1)$ such that for all $x, y \in X$ we have

$$\|\phi(x) - \phi(y)\| \leq \alpha \|x - y\|.$$

Theorem 2.23 (Banach fixed point theorem)**.** *Let $X \neq \emptyset$ be a complete metric space and let $\phi : X \to X$ be a contraction. Then there exists a unique fixed point of ϕ, that is, a unique point $x \in X$ such that $\phi(x) = x$.*

Proof. Let $x_0 \in X$. Define a sequence (x_n) in X via the recursive relation $x_n := \phi(x_{n-1})$ for $n \in \mathbb{N}$. First we show that (x_n) converges. Let $n, k \in \mathbb{N}$. Then

$$\|x_{n+k} - x_n\| = \|\phi(x_{n-1+k}) - \phi(x_{n-1})\| \leq \alpha \|x_{n-1+k} - x_{n-1}\|,$$

where in the last inequality we used that ϕ is a contraction. We can again write $\|x_{n-1+k} - x_{n-1}\| = \|\phi(x_{n-2+k}) - \phi(x_{n-2})\|$ and apply the contraction property to obtain

$$\|x_{n+k} - x_n\| \leq \alpha^2 \|x_{n-2+k} - x_{n-2}\|.$$

Performing this step n times we arrive to

$$\|x_{n+k} - x_n\| \leq \alpha^n \|x_k - x_0\|. \tag{17}$$

Now consider the right hand-side of (17). By the triangle inequality we have

$$\begin{aligned} \|x_k - x_0\| = \|x_k - x_1 + x_1 - x_0\| &\leq \|x_k - x_1\| + \|x_1 - x_0\| \\ &\leq \|x_k - x_2\| + \|x_2 - x_1\| + \|x_1 - x_0\| \end{aligned}$$

Iterating we obtain

$$\|x_k - x_0\| \leq \sum_{j=0}^{k-1} \|x_{j+1} - x_j\| \leq \sum_{j=0}^{k-1} \alpha^j \|x_1 - x_0\|.$$

Altogether we have for (17) that for each $k, n \in \mathbb{N}$

$$\begin{aligned}\|x_{n+k} - x_n\| &\leq \alpha^n \sum_{j=0}^{k-1} \alpha^j \|x_1 - x_0\| \\ &= \alpha^n \|x_1 - x_0\| \sum_{j=0}^{k-1} \alpha^j \\ &\leq \alpha^n \|x_1 - x_0\| \sum_{j=0}^{\infty} \alpha^j = \frac{\alpha^n}{1-\alpha} \|x_1 - x_0\|.\end{aligned}$$

In the last step we used that $\alpha < 1$, so $\sum_{j=0}^{\infty} \alpha^j$ is a convergent geometric series. Now, $\alpha \in (0,1)$ implies that $\alpha^n \to 0$ as $n \to \infty$, so the sequence (x_n) is a Cauchy sequence in X. Since X is complete, (x_n) converges. That is, there exists $a \in X$ such that $x_n \to a$.
We claim that the limit a is the fixed point of ϕ. To see this we consider

$$\|\phi(a) - a\| \leq \|\phi(a) - \phi(x_n)\| + \|\phi(x_n) - a\| \leq \alpha \|a - x_n\| + \|x_{n+1} - a\|.$$

Since $x_n \to a$ as $n \to \infty$, we that $\alpha\|a - x_n\| + \|x_{n+1} - a\| \to 0$ as n tends to ∞. Thus, $\|\phi(a) - a\| = 0$ which is equivalent to $\phi(a) = a$. So a is really a fixed point of ϕ.
It remains to show uniqueness. Assume there is another fixed point b of ϕ. That is, there is $b \in X$ such that $\phi(b) = b$. Then

$$\|a - b\| = \|\phi(a) - \phi(b)\| \leq \alpha \|a - b\|.$$

This implies $(1 - \alpha)\|a - b\| \leq 0$. Since $\alpha < 1$, we have $(1 - \alpha) > 0$. A norm is always non-negative, so we must have $\|a - b\| = 0$, i.e. $a = b$. □

Remark. From the proof it follows that the sequence $x_n = \phi(x_{n-1})$ converges to the unique fixed point of ϕ for an *arbitrary* initial value $x_0 \in X$.

Example. Let $X = [1, 2] \subset \mathbb{R}$. This is a complete metric space with norm being the absolute value. For $x \in X$ consider

$$\phi(x) := \frac{x+2}{x+1}$$

We show that ϕ satisfies the assumptions of the Banach fixed point theorem.

- The function ϕ maps X to itself, i.e. $\phi : X \to X$:
  This holds since $1 \leq x \leq 2$ implies the bounds

$$\phi(x) \geq \frac{x+2}{x+2} = 1$$

and

$$\phi(x) \leq \frac{2+2}{x+1} \leq \frac{4}{2} = 2.$$

Thus, $1 \leq \phi(x) \leq 2$.

- ϕ is a contraction:

$$\phi(x) - \phi(y) = \frac{x+2}{x+1} - \frac{y+2}{y+1} = \frac{y-x}{(x+1)(y+1)}$$

Bounding x from above and below gives

$$|\phi(x) - \phi(y)| \leq \frac{1}{4}|x-y|.$$

Thus, the sequence $x_n = \phi(x_{n-1})$ converges to the fixed point $x_* = \phi(x_*)$ for an arbitrary initial value x_0. The fixed point satisfies

$$x_* = \frac{x_*+2}{x_*+1} \;\Rightarrow\; x_*^2 = 2 \;\Rightarrow\; x_* = \sqrt{2} \in [1,2].$$

This example yields explicit approximations of $\sqrt{2}$.

2.4 Inverse function theorem

From Analysis I we recall the following fact. Let $f : \mathbb{R} \to \mathbb{R}$ be continuously differentiable with $f'(x_0) \neq 0$ for some $x_0 \in \mathbb{R}$. Then there exists an interval $I = (x_0 - \varepsilon, x_0 + \varepsilon)$ such that f is monotone on I. In particular, $f|_I : I \to f(I)$ is bijective, so there is a local inverse[4] $f^{-1} : f(I) \to I$. Moreover, f^{-1} is differentiable in $f(x_0)$ and by the chain rule we have

$$(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}.$$

The following theorem generalizes this fact to functions $f : \mathbb{R}^n \to \mathbb{R}^n$.

Theorem 2.24 (Inverse function theorem)**.** *Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^n$ be continuously differentiable. Let $a \in U$ and let the matrix $Df(a)$ be invertible. Then there exists an open set $V \subset U$ with $a \in V$ and an open set $W \subset \mathbb{R}^n$ with $b := f(a) \in W$ such that f is a bijection from V to W. The inverse mapping $g := (f|_V)^{-1}$ is differentiable in b and*[5]

$$Dg(f(a)) = Df(a)^{-1}.$$

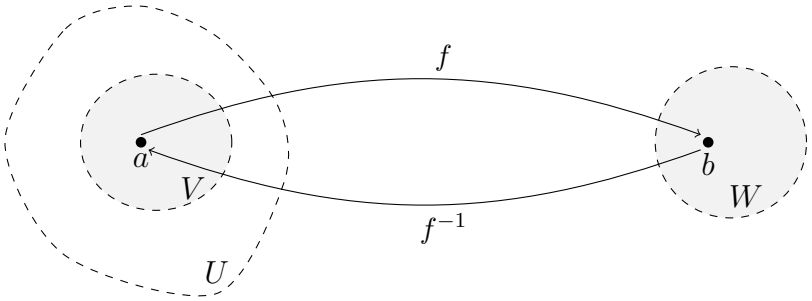


Figure 9: Inverse function theorem

Corollary 2.25. *If f is k-times continuously differentiable, so is g.*

Before proving Theorem 2.24 we state the following lemma, which can be seen as a converse of Theorem 2.24.

Lemma 2.26. *Let $U \subset \mathbb{R}^n$ be open and $V \subset \mathbb{R}^k$ open. Let $f : U \to V$ be bijective and let both f and f^{-1} be continuously differentiable. Then $Df(x) : \mathbb{R}^n \to \mathbb{R}^k$ is at each $x \in U$ non-singular (invertible). In particular, $k = n$.*

A bijective map f such that both f and f^{-1} are continuously differentiable is called a C^1-*diffeomorphism*.

Proof of Lemma 2.26. We have $\mathrm{id}_U = f \circ f^{-1}$. Applying the chain rule we obtain that $I = Df^{-1}(f(x))Df(x)$ for each $x \in U$. □

Remark. A natural question is whether one still has $k = n$ if one weakens the assumption of f and f^{-1} being continuously differentiable to just being continuous. The answer is yes. If $f : U \to V$ is bijective and both f and f^{-1} are continuous, then necessarily $k = n$. Such a map f is called a *homeomorphism*. This result is much harder to prove and the proof uses tools from algebraic topology.

Let us now discuss the idea of proof of Theorem 2.24, which will be integrated into a detailed proof afterwards. Assume that $a = b = 0$, that is, $f(0) = 0$. Assume also $Df(0) = I$. So near 0, f is approximately the identity map. Consider

$$f(x) = x + h(x)$$

where for small $\|x\|$, the functions h is small. We also have $Dh(0) = Df(0) - I = 0$. Therefore we should expect that for small x, y we have $\|h(x)-h(y)\| \leq$

[4]For the inverse one should more precisely write $(f|_I)^{-1}$.

[5]$Df(a)^{-1}$ is the matrix inverse of $Df(a)$

$\frac{1}{2}\|x-y\|$, so h is a contraction. (If $n=1$, this would follow from the mean value theorem $|h(x)-h(y)| = |h'(\xi)|(x-y)| \leq \frac{1}{2}|x-y|$ as $h'(\xi)$ is small for ξ near 0.) Let z be fixed and small and define a perturbation of h by

$$\phi(x) := z - h(x).$$

Since h is a contraction, ϕ is a contraction and $\|h(x)-h(0)\| \leq \frac{1}{2}\|x-0\|$. Using $h(0)=0$ we get $\|h(x)\| \leq \frac{1}{2}\|x\|$. Thus for small z and small x the function $\phi(x)$ is small, i.e. it maps a ball around zero into itself. Applying the Banach fixed point theorem we find a unique a in a neighborhood of zero such that $\phi(a) = a = z - h(a)$. That is, a unique a such that $z = f(a)$. So near 0, the function f is invertible.

Proof of Theorem 2.24 . First we make the simplifications which we already assumed in the above discussion. Let us show that we may assume $a = b = 0$. Set $\widetilde{f}(x) := f(x) - b$. Then for each $x \in U$ we have $D\widetilde{f}(x) = Df(x)$. Assuming $\widetilde{f}^{-1}$ exists, we have $f^{-1}(y) = \widetilde{f}^{-1}(y-b)$. Indeed, this follows from

$$\begin{aligned} f^{-1}(f(x)) &= \widetilde{f}^{-1}(f(x)-b) = \widetilde{f}^{-1}(\widetilde{f}(x)) = x \\ f(f^{-1}(x)) &= f(\widetilde{f}^{-1}(y-b)) = \widetilde{f}(\widetilde{f}^{-1}(y-b)) + b = y. \end{aligned}$$

Thus, it suffices to prove the theorem for $\widetilde{f}$. Since $\widetilde{f}(a) = f(a) - b$, we may then assume $b = 0$. Similarly, we can consider the function $\widetilde{f}(x) = f(x+a)$ to see that we may suppose $a = 0$. So our assumptions for now are that $f(0) = 0$ and $Df(0)$ is invertible. Without loss of generality we may then suppose $Df(0) = I$, as otherwise we replace $f(x)$ by $Df^{-1}(0)f(x)$.

Define now $h(x) := f(x) - x$. Our goal is to show that h is a contraction. We have $h(0) = 0$ and $Dh(0) = Df(0) - I = 0$. This implies that all partial derivatives of h at 0 are 0, i.e. $D_j h(0) = 0$ for each $j = 1, \ldots, n$. Since h is continuously differentiable, there exists a ball $\overline{B_r(0)} \subset U$ such that $\|D_j h(x)\| \leq \frac{1}{2n^2}$ for each $x \in \overline{B_r(0)}$. Let now $x \in \overline{B_r(0)}$ and $v \in \mathbb{R}^n$. Consider the directional derivative $D_v h$ and estimate

$$\|D_v h(x)\| = \Big\| \sum_{j=1}^n v_j D_j h(x) \Big\| \leq \sum_{j=1}^n |v_j| \|D_j h(x)\| \leq \sum_{j=1}^n \|v\| \frac{1}{2n^2} \leq \|v\| \frac{1}{2n}.$$

By the fundamental theorem of calculus we have

$$h(y) - h(x) = \int_0^1 \frac{d}{dt} h(x + t(y-x))dt = \int_0^1 D_{y-x} h(x + t(y-x))dt.$$

The integral on the right hand side is an integral of the vector valued function $D_{y-x}h(x+t(y-x))$. (Just as an reminder, in each component it equals $\frac{d}{dt}h_j(x+t(y-x)) = \langle Dh_j(x+t(y-x)), y-x\rangle$.) One understands such integrals of vector valued continuous functions $\varphi : [0,1] \to \mathbb{R}^n$ componentwise. That is,

$$\int_0^1 \varphi(t)dt = \int_0^1 (\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t))dt$$
$$:= (\int_0^1 \varphi_1(t)dt, \int_0^1 \varphi_2(t)dt, \dots, \int_0^1 \varphi_n(t)dt) = \sum_{j=1}^n \int_0^1 \varphi_j(t)dt\, e_j$$

where e_j are the standard unit vectors $e_1 = (1,0,\dots,0)$, $e_2 = (0,1,0,\dots,0), \dots,$ $e_n = (0,\dots,0,1)$. Therefore,

$$\begin{aligned}\Big\| \int_0^1 \varphi(t)dt\Big\| &= \Big\| \sum_{j=1}^n \int_0^1 \varphi_j(t)dt\, e_j\Big\| \\ &\le \sum_{j=1}^n \Big\| \int_0^1 \varphi_j(t)dt\, e_j\Big\| \\ &= \sum_{j=1}^n \Big| \int_0^1 \varphi_j(t)dt\Big| \|e_j\| \\ &\le \sum_{j=1}^n \int_0^1 |\varphi_j(t)|dt \le \sum_{j=1}^n \int_0^1 \|\varphi(t)\|dt \le n\int_0^1 \|\varphi(t)\|dt.\end{aligned}$$

Applying this to $\varphi(t) = D_{y-x}h(x+t(y-x))$ we obtain

$$\begin{aligned}\|h(y)-h(x)\| &= \Big\| \int_0^1 D_{y-x}h(x+t(y-x))dt\Big\| \\ &\overset{(*)}{\le} n\int_0^1 \|D_{y-x}h(x+t(y-x))\|dt \le n\frac{1}{2n}\|y-x\| = \frac{1}{2}\|y-x\|,\end{aligned}$$

where in $(*)$ we used our previously established bound on $\|D_v h\|$. This shows that h is indeed a contraction.

Now we show the following two claims.
a) f is injective on $\overline{B_r(0)}$.
b)$B_{r/2}(0) \subset f(\overline{B_r(0)})$.

Proof of a). Take $x, y \in \overline{B_r(0)}$ such that $f(x) = f(y)$. We need to show that $x = y$. This follows from the reverse triangle inequality and the fact that h is a contraction:

$$\begin{aligned} 0 = \|f(x) - f(y)\| &= \|x - y - (h(y) - h(x))\| \\ &\geq \|x - y\| - \|h(y) - h(x)\| \\ &\geq \|x - y\| - \frac{1}{2}\|x - y\| = \frac{1}{2}\|x - y\| \end{aligned}$$

So $\|x - y\| \leq 0$, which implies $x = y$.

Proof of b). Let $z \in B_{r/2}(0)$ and set $\phi(x) := z - h(x)$. For $x \in \overline{B_r(0)}$ we have

$$\|\phi(x)\| \leq \|z\| + \|h(x)\| \leq \frac{r}{2} + \|h(x) - h(0)\| \leq \frac{r}{2} + \frac{1}{2}\|x - 0\| \leq \frac{r}{2} + \frac{r}{2} = r.$$

This shows that ϕ maps $\overline{B_r(0)}$ to itself. Moreover, since h is a contraction, for every $x, y \in \overline{B_r(0)}$ we have

$$\|\phi(x) - \phi(y)\| = \|h(x) - h(y)\| \leq \frac{1}{2}\|x - y\|$$

Thus ϕ is a contraction on the complete metric space $\overline{B_r(0)}$. By the Banach fixed-point theorem, there exists a unique $a \in \overline{B_r(0)}$ such that $a = \phi(a) = z - h(a)$. That is, $z = a + h(a) = f(a)$. Since z was arbitrary, for each $z \in B_{r/2}(0)$ there exists a unique $a \in \overline{B_r(0)}$ such that $f(a) = z$, which implies $B_{r/2}(0) \subset f(\overline{B_r(0)})$.

Set now $W := B_{r/2}(0)$ and $V := f^{-1}(B_{r/2}(0))$. Since f is continuous and $f(0) = 0$, V is an open neighborhood of 0. Also, $B_{r/2}(0) \subset f(\overline{B_r(0)})$ implies $V = f^{-1}(B_{r/2}(0)) \subset \overline{B_r(0)}$. From a) and b) it follows that $f : V \to W$ is bijective, thus it is invertible.

It remains to show differentiability of f. We know that $Df(0) = I$. If f^{-1} is differentiable in 0, by the chain rule necessarily $Df^{-1}(0) = I$. So we need to show that for each $\varepsilon > 0$ there is a $\delta > 0$ such that for all $\|k\| < \delta, k \in W$ we have

$$\|f^{-1}(k) - f^{-1}(0) - Df^{-1}(0)k\| \leq \varepsilon\|k\|,$$

i.e

$$\|f^{-1}(k) - k\| \leq \varepsilon\|k\|.$$

Let $\varepsilon > 0$. By differentiability of f, there exists a $\delta' > 0$ such that for all $\|k'\| < \delta', k' \in V$ we have $\|k' - f(k')\| \leq \frac{\varepsilon}{2}\|k'\|$. Let $\|k\| < \frac{\delta'}{2}, k \in W$. We have $f^{-1}(k) \in V$. Moreover, for each $x \in B_r(0) \supset V$ we have

$$\|f(x)\| = \|x - (-h(x))\| \geq \|x\| - \|h(x)\| \overset{(*)}{\geq} \frac{1}{2}\|x\|$$

where in $(*)$ we used that h is a contraction. Applying this with $x = f^{-1}(k)$ we obtain

$$\|f^{-1}(k)\| \leq 2\|k\| < \delta'.$$

Set now $k' := f^{-1}(k)$. Since $\|k'\| < \delta', k' \in V$, by differentiability of f

$$\|f^{-1}(k) - k\| = \|k' - f(k')\| \leq \frac{\varepsilon}{2}\|k'\| \leq \varepsilon\|k\|,$$

so f^{-1} is differentiable at 0. This finishes the proof. □

Example. (Polar coordinates)
Let $f : \{(x, y) \in \mathbb{R}^2 : x > 0\} \to \mathbb{R}^2$ be given by

$$f(x, y) = (\sqrt{x^2 + y^2}, \operatorname{atan}\frac{y}{x}) = (f_1, f_2).$$

We compute the partial derivatives of f:

$$D_1 f_1(x, y) = \frac{\partial}{\partial x}\sqrt{x^2 + y^2} = \frac{x}{\sqrt{x^2 + y^2}}, \quad D_2 f_1(x, y) = \frac{\partial}{\partial y}\sqrt{x^2 + y^2} = \frac{y}{\sqrt{x^2 + y^2}}$$

$$D_1 f_2(x, y) = \frac{\partial}{\partial x}\operatorname{atan}\frac{y}{x} = -\frac{y}{x^2 + y^2}, \quad D_2 f_2(x, y) = \frac{\partial}{\partial y}\operatorname{atan}\frac{y}{x} = \frac{x}{x^2 + y^2}$$

The Jacobian of f then equals

$$Df(x, y) = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix}$$

and has the determinant

$$\det Df(x, y) = \frac{x^2}{(x^2 + y^2)^{3/2}} + \frac{y^2}{(x^2 + y^2)^{3/2}} = \frac{1}{\sqrt{x^2 + y^2}} > 0.$$

Thus, at every point $(x, y), x > 0$, the function f is locally invertible.

We remark that local invertibility at every point does *not* imply global invertibility. Nevertheless, the function from this example is globally invertible with the inverse $f^{-1} : (0, \infty) \times (-\frac{\pi}{2}, -\frac{\pi}{2}) \to \{(x, y) \in \mathbb{R}^2 : x > 0\}$ given by

$$f^{-1}(r, \varphi) = (r\cos\varphi, r\sin\varphi).$$

Remark. From the existence of the local inverse it does *not* follow that one can explicitly "write it down". For instance, consider $f : (0, \infty) \to (0, \infty)$ given by $f(x) = xe^x$. For each $x > 0$ we have $f'(x) = (1 + x)e^x > 0$, so f is at locally invertible each point. Moreover, f is even globally invertible, as it is strictly monotone. The inverse function of f is called the *Lambert W function* and it cannot be expressed in terms of elementary functions.

2.5 Implicit function theorem

Suppose we are given an equation $F(x, y) = 0$ for some function $F : \mathbb{R}^2 \to \mathbb{R}$. Question: Is there a (unique) $g : \mathbb{R} \to \mathbb{R}$ such that for each x we have $y = g(x)$, i.e. $F(x, g(x)) = 0$? In other words, can one "solve" the equation for y? If the answer is yes, one says that g is given *implicitly* through F.

Example. Let $F(x, y) = x + y$. From $F(x, y) = 0$ it follows $y = -x$, so $g(x) = -x$.

Example. Let $F(x, y) = x^2 + y^2 - 1$. From $F(x, y) = 0$ it follows $y = \pm\sqrt{1 - x^2}$. If $|x| > 1$, the equation $F(x, y) = 0$ has no solutions. On the other hand, if $|x| < 1$, the equation has two different solutions. If one restricts oneself to the region $y > 0, |x| < 1$, the solution is unique. This example shows that one should consider this problem *locally*.

We make the following observation. If $F(x, g(x)) = 0$ and all functions involved are sufficiently differentiable, then it follows from the chain rule

$$0 = \frac{d}{dx}F(x, g(x)) = D_1F(x, g(x))1 + D_2F(x, g(x))g'(x)$$

and so

$$g'(x) = -\frac{D_1F(x, g(x))}{D_2F(x, g(x))}.$$

Thus, if we want g to be differentiable, the condition $D_2F(x, y) \neq 0$ is necessary. In the following theorem we will see that it is also sufficient.

Let us set up some notation. Let $F : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$. Define

$$F^1_y(x) := F(x, y) \text{ and } F^2_x(y) := F(x, y).$$

For every $y \in \mathbb{R}^m$ we have $F^1_y : \mathbb{R}^k \to \mathbb{R}^m$ and for every $x \in \mathbb{R}^k$ we have $F^2_x : \mathbb{R}^m \to \mathbb{R}^m$. If F is differentiable in $(x, y) \in \mathbb{R}^k \times \mathbb{R}^m$, then we write

$$D_1F(x, y) := DF^1_y(x) \text{ and } D_2F(x, y) := DF^2_x(y).$$

Theorem 2.27 (Implicit function theorem)**.** *Let $U \subset \mathbb{R}^k \times \mathbb{R}^m$ be open and $F : U \to \mathbb{R}^m$ continuously differentiable. Let $(a, b) \in U$ be such that $F(a, b) = 0$ and such that $D_2F(a, b)$ is invertible. Then there is an open neighborhood V_1 of a and an open neighborhood V_2 of b with $V_1 \times V_2 \subset U$, and a continuously differentiable map $g : V_1 \to V_2$, such that $F(x, g(x)) = 0$ for each $x \in V_1$. If $(x, y) \in V_1 \times V_2$ is such that $F(x, y) = 0$, then $y = g(x)$. Moreover,*

$$Dg(x) = -D_2F(x, y)^{-1}D_1F(x, y).$$

End of lectures 11 and 12. May 18 and May 21, 2015

3 Integration in $\mathbb{R}^d$

3.1 Integrals depending on parameters

We well now approach the question of evaluating integrals of functions that depend on parameters. We will see that a particularly relevant property of the domain of definition of a function that intervenes in studying parameter-dependent expressions is compactness. When working on a compact domain it is possible to deduce some sort of uniformity of continuity estimates on the whole domain. This is necessary, for example, to be able to pass to the limit in expressions involving values of a function on the full domain like the integral.
Let us recall that a subset of $\mathbb{R}^d$ is compact if and only if it is bounded and closed via Heine-Borel's Theorem. Let $K \subset \mathbb{R}^d$ be a compact set, we define $C(K) = \{f : K \to \mathbb{R} \mid \text{f is a continuous function defined on } K\}$. Continuity is as usual expressed as $\forall x \in K\, \forall \epsilon > 0\, \exists \delta > 0$ such that for any $y \in K$ with $\|x - y\| < \delta$ one has $|f(x) - f(y)| < \epsilon$. A uniform notion of continuity is obtained if we require that the choice of δ be uniform, or independent, of the point $x \in K$ where we are studying continuity. Thus a function f is uniformly continuous if $\forall \epsilon > 0\, \exists \delta > 0$ such that $\forall x \in K$ and $\forall y \in K$ with $\|x - y\| < \delta$ one has $|f(x) - f(y)| < \epsilon$.

Theorem 3.1. *A continuous function $f \in C(K)$ on a compact domain K is uniformly continuous.*

Proof. We reason by contradiction. Suppose that the statement is false, then there exists an $\epsilon > 0$ such that for any choice of $\delta > 0$ we can find a pair of points x, $y \in K$ such that $\|x-y\| < \delta$ but $|f(x)-f(y)| \geq \epsilon$. Let us fix such an

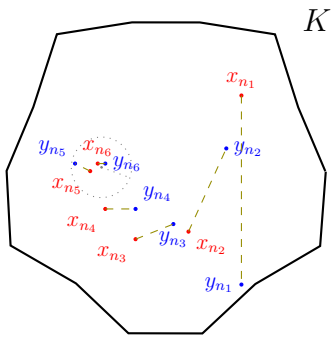


Figure 10: A compact domain K and jumps of a continuous function.

ϵ and for $n \in \mathbb{N}$ let us construct sequences x_n, y_n given by the contradiction statement with $\delta_n = \frac{1}{n}$ so that $\|x_n - y_n\| < \frac{1}{n}$ and $\|f(x_n) - f(y_n)\| \geq \epsilon$.
Since K is compact we can select a subsequence x_{n_k} so that it converges to a point $x \in K$. We will now show that the continuity of f contradicts the assumption on x_n and y_n close to this point. Continuity of f in x allows us to select $\delta > 0$ so that for all $z \in K$ we have $\|x - z\| < \delta$ implies $\|f(x) - f(z)\| < \frac{\epsilon}{3}$. Also choose $k > 0$ sufficiently large so that $\|x - x_{n_k}\| < \frac{\delta}{3}$ and such that $\frac{1}{n_k} < \frac{\delta}{3}$ so that $\|x - y_{n_k}\| \leq \|x - x_{n_k}\| + \|x_{n_k} - y_{n_k}\| < \frac{2\delta}{3}$. We get a contradiction from

$$\underbrace{\epsilon \leq |f(y_{n_k}) - f(x_{n_k})|}_{\text{contradiction hypothesis}} \leq \overbrace{|f(y_{n_k}) - f(x)| + |f(x) - f(x_{n_k})| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3}}^{\text{continuity of } f \text{ in } x}.$$

□

Notice that $C(K)$ is naturally a normed (and thus metric) space with the norm given by the supremum norm $\|f\|_\infty = \sup_{x \in K} |f(x)|$. The distance is thus as usual given by $(f, g) \mapsto \|f - g\|_\infty$.

Lemma 3.2. *Let $K \subset \mathbb{R}^d$ be a compact set and $a, b \in \mathbb{R}$ with $a < b$ then $[a, b] \times K \subset \mathbb{R}^{d+1}$ is also compact. Recall that $[a, b] \times K = \{(x, y) \,|\, x \in [a, b], y \in K\}$.*
Furthermore let $f : [a, b] \times K \to \mathbb{R}$ be a continuous function. Let us defined $F : K \to C([a, b])$ given by $F(y)(\cdot) = f(\cdot, y)$. F is a continuous map from K to $C([a, b])$ endowed with $\|\cdot\|_\infty$ as a normed vector space i.e. $\forall y \in K$ and $\forall \epsilon > 0$ there exists a $\delta > 0$ such that if $y' \in K$ with $\|y' - y\| < \delta$ then $\|F(y')(\cdot) - F(y)(\cdot)\|_\infty < \epsilon$.

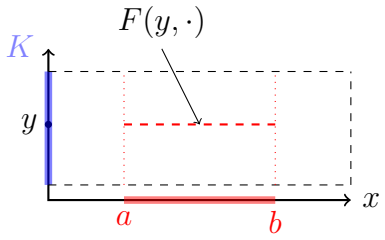


Figure 11: The function $F(y)(\cdot) = f(\cdot, y)$.

Proof. To begin with notice that $[a,\, b] \times K \subset \mathbb{R}^{d+1}$ is both bounded and closed so the first part of the Lemma follows.
We now check that F is continuous. We write $M = [a,\, b] \times K$. Let $\epsilon > 0$ be given; since M is compact then the continuity of f implies uniform continuity: there exists a $\delta > 0$ so that $\forall (x, y)$, $(x', y') \in M$ we have that $\|(x, y) - (x', y')\| < \delta$ implies $|f(x, y) - f(x', y')| < \epsilon$. By setting $x = x'$ for any $x \in K$ and rewriting the above inequality we see that

$$|F(y)(x) - F(y')(x)| = |(F(y) - F(y'))\,(x)| < \epsilon \quad \text{so} \quad \|F(y) - F(y')\|_\infty < \epsilon.$$

This is precisely the statement of continuity of F as a map from K to $C([a, b])$ because we have obtained that for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any y, $y' \in K$ such that $\|y - y'\| < \delta$ one has $\|F(y) - F(y')\|_\infty < \epsilon$. □

Let us consider the example of a function on $\mathbb{R}^2$ given by $(x,\, y) \mapsto f(x, y) = e^{ixy}$ so that $F(y)(x) = f(x, y)$. We have that $(F(y) - F(y'))\,(x) = e^{ixy} - e^{ixy'} = e^{ixy}(1 - e^{ix(y'-y)})$. But if x is an arbitrary point of $\mathbb{R}$ we have that this quantity can be large even if $y - y'$ is small; for example take $x = \frac{\pi}{y'-y}$. In particular we have that $\|F(y) - F(y')\|_\infty = 2$ for any $y \neq y'$ and this implies that F is NOT continuous. However if x were to be allowed to assume values only in a bounded set then the above function would, as a matter of fact, be continuous.
Now let us consider integrals depending on parameters and integrals over in more than one variables. Let $f \in C([a, b])$ and set

$$If = \int_a^b f(x)dx = \lim_{n\to\infty} \frac{b-a}{n} \sum_{k=1}^{n} f\left(a + \frac{b-a}{n}k\right).$$

The map $I : (f) \mapsto If$ is continuous as a map $I : C([a, b]) \to \mathbb{R}$. It suffices

to see that

$$\left|\int_a^b f(x)dx - \int_a^b g(x)dx\right| = \left|\int_a^b (f(x) - g(x))\,dx\right|$$
$$\leq \int_a^b |f(x) - g(x)|\,dx \leq (b-a)\|f-g\|_\infty$$

Theorem 3.3. *Let $[a,b] \subset \mathbb{R}$ and let $K \subset \mathbb{R}^d$ be a compact set. Given a continuous function $f : [a,b] \times K \to \mathbb{R}$ let us define $\phi(y) = \int_a^b f(x,y)dx$ for every $y \in K$. We then have that $\phi : K \to \mathbb{R}$ is continuous.*

Proof. It suffices to see that setting $F(y)(\cdot) = f(\cdot, y)$ we have that $\phi = I \circ F$. Since $F : K \to C([a,b])$ and $I : C([a,b]) \to \mathbb{R}$ are both continuous, their composition ϕ is also continuous. □

Now let $K = [c,d]$ so that $\int_c^d \left(\int_a^b f(x,y)dx\right) dy = \int_c^d \phi(y)dy$.

Lemma 3.4. *Set $[a,b]$, $[c,d]$ two compact intervals and let $f : [a,b] \times [c,d] \to \mathbb{R}$ be a continuous function. Suppose also that $D_2 f$ exists and is continuous in all points of the domain. Let $y_0 \in [c,d]$. Consider the function $g : [a,b] \times [c,d] \to \mathbb{R}$ given by*

$$g(x,y) = \begin{cases} \dfrac{f(x,y) - f(x,y_0)}{y - y_0} & y \neq y_0 \\ D_2 f(x,y_0) & y = y_0 \end{cases}.$$

Then g is a continuous function on $[a,b] \times [c,d]$.

Proof. Choose any $(x,y) \in [a,b] \times [c,d]$; we must show that g is continuous in (x,y). The case in which $y \neq y_0$ is left as an exercise. Suppose $y = y_0$. Let $\epsilon > 0$ and choose $\delta > 0$ so that $\forall (x',y')$ such that $\|(x',y') - (x,y_0)\| < \delta$ one has $|D_2 f(x',y') - D_2 f(x,y_0)| < \epsilon$. For any such (x',y') choose $y'' \in (y,y')$ so that $g(x',y') = \frac{f(x',y') - f(x',y_0)}{y' - y_0} = D_2 f(x',y'')$ via the Lagrange Theorem in one dimension. We then have that $|g(x',y') - g(x,y_0)| = |D_2 f(x',y'') - D_2 f(x,y_0)| < \epsilon$ where the last bound comes from continuity of $D_2 f$ and the fact that $|y'' - y_0| < |y' - y_0|$. □

Theorem 3.5. *Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be a continuous function such that $D_2 f$ exists and is continuous. Consider $\phi : [c,d] \to \mathbb{R}$ to be $\phi(y) = \int_a^b f(x,y)dx$. Then ϕ is differentiable with $\phi'(y) = \int_a^b D_2 f(x,y)dx$.*

Proof. Let g as in Lemma 3.4 be continuous so that $\int_a^b g(x,y)dx$ is also continuous. This allows us to state that $\forall \epsilon > 0$ there exists a $\delta > 0$ such that for all points y' so that $|y_0 - y'| < \delta$ we have that $\left| \int_a^b g(x,y')dx - \int_a^b D_2 f(x,y_0)dx \right| < \epsilon$. Thus

$$\left| \frac{\int_a^b f(x,y')dx - \int_a^b f(x,y_0)dx}{y'-y} - \int_a^b D_2 f(x,y_0)dx \right| < \epsilon.$$

This means that $\lim_{y' \to y_0} \frac{\int_a^b f(x,y')dx - \int_a^b f(x,y_0)dx}{y'-y} = \int_a^b D_2 f(x,y_0)dx$ and thus the derivative of $\phi(y)$ exists and is given by $\phi'(y) = \int_a^b D_2 f(x,y)dx$. □

Theorem 3.6. *Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be a continuous function then we have that*

$$\int_c^d \left(\int_a^b f(x,y)dx \right) dy = \int_a^b \left(\int_c^d f(x,y)dy \right) dx.$$

End of lecture 13. June 01, 2015

Theorem (3.6) follows from the following result.

Theorem 3.7. *Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be continuous. Then*

$$\int_c^d \left(\int_a^b f(x,y)dx \right) dy$$
$$= \lim_{N \to \infty} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(b-a)(d-c)}{N^2} f\Big(a + \frac{j}{N}(b-a), c + \frac{i}{N}(d-c) \Big). \quad (18)$$

The expression on the right hand-side of (18) is a two-dimensional Riemann sum.

Proof. Denote the left hand-side of (18) by L and the double sum on the right hand-side by R. First we show that for every $\varepsilon > 0$ there is N_0 such that for all $N > N_0$:

$$L \leq R + \varepsilon.$$

Let $\varepsilon > 0$. Since f is uniformly continuous on $[a,b] \times [c,d]$, there exists $\delta > 0$ such that for all $(x,y), (x',y') \in [a,b] \times [c,d]$:

$$|(x,y) - (x',y')| < \delta \Rightarrow |f(x,y) - f(x',y')| < \frac{\varepsilon}{(b-a)(d-c)}.$$

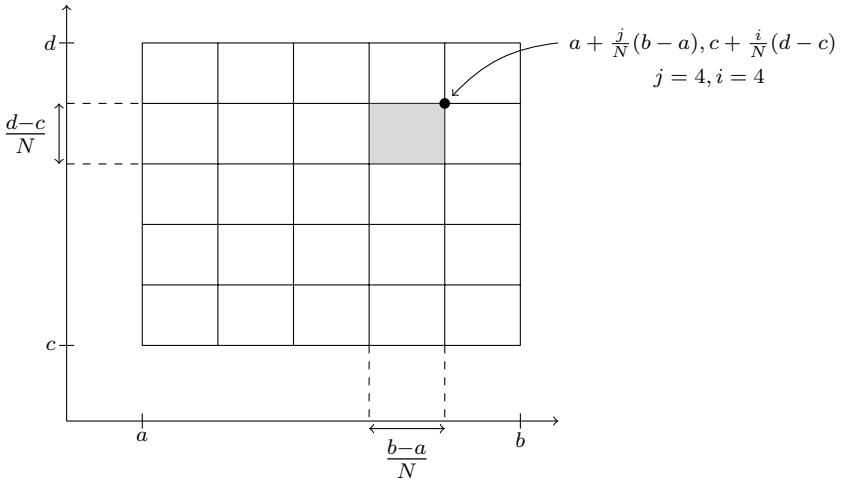


Figure 12: Discretization, $N = 5$.

Denote $N_0 := \frac{\sqrt{(b-a)(d-c)}}{\delta}$ and let $N > N_0$. Then for all $x \in [a + \frac{j-1}{N}(b-a), a + \frac{j}{N}(b-a)]$, $y \in [c + \frac{i-1}{N}(d-c), c + \frac{i}{N}(d-c)]$,

$$f(x,y) \le f\Big(a + \frac{j}{N}(b-a), c + \frac{i}{N}(d-c)\Big) + \frac{\varepsilon}{(b-a)(d-c)}. \tag{19}$$

By the definition of the integral ($\int_a^b f(x,y)dx$ as the infimum of upper sums of $f(\cdot, y)$ over all partitions of $[a,b]$) and by (19), for all $y \in [c + \frac{i-1}{N}(d-c), c + \frac{i}{N}(d-c)]$ we have

$$\int_a^b f(x,y)dx \le \sum_{j=1}^{N} f\Big(a + \frac{j}{N}(b-a), c + \frac{i}{N}(d-c)\Big) + \frac{\varepsilon}{d-c}.$$

The same argument in the y-direction gives

$$\int_c^d \Big(\int_a^b f(x,y)dx\Big)dy \le \sum_{i=1}^{N}\sum_{j=1}^{N} f\Big(a + \frac{j}{N}(b-a), c + \frac{i}{N}(d-c)\Big) + \varepsilon,$$

which establishes $L \le R + \varepsilon$. We leave it as an exercise to show the reverse inequality $L \ge R - \varepsilon$, which then finishes the proof. □

Note that we could have evaluated f at any point of the rectangle $[a + \frac{j-1}{N}(b-a), a + \frac{j}{N}(b-a)] \times [c + \frac{i-1}{N}(d-c), c + \frac{i}{N}(d-c)]$, not necessarily at its upper right corner.

The following theorem generalizes Theorem (3.6) to $\mathbb{R}^d$. Its proof is a straightforward generalization of the proof for $d = 2$.

Theorem 3.8. *Let $Q = [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$ and $f : Q \to \mathbb{R}$ continuous. We define*

$$\int_Q f(x)dx := \int_{a_1}^{b_1} \left(\dots \left(\int_{a_d}^{b_d} f(x_1, \dots, x_d)dx_d \right) \dots \right) dx_1$$

Then for any bijection (permutation) $\sigma : \{1, \dots, d\} \to \{1, \dots, d\}$,

$$\int_Q f(x)dx = \int_{a_{\sigma(1)}}^{b_{\sigma(1)}} \left(\dots \left(\int_{a_{\sigma(d)}}^{b_{\sigma(d)}} f(x_1, \dots, x_d)dx_{\sigma(d)} \right) \dots \right) dx_{\sigma(1)}.$$

3.2 Abstract characterization of the integral

First let us state some definitions.

The *support* of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\operatorname{supp}(f) := \overline{\{x : f(x) \neq 0\}}$$

That is, the closure of the set of all points in $\mathbb{R}^d$ where f is non-zero. The support is by definition closed.

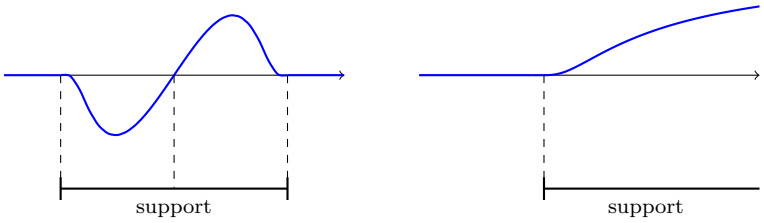


Figure 13: Compact and non-compact support, respectively.

We denote

$$C_c(\mathbb{R}^d) := \{f : \mathbb{R}^d \to \mathbb{R} \text{ continuous, } \operatorname{supp}(f) \text{ compact}\}$$

Note that if $f \in C_c(\mathbb{R}^d)$, then there exists a box Q such that $\operatorname{supp}(f) \subset Q$. We write [6]

$$\int_{\mathbb{R}^d} f(x)dx := \int_Q f(x)dx.$$

We leave it as an exercise to show that this definition does not depend on the choice of Q.

[6]Note that in the notation $\int_{\mathbb{R}^d} f(x)dx$ all the information is encoded in f.

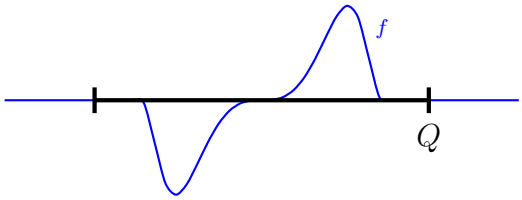


Figure 14: An interval Q containing $\mathrm{supp}(f)$.

Theorem 3.9. *The map $J : C_c(\mathbb{R}^d) \to \mathbb{R}$ defined by*

$$J(f) := \int_{\mathbb{R}^d} f(x)dx$$

is

1. *linear: $J(\lambda f + g) = \lambda J(f) + J(g)$ for all $f, g \in C_c(\mathbb{R}^d)$, $\lambda \in \mathbb{R}$*
2. *positive: $\forall x : f(x) \geq 0 \Rightarrow J(f) \geq 0$*
3. *translation invariant: $J(\tau_y f) = J(f)$ for all $f \in C_c(\mathbb{R}^d), y \in \mathbb{R}^d$, where $\tau_y : C_c(\mathbb{R}^d) \to C_c(\mathbb{R}^d)$ is defined by $\tau_y f(x) = f(x - y)$*

Note that τ_y translates the above mentioned boxes Q as well and that $\tau_y(y) = f(y-y) = f(0)$. Moreover, note that positivity and linearity imply $h \leq k \Rightarrow I(k) \geq I(h)$. To see this last fact, $k - h \geq 0 \Rightarrow I(k) - I(h) = I(k - h) \geq 0 \Rightarrow I(k) \geq I(h)$.

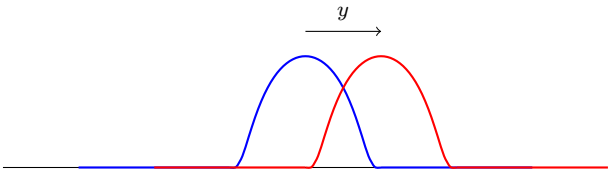


Figure 15: Translation for $y \in \mathbb{R}$.

Proof. 1.-3. follow from iterative applications of the same properties for the one-dimensional integral. In case of 3., this is $\int_a^b f(x)dx = \int_{a+y}^{b+y} f(x + y)dy$. We leave the details as an exercise. □

For a map satisfying 1.-3. we have the following uniqueness result, which gives an abstract (axiomatic) characterization of the integral.

Theorem 3.10. *If $I : C_c(\mathbb{R}^d) \to \mathbb{R}$ is any map satisfying the properties $1. - 3.$, there exists $c \geq 0$ such that $I = c\, J$.*

A map $I : C_c(\mathbb{R}^d) \to \mathbb{R}$ is called a *functional on* $C_c(\mathbb{R}^d)$. The statement of the last theorem can be rephrased such that any positive translation invariant linear functional on $C_c(\mathbb{R}^d)$ is a non-zero constant multiple of the integral. Before we start with the proof we need some preparation.

Lemma 3.11. *Let I be a positive linear functional on $C_c(\mathbb{R}^d)$. Let f_n be a sequence of functions whose support lies in a compact set K. Suppose that f_n converges uniformly (that is, in $\|\cdot\|_\infty$ norm) to a function f. Then $f \in C_c(\mathbb{R}^d)$ and $\lim_{n\to\infty} I(f_n) = I(f)$.*

Proof. Since f is the uniform limit of a sequence of continuous functions supported in K, f is continuous and its support lies in K.
It remains to see $\lim_{n\to\infty} I(f_n) = I(f)$. Choose a function $g \in C_c(\mathbb{R})$ such that $g|_K = 1$, $g \geq 0$. If $d = 1$ and $[a, b]$ we take g to be (see Figure 16)

$$h_{[a,b]}(x) := \begin{cases} 0 & x < a \\ x - (a-1) & a - 1 \leq x < a \\ 1 & x \in [a, b) \\ -x + (b+1) & b \leq x < b + 1 \\ 0 & b + 1 \leq x \end{cases}$$

for some interval $[a, b]$ containing K.

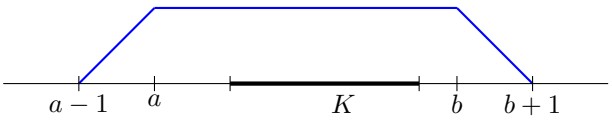


Figure 16: Function $h_{[a,b]}$ for some $[a, b] \supset K$.

If $d > 1$, we take $g(x) := h_{[a,b]}(x_1)h_{[a,b]}(x_2)\cdots h_{[a,b]}(x_d)$, for some $[a, b]$ such that $K \subset [a, b]^d$.
Let $\varepsilon > 0$. Choose N large enough such that for all $n > N$, $\|f_n - f\|_\infty < \varepsilon$. Since $g|_K = 1$, on K we have $|f_n - f| < \varepsilon = \varepsilon g$. On K^c we have $|f_n - f| = 0 < \varepsilon g$. Thus, on $\mathbb{R}^d$ we have

$$-\varepsilon g \leq f_n - f \leq \varepsilon g.$$

By the remark after Theorem 3.9, $-\varepsilon I(g) \leq I(f_n - f) \leq \varepsilon I(g)$. By linearity of I we obtain

$$-\varepsilon I(g) \leq I(f_n) - I(f) \leq \varepsilon I(g)$$

and hence

$$|I(f_n) - I(f)| \leq \varepsilon I(g).$$

Since $I(g)$ is independent of n, this finishes the proof. □

Lemma 3.12. *Let $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be continuous with compact support. Denote by $I_x F$ the action of I on the function $F(\cdot, y)$ (for a fixed y). Then $I_x F \in C_c(\mathbb{R}^d)$ and*

$$\int_{\mathbb{R}^d} I_x F(y)dy = I\left(\int_{\mathbb{R}^d} F(\cdot, y)dy\right).$$

This lemma states that we can "interchange" I and $\int$. If $I = \int$, we obtain Theorem 3.6.

Proof. The idea of the proof is to approximate $\int$ with Riemann sums and use linearity of I to interchange it with $\sum$. We use uniform continuity of F and the previous lemma. Assuming the support of F is contained in the box $[-a, a]^{2d}$, the Riemann sums one considers are

$$R_N F(x) = \sum_{j_1=1}^{N} \cdots \sum_{j_d=1}^{N} F\Big(x, -a + \frac{j_1}{N}2a, \ldots, -a + \frac{j_d}{N}2a\Big)\Big(\frac{2a}{N}\Big)^d.$$

The details are left as an exercise. □

Definition 3.13. The *convolution* $f * g$ of functions $f, g \in C_c(\mathbb{R}^d)$ is the function

$$f * g(x) := \int_{\mathbb{R}^d} f(y)g(x - y)dy$$

We have $f * g = g * f$, i.e. the convolution product is commutative. This fact follows from the change of variables

$$\begin{aligned} f * g(x) &= \int_{\mathbb{R}^d} f(y)g(x - y)dy \\ &= \int_{\mathbb{R}^d} f(x + y)g(-y)dy = \int_{\mathbb{R}^d} f(x - y)g(y)dy = g * f(x). \end{aligned}$$

Proof of Theorem 3.10. Let $f \in C_c(\mathbb{R}^d)$. Let $g \in C_c(\mathbb{R}^d)$ such that $\int g > 0$. Then

$$\begin{aligned} I(g)\int_{\mathbb{R}^d} f(y)dy &= \int_{\mathbb{R}^d} I(g)f(y)dy \\ &\overset{(1)}{=} \int_{\mathbb{R}^d} I(\tau_y g)f(y)dy \\ &\overset{(2)}{=} I\left(\int_{\mathbb{R}^d} f(y)\tau_y g dy\right) \\ &\overset{(3)}{=} I(f * g) \\ &= I(g * f) \\ &\overset{(4)}{=} I(f)\int_{\mathbb{R}^d} g(y)dy. \end{aligned}$$

Explanation of steps: (1) translation invariance of I. (2) Lemma 3.12. Note that $\tau_y g$ is a function of x and y. (3) Note that the integral in the bracket equals $\int_{\mathbb{R}^d} f(y)g(x-y)$. (4) Repeating the steps backwards.
Therefore, we have obtained

$$I(f) = \underbrace{\frac{I(g)}{\int_{\mathbb{R}^d} g(x)dx}}_{=:c} \int_{R^d} f(y)dy.$$

□

End of lecture 14. June 8, 2015

3.3 Change of variables formula

For a function $f \in C_c(\mathbb{R}^d)$ we have defined the integral $I(f) = \int_{\mathbb{R}^d} f(x)dx$. The functional I has an abstract characterization. $I : C_c(\mathbb{R}^d) \to \mathbb{R}$ is

1. Linear: $I(f + cg) = I(f) + cI(g)$
2. Positive: if $f \geq 0$ then $I(f) \geq 0$
3. Translation invariant: $I(\tau_y f) = I(f)$ where $\tau_y f(x) = f(x - y)$

We have showed that if $J : C_c(\mathbb{R}^d) \to \mathbb{R}$ is linear, positive and translation invariant, then there exists a constant $c \in \mathbb{R}^+$ such that $J = cI$.

This characterization of the integral will allow us to determine the behavior of integrals with respect to changes of variables. We begin by studying the simplest case when the change of variables is linear i.e. it is given by a matrix.

Theorem 3.14. *Let $A \in \mathbb{R}^{d\times d}$ be an invertible matrix, then for all $f \in C_c(\mathbb{R}^d)$ we have that $\int_{\mathbb{R}^d} f(Ax)dx = |\det A|^{-1} \int_{\mathbb{R}^d} f(x)dx$.*

Proof. The proof consists of two parts. Initially we will prove that $\int_{\mathbb{R}^d} f(Ax)dx$ is up to a constant the integral of f. This is done using the axiomatic characterization of the integral. Subsequently, to determine the constant we will need several intermediate results.
Let us set $J(f) := \int_{\mathbb{R}^d} f(Ax)dx$. J is a functional that possesses the three defining properties of the integral: linearity, positivity, and translation invariance. As a matter of fact $(f + cg)(Ax) = f(Ax) + cg(Ax)$ so linearity of J follows from the linearity of the integral. Positivity is also conserved

under composition with the linear operator A: if $f \geq 0$ then $f \circ A \geq 0$. Finally, for translation invariance we have that $(\tau_y f) \circ A(x) = f(Ax - y) = f\left(A(x - A^{-1}y)\right) = \tau_{A^{-1}y}\left(f \circ A\right)(x)$ so $J(\tau_y f) = \int_{\mathbb{R}^d} \tau_{A^{-1}y}\left(f \circ A\right)(x)dx = \int_{\mathbb{R}^d} \tau_{A^{-1}y}\left(f \circ A\right)(x)dx = \int_{\mathbb{R}^d} f(Ax)dx = J(f)$. Here we used that A is invertible.
By the previous theorem we can state that $J(f) = C_A I(f)$ for some constant $C_A \geq 0$. We now need to determine C_A and show that $C_A = |\det A|^{-1}$ to conclude the proof. □

We first determine C_A for special classes of matrixes.

Definition 3.15. A matrix $O \in \mathbb{R}^{d\times d}$ is called orthogonal if $\forall x \in \mathbb{R}^d$ one has $\|Ox\| = \|x\|$.

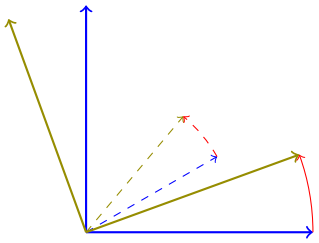


Figure 17: An orthogonal transformation.

Let us define $f(x) = \begin{cases} 1 - \|x\| & \text{if } \|x\| < 1 \\ 0 & \text{otherwise} \end{cases}$. If O is orthogonal we use that $\|Ox\| = \|x\|$ to obtain

$$\int_{\mathbb{R}^d} f(Ox)dx = \int_{\|Ox\|<1} \left(1 - \|Ox\|\right) dx = \int_{\|x\|<1} \left(1 - \|x\|\right) dx = \int_{\mathbb{R}^d} f(x)dx$$

So $C_O = 1$. Notice that the determinant of orthogonal matrix satisfies $\det O \in \{\pm 1\}$ since orthogonality is equivalent to $O^tO = \mathrm{I}$ and thus $1 = \det(O^tO) = \det(O^t)\det(O) = (\det O)^2$. So, as a matter of fact $C_O = 1 = |\det O|^{-1}$.
The second case we consider is that of lower triangular matrixes.

Definition 3.16. A matrix $A \in \mathbb{R}^{d\times d}$ is lower triangular if $a_{i,j} = 0$ for $j > i$. For a lower triangular matrix one has $(Ax)_i = \sum_{j=1}^{i} a_{i,j}x_j$.

$$A = \begin{pmatrix} a_{1,1} & & & & \\ a_{2,1} & a_{2,2} & & 0 & \\ \dots & \dots & \ddots & & \\ \dots & \dots & \dots & a_{d-1,d-1} & \\ \dots & \dots & \dots & a_{d,d-1} & a_{d,d} \end{pmatrix}$$

We now evaluate $J(f)$ for A a lower triangular matrix.

$$\int_{\mathbb{R}^d} f(Ax) = \int_{\mathbb{R}} \Big[\dots \Big[\int_{\mathbb{R}} f(a_{11}x_1, a_{2,1}x_1 + a_{2,2}x_2, \dots \\ \dots, a_{d,1}x_1 + \dots + a_{d,d}x_d)\, dx_d \Big] \dots \Big] dx_1.$$

Now notice that the innermost integral is a one-dimensional integral over the domain $\mathbb{R}$. Fixing $x_1 \dots x_{d-1}$ as parameters we can apply theorems on changes of variables in one dimension to get that

$$\int_{\mathbb{R}^d} f\left(Ax\right) = \int_{\mathbb{R}^d} f\left(A^{(d-1)}x^{(d-1)}, a_{d,d}x_d\right) dx = \\ \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} f\left(A^{(d-1)}x^{(d-1)}\right) dx_d dx^{(d-1)} = |a_{d,d}|^{-1} \int_{\mathbb{R}^{d-1}} \tilde{f}^{(d-1)}\left(A^{(d-1)}x^{(d-1)}\right)$$

where $x^{(d-1)} \in \mathbb{R}^{d-1}$ is the vector of the first $d-1$ coordinates of x so that $x^{(d-1)} = (x_1, \dots, x_{d-1})$, $A^{(d-1)} \in \mathbb{R}^{d-1,d-1}$ is the principal leading $(d-1) \times (d-1)$ minor of A i.e. $A^{(d-1)}_{i,j} = A_{i,j}$ for $i, j \in \{1, \dots, d-1\}$. We write $\tilde{f}^{(d-1)}(x^{(d-1)}) = \int_{\mathbb{R}} f(x^{(d-1)}, x_d) dx_d$ and the above equality holds because we are integrating over the whole $\mathbb{R}$ in x_d independently of the translation induced by $a_{d,1}x_1 + \dots + a_{d,d-1}x_{d-1}$. Applying this reasoning inductively we can show that $\int_{\mathbb{R}^d} f(Ax)dx = \prod_{i=1}^d |a_{j,j}|^{-1} \int_{\mathbb{R}^d} f(x)dx$. Since for lower triangular matrixes $\det A = \prod_{i=1}^d a_{i,i}$ we have effectively shown that for lower triangular matrixes A we have that $C_A = \prod_{j=1}^d |a_{j,j}|^{-1} = |\det A|^{-1}$.
We now want to describe any matrix in terms of these simpler matrixes. Notice that given any two invertible matrixes $A, B \in \mathbb{R}^{d\times d}$ the matrix AB is also invertible and we have that $C_{AB} = C_A C_B$ since setting $g(y) = f(Ay)$ we have

$$\underbrace{\int_{\mathbb{R}^d} f(ABx)dx}_{C_{AB}\int_{\mathbb{R}^d} f(x)dx} = \int_{\mathbb{R}^d} g(Bx)dx = C_B \int_{\mathbb{R}^d} g(x)dx = \\ C_B \int_{\mathbb{R}^d} f(Ax)dx = C_B C_A \int_{\mathbb{R}^d} f(x)dx$$

Theorem 3.17. *Any invertible matrix $A \in \mathbb{R}^{d\times d}$ can be represented as $A = OR$ with O an orthogonal matrix and R a lower triangular matrix.*

Proof. The proof is based on the Gram Schmidt algorithm. Let $a_1, \dots, a_d$ be

the column vectors of A i.e. $(a_j)_i = A_{i,j}$ so that

$$A = \left(\begin{bmatrix} \vdots \\ a_1 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \vdots \\ a_2 \\ \vdots \end{bmatrix} \quad \cdots \quad \begin{bmatrix} \vdots \\ a_d \\ \vdots \end{bmatrix} \right).$$

Since A is invertible, these vectors are linearly independent. We now apply the Gram-Schmidt formula starting from the last vector to the first one:

$$\begin{aligned} e_d &= \frac{a_d}{\|a_d\|} \\ e_{d-1} &= \frac{a_{d-1} - \langle a_{d-1}; e_d \rangle e_d}{\|a_{d-1} - \langle a_{d-1}; e_d \rangle e_d\|} \\ e_{d-2} &= \frac{a_{d-2} - \langle a_{d-2}; e_{d-1} \rangle e_{d-1} - \langle a_{d-2}; e_d \rangle e_d}{\|a_{d-2} - \dots\|} \\ &\vdots \\ e_1 &= \frac{a_1 - \dots}{\|a_1 - \dots\|} \end{aligned}$$

The system of vectors $(e_1, \dots, e_d)$ thus obtained is an orthonormal basis that together form a matrix

$$O = \left(\begin{bmatrix} \vdots \\ e_1 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \vdots \\ e_2 \\ \vdots \end{bmatrix} \quad \cdots \quad \begin{bmatrix} \vdots \\ e_d \\ \vdots \end{bmatrix} \right)$$

that is also orthonormal. We leave checking this as an exercise.
Now let us construct the matrix R that represents the Gram Schmidt algorithm. We set

$$r_d = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \|a_d\| \end{bmatrix} \qquad r_{d-1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \|a_{d-1} - \langle a_{d-1}; e_d \rangle e_d\| \\ \langle a_{d-1}; e_d \rangle \end{bmatrix}$$

and so on. The matrix

$$R = \left(\begin{bmatrix} \vdots \\ r_1 \\ \vdots \end{bmatrix} \quad \cdots \quad \begin{bmatrix} \vdots \\ r_{d-1} \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \vdots \\ r_d \\ \vdots \end{bmatrix} \right)$$

is clearly lower triangular and it is left as an exercise to see that $A = OR$. $\square$

The above decomposition property together with the property that $C_{AB} = C_A C_B$ and $\det(AB) = \det(A)\det B$ allows us to conclude from the fact that $C_A = |\det A|^{-1}$ for all orthogonal and lower triangular matrixes that $C_A = |\det A|^{-1}$ for any invertible matrix A. This concludes the proof of Theorem 3.14
Notice that the determinant function $\det : \mathbb{R}^{d\times d} \to \mathbb{R}$ is a polynomial in the entries of the matrix. In particular one has that

$$\det A = \sum_{\substack{\sigma \text{ permutation} \\ \text{of } \{1,\dots,d\}}} \operatorname{sign}(\sigma) \prod_{i=1}^{j} a_{i,\sigma(i)}$$

where $\operatorname{sign}(\sigma) \in \{\pm 1\}$ is the sign of the permutation σ, a combinatorial quantity. In particular this means that the determinant is a polynomial function on $\mathbb{R}^{d\times d}$ and thus it is C^∞. The function $A \mapsto |\det A|^{-1}$ is thus also smooth away from the set of matrixes A that are not invertible i.e. where $\det A = 0$. Notice that for a linear map $g(x) = Ax$ we have the property that it is equal to it's differential i.e. $Dg = DA = A$. It is thus possible to rewrite the formula for the change of variables as

$$\int_{\mathbb{R}^d} f \circ g(x)dx = \int_{\mathbb{R}^d} |\det(Dg(x))|^{-1} f(x)dx$$

as long as g is a linear map invertible map (i.e. it is given by an invertible matrix). The fact smoothness of the map $A \mapsto |\det A|^{-1}$ actually guarantees that even if the function g is non linear but just continuously differentiable then both sides of the above equality are well defined. We will show next time that the equality continues to hold extending the formula for changes of variable to non-linear maps g.

Definition 3.18. Let $U \subset \mathbb{R}^d$ be an open set in $\mathbb{R}^d$. We define

$C_c(U) = \{f \in C_c(\mathbb{R}^d) |\ \operatorname{supp} f \subset U\}$
$C_c^k(U) = \{f \in C_c(\mathbb{R}^d) |\ \operatorname{supp} f \subset U \text{ and } f \text{ is } k \text{ times continuously differentiable}\}$
$C_c^\infty(U) = \bigcap_{k=1}^{\infty} C_c^k(U) = \{f \in C_c(\mathbb{R}^d) |\ \operatorname{supp} f \subset U \text{ and } f \text{ is infinitely continuously differentiable}\}$

Theorem 3.19. *Let $f \in C_c(U)$, then there is a sequence $f_i \in C_c^\infty(U)$ so that $\limsup_{i\to\infty} \|f - f_i\|_\infty = 0$*

To show this result we must recall the definition and some properties of convolution:

$$f * g(x) = \int_{\mathbb{R}^d} f(y)g(x-y)dy \qquad \text{on } \mathbb{R}^d$$
$$f * g(x) = \sum_{y\in\mathbb{Z}} f(y)g(x-y) \qquad \text{on } \mathbb{Z}$$

On $\mathbb{Z}$ we have the important functions δ_n that are defined by

$$\delta_n(z) = \begin{cases} \delta_n(n) = 1 \\ \delta_n(z) = 0 \qquad z \neq n. \end{cases}$$

These functions are called Dirac deltas and have the property that they form a group under convolution: $\delta_n * \delta_m = \delta_{n+m}$. More in general we have that $f * \delta_n(z) = f(z-n)$ and thus if $n = 0$ we have that $f * \delta_0(z) = f(z)$ so δ_0 acts as the identity operator via convolution. The above holds, as noted, for functions on $\mathbb{Z}$. The problem is that on $\mathbb{R}$ and $\mathbb{R}^n$ there are no functions in the standard sense that play a role of Dirac deltas. The function δ_0 would have to be supported on $x = 0$ and be zero everywhere else. We thus chose to substitute these elements by Dirac sequences i.e. sequences that approximate Dirac deltas in an appropriate sense. Heuristically, a Dirac sequence will be a sequence of non-negative functions that are supported on small balls around 0 and all have integral 1.

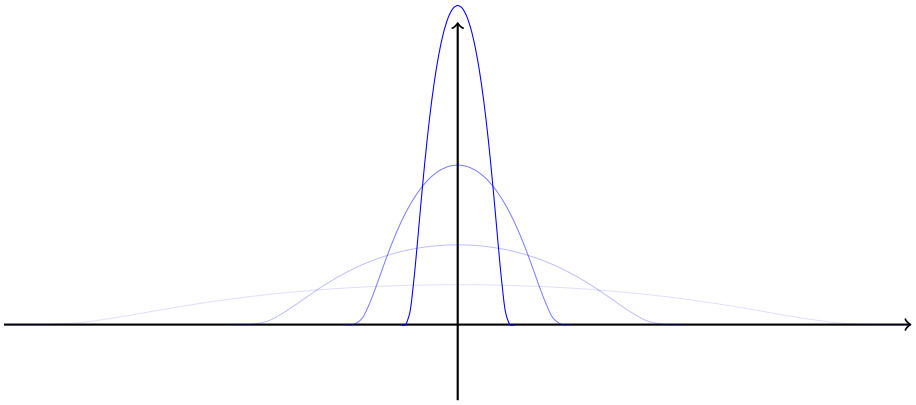


Figure 18: A Dirac sequence.

To construct a smooth Dirac sequence on $\mathbb{R}$ we can take a $C_c^\infty(B_1(0))$ function $\phi \geq 0$ with $\operatorname{supp}\phi \subset \{|x| < 1\}$ and with $\int_{\mathbb{R}} \phi = 1$ and by setting $\phi_\epsilon(x) = \epsilon^{-1}\phi\left(\frac{x}{\epsilon}\right)$. If we want to construct a smooth Dirac sequence on $\mathbb{R}^d$ it is sufficient to set $\psi_\epsilon(x) = \prod_{i=1}^d \phi_\epsilon(x_i) = \epsilon^{-d}\prod_{i=1}^d \phi\left(\frac{x_i}{\epsilon}\right)$.

We will show that given any function $f \in C_c(\mathbb{R}^d)$ we have that $\|f*\psi_\epsilon - f\|_\infty \to 0$ as $\epsilon \to 0$ and by properties of convolution each function $f * \psi_\epsilon \in C_c^\infty(\mathbb{R}^d)$. Before proceeding to that we will explicitly construct a smooth compactly supported function on $\mathbb{R}$. Our candidate is going to be the function $\eta : \mathbb{R} \to \mathbb{R}$ given by

$$\eta(x) = \begin{cases} 0 & x \leq 0 \\ e^{-\frac{1}{x}} & x > 0 \end{cases}$$

A special property is that for $x > 0$ we have that

$$\left(e^{-\frac{1}{x}}\right)' = \frac{1}{x^2} e^{-\frac{1}{x}} \qquad \left(e^{-\frac{1}{x}}\right)'' = \frac{1}{x^2} e^{-\frac{1}{x}}$$
$$\left(e^{-\frac{1}{x}}\right)^{(n)} = P_n\left(\frac{1}{x}\right) e^{-\frac{1}{x}}$$

where P_n is some polynomial. This can be prove by induction and is left as an exercise. Trivially, for $x < 0$ we have that $\eta(x) = 0$ and thus $\eta^{(n)} = 0$. The fact that $\eta^{(n)}(0) = 0$ and that $\eta \in C^\infty$ follows from the fact that for any $n \in \mathbb{N}$ we have that $\lim_{y \to +\infty} y^n e^{-y} = 0$. As a matter of fact using the series expansion of the exponential we have that $e^y = \sum_{k=0}^\infty \frac{1}{k!} y^k$ and in particular $e^y > \frac{1}{(n+1)!} y^{n+1}$ so $\lim_{y \to +\infty} \frac{e^y}{y^n} = \infty$ as required. The conclusion follows and is left as an exercise.
Finally setting $\rho(x) = \eta(x+1)\eta(1-x)$ we obtain that $\rho \in C_c^\infty(B_1(0))$ thus setting $\phi(x) = \rho(x) \left(\int_{\mathbb{R}} \rho(y) dy\right)^{-1}$ gives a function that is in $C_c^\infty(B_1(0))$ and with $\int_{\mathbb{R}} \phi(x) dx = 1$.

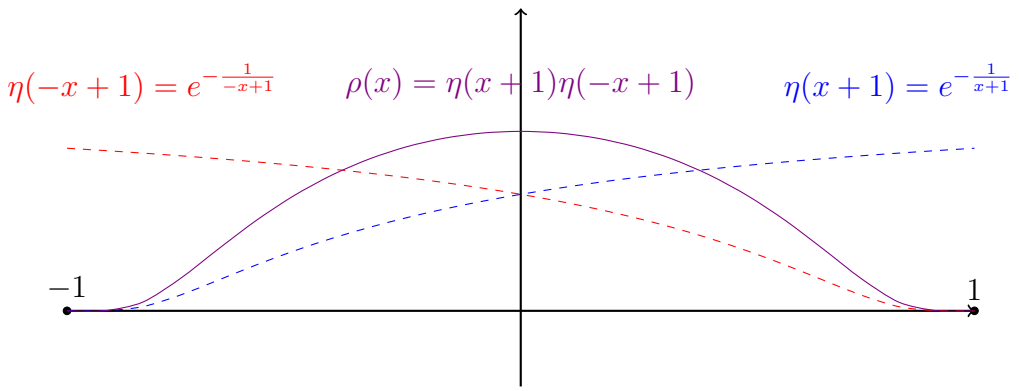


Figure 19: A $C_c^\infty(B_1(0))$ function.

End of lecture 15. June 15, 2015

We are set to establish a more general change of variables formula (transformation formula). The first step towards it is to prove Theorem 3.19, which says that $C_c^\infty(U)$ is dense in $C_c(U)$ in the supremum norm. Note that Theorem 3.19 could also be seen as a variant of the Stone-Weierstrass theorem from Analysis 1.

We remark that infinitely continuously differentiable functions (i.e. belonging to C^∞) are called *smooth*.

Before proceeding with the proof of Theorem 3.19 we need the following lemma.

Lemma 3.20. *Let $K := \mathrm{supp}(f)$. Then there is $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$ for all $x \in K$.*

Proof. The proof relies on a compactness argument. Since U is open, for all $x \in K$ there exists $\varepsilon_x > 0$ such that $B_{2\varepsilon_x}(x) \subset U$. We write ε_x to stress that it depends on the point x. What we need to show is that we can choose an epsilon, universal for all x, such that for each $x \in K$ we have $B_\varepsilon(x) \subset U$. Consider the set

$$\{B_{\varepsilon_x}(x) : x \in K\}.$$

This is an open covering of K, i.e. $K \subset \cup_{x\in K} B_{\varepsilon_x}(x)$. By compactness of K there exists a finite subcovering, i.e. there is $N \geq 0$ such that

$$K \subset \bigcup_{i=1}^{N} B_{\varepsilon_{x_i}}(x_i).$$

Set

$$\varepsilon := \min_{i=1,\ldots,N} \varepsilon_{x_i} > 0.$$

We claim that $\forall x \in K : B_\varepsilon(x) \subset U$. Let $x \in K$. Then $x \in B_{\varepsilon_{x_i}}(x_i)$ for some i. It suffices to show $B_\varepsilon(x) \subset B_{2\varepsilon_{x_i}}(x_i)$, which follows from the triangle inequality. Indeed, let $z \in B_\varepsilon(x)$, i.e. $\|z - x\| < \varepsilon$. Then $\|z - x_i\| \leq \|z - x\| + \|x - x_i\| \leq \varepsilon + \varepsilon_{x_i} \leq 2\varepsilon_{x_i}$ as desired. □

Observe that without compactness of K we could define ε only as the infimum of all ε_x. However, then possibly $\varepsilon = 0$.

Proof of Theorem 3.19. Choose a function $\varphi \in C_c^\infty(B_1(0))$ with $\varphi \geq 0$ and $\int_{\mathbb{R}^d} \varphi(x)dx = 1$. We have constructed such a function in the previous lecture. Define $\varphi_\varepsilon(x) := \varepsilon^{-d}\varphi(\varepsilon^{-1}x)$ and note $\mathrm{supp}(\varphi_\varepsilon) \subset B_\varepsilon(0)$ and $\int_{\mathbb{R}^d} \varphi_\varepsilon(x)dx = 1$. The last fact can be seen from the change of variables formula for constant

matrices, as the transformation matrix is in this case a diagonal matrix, with diagonal entries equal to ε^{-1}.

Claim: For ε as in Lemma 3.20, $\operatorname{supp}(f * \varphi_{\frac{\varepsilon}{2}}) \subset U$.
To see this we need to show that for all $x \in U^c$ and all $y \in B_{\frac{\varepsilon}{2}}(x)$: $f * \varphi_{\frac{\varepsilon}{2}} = 0$. Suppose the opposite, i.e.

$$\int_{\mathbb{R}^d} f(y-z)\varphi_{\frac{\varepsilon}{2}}(z)dz \neq 0.$$

Then there is z such that $f(y-z) \neq 0,\ \varphi_{\frac{\varepsilon}{2}}(z) \neq 0$. By the information on the support of $\varphi_{\frac{\varepsilon}{2}}$ we have $\|z\| < \frac{\varepsilon}{2}$. Since $y \in B_{\frac{\varepsilon}{2}}(x)$, this implies $x \in B_\varepsilon(y-z)$. By the support of f we have $y - z \in K$. By the previous lemma then $x \in U$, which is a contradiction.

The functions $f * \varphi_\varepsilon$ are smooth. Now we show that $f * \varphi_\varepsilon$ converge uniformly (i.e. in the ∞-norm) to f. Let $\delta > 0$. To show is that there is $\varepsilon > 0$ such that $\|f * \varphi_\varepsilon - f\|_\infty < \delta$. Since f is uniformly continuous, there is $\varepsilon > 0$ such that for all x, y: $\|x - y\| < \varepsilon$ implies $|f(x) - f(y)| < \delta$. If necessary make ε smaller such that by the above claim $\operatorname{supp}(f * \varphi_\varepsilon) \subset U$. We have

$$\begin{aligned}
|(f * \varphi_\varepsilon - f)(x)| &= \Big| \int_{\mathbb{R}^d} f(x-y)\varphi_\varepsilon(y)dy - f(x) \Big| \\
&\overset{(1)}{=} \Big| \int_{\mathbb{R}^d} (f(x-y) - f(x))\varphi_\varepsilon(y)dy \Big| \\
&\overset{(2)}{\leq} \int_{B_\varepsilon(0)} |f(x-y) - f(x)|\varphi_\varepsilon(y)dy \\
&\overset{(3)}{\leq} \int_{B_\varepsilon(0)} \delta\varphi_\varepsilon(y)dy = \delta.
\end{aligned}$$

Explanation: (1): $\int_{\mathbb{R}^d} \varphi_\varepsilon(y)dy = 1$. (2): triangle inequality, $\varphi \geq 0$ and $\operatorname{supp}(\varphi_\varepsilon) \subset B_\varepsilon(0)$. (3): Uniform continuity of f, as $\|x-y-x\| = \|y\| < \varepsilon$. $\square$

Now we are ready to state the transformation formula.

Theorem 3.21. *Let U, V be open in $\mathbb{R}^d$, $\eta : U \to V$ invertible and continuously differentiable and η^{-1} continuously differentiable. Let $f \in C_c(V)$. Then*

$$\int_U f(\eta(x))|\det D\eta(x)|dx = \int_V f(y)dy.$$

We remark that $\det D\eta$ is also called the *Jacobi determinant*. In case $d = 1$ this is simply η'.

Proof. Let $\varphi \in C_c^\infty(B_1(0))$ with $\varphi \geq 0$, $\int_{\mathbb{R}^d} \varphi(x)dx = 1$. Denote $\varphi_j(x) := j^d\varphi(jx)$. Consider

$$\begin{aligned}|\varphi_j(x) - \varphi_j(x')| &= j^d|\varphi(jx) - \varphi(jx')| \\ &\overset{(*)}{\leq} j^d\|D\varphi\|_\infty\|jx - jx'\| \\ &\leq j^{d+1}\|D\varphi\|_\infty\|x - x'\|,\end{aligned}$$

the inequality $(*)$ following from the mean value theorem. We write

$$\begin{aligned}\int_U (f * \varphi_j)(\eta(x))|\det D\eta(x)|dx &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(y)\varphi_j(\eta(x) - y)|\det D\eta(x)|dydx \\ &= \int_V f(y)h_j(y)dy,\end{aligned}$$

where $h_j(y) := \int_{\mathbb{R}^d} \varphi_j(\eta(x) - y)|\det D\eta(x)|dx$. As $j \to \infty$, $f * \varphi_j \to f$ uniformly, so for the left hand-side we have convergence

$$\int_U (f * \varphi_j)(\eta(x))|\det D\eta(x)|dx \longrightarrow \int_U f(\eta(x))|\det D\eta(x)|dx.$$

Thus it remains to show that $\|h_j - 1\|_{L^\infty(K)} \to 0$, where $K = \operatorname{supp}(f)$. Let $\varepsilon > 0$. We need to show that there is j such that $\|h_j - 1\|_{L^\infty(K)} < \varepsilon$. For $y \in V$ let z be such that $\eta(z) = y$. Choose δ small enough such that for $\|x - z\| < \delta$ we have

1. (by differentiability of η) $\|\eta(x) - \eta(z) - D\eta(z)(x - z)\| < \varepsilon\|x - z\|$
2. (by continuity of $\det D\eta$) $|\det(D\eta(x)) - \det(D\eta(z))| < \varepsilon$.

Choose C_0 large enough (depending on φ, η), such that for $j = C_0\delta^{-1}$, $\varphi_j(\eta(x) - \eta(z)) \neq 0$ or $\varphi_j(D\eta(z)(x - z)) \neq 0$ imply $\|x - z\| < \delta$. Now, for $y \in K$,

$$\begin{aligned}h_j(y) - 1 &= h_j(\eta(z)) - 1 \\ &= \int_{B_{\frac{1}{j}}(0)} \varphi_j(\eta(x) - \eta(z))\det D\eta(x)dx - 1\end{aligned}$$

We split the integral whether $\det D\eta(x) \geq 0$ or $\det D\eta(x) \geq 0$, both parts are treated similarly. So assume for simplicity $\det D\eta(x) \geq 0$ for all $x \in U$. Since $D\eta(z)$ is constant in x, by $1 = \int_{\mathbb{R}^d} \varphi(y)dy$ and by the transformation formula for linear maps the last display equals

$$\int_{B_{\frac{1}{j}}(0)} \varphi_j(\eta(x) - \eta(z))\det D\eta(x)dx - \int_{B_{\frac{1}{j}}(0)} \varphi_j(D\eta(z)(x - z))\det D\eta(z)dx.$$

The remaining part of the proof consists of using linearity of the integral, writing the difference of the involved products in the way $aB - bA = (a - A)B - (b - B)A$, applying the triangle inequality and using 1. and 2. The details are left as an exercise. □

Our (abstract) definition of the integral is valid only for functions $C_c(U)$. For now we extend the definition to the following two cases:

- (Non-vanishing at the boundary.) Let $K \subset U$, K compact, $U \subset \mathbb{R}^d$ open, $f \in C(U)$, $f \geq 0$. We define

$$\int_K f(x)dx := \inf_{\substack{h \in C_c(U) \\ h \geq \mathbf{1}_K}} \int_U f(x)h(x)dx,$$

  where $\mathbf{1}_K$ is the characteristic function of the set K.

- (Non-compact support.) Let $U \subset \mathbb{R}^d$ open, $f \in C(U)$, $f \geq 0$. Define

$$\int_U f(x)dx = \sup_{\substack{K \subset U \\ K \text{ compact}}} \int_K f(x)dx.$$

If we do not have $f \geq 0$, we split f into positive and negative parts and use these definitions on each of the parts separately.

End of lecture 16. June 18, 2015

4 Curves in $\mathbb{R}^n$ and path integrals

We have so far worked on functions in several variables and most of the theorems we have obtained are valid for general functions $f : \mathbb{R}^n \to \mathbb{R}^m$. A particularly important class of functions is given by paths or curves: functions defined on one-dimensional domains like $\gamma : \mathbb{R} \to \mathbb{R}^d$ or $\gamma : [a, b] \to \mathbb{R}$.

Definition 4.1. A curve (or path) is a map $\gamma : [a, b] \to \mathbb{R}^d$. We restrict our attention to curves that are at least continuous $C\left([a, b]; \mathbb{R}^d\right)$ but most of the time we will concentrate on continuously differentiable curves $C^1\left([a, b]; \mathbb{R}^d\right)$.

The derivative of a C^1 curve γ in a point $t \in [a, b]$ is given by the vector

$$\gamma'(t) = \begin{pmatrix} \gamma_1'(t) \\ \gamma_2'(t) \\ \vdots \\ \gamma_d'(t) \end{pmatrix}.$$

γ' is also a curve called the curve of tangent vectors since the vector $\gamma'(t)$ is tangent to the curve γ in the point $\gamma(t)$.

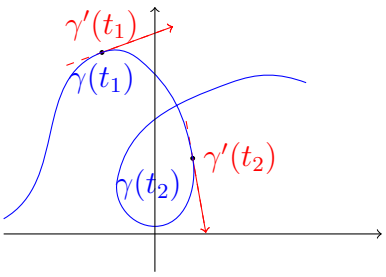


Figure 20: A curve γ and its tangent γ'.

Lemma 4.2. *Let $x_0, x_1, \ldots, x_n \in \mathbb{R}^d$ then there exists a C^1 curve $\gamma : [0, n] \to \mathbb{R}^d$ such that $\gamma(i) = x_i$ for $i \in \{0, \ldots, n\}$ and such that it is given by a linear segment on $[x_i, x_{i+1}]$ for $i \in \{0, \ldots, n-1\}$.*

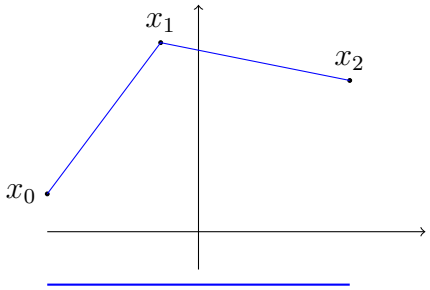


Figure 21: A piecewise linear C^1 curve

Proof. Define $\gamma(t) = x_i + \frac{1}{2}(x_{i+1} - x_i)(1 - cos((t-i)\pi))$ for $t \in [i, i+1]$, $i \in \{0, \ldots, n-1\}$. This functions is well defined at integer times with $\gamma(i) = x_i$ and it is C^1 since close to the right and to the left of the integer point $i \in \{0, \ldots, n-1\}$ the derivative γ' is continuous and $\gamma'(i) = 0$ as can be seen explicitly. □

Let $\gamma : [a, b] \to \mathbb{R}^d \in C^1$ be a continuously differentiable curve and let $f : \mathbb{R}^d \to \mathbb{R} \in C^1$ be a scalar-valued function, then $f \circ \gamma : [a, b] \to \mathbb{R}$ is a scalar function of one real variable and it is also C^1). In particular, the Fundamental Theorem of Calculus holds for this function and so we have

that

$$f \circ \gamma(b) - f \circ \gamma(a) = \int_a^b (f \circ \gamma)'(t)dt =$$
$$\int_a^b \sum_{i=1}^d \overbrace{D_i f(\gamma(t))}^{\nabla f(\gamma(t))} \gamma_i'(t)dt = \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)dt.$$

We call the vector

$$\nabla f(x) = \begin{pmatrix} D_1 f(x) \\ D_2 f(x) \\ \vdots \\ D_d f(x) \end{pmatrix}$$

the gradient vector of the scalar-valued function f. This is a functions defined on $\mathbb{R}^d$ (or an open subset of thereof) with values in $\mathbb{R}^d$, provided that f is sufficiently regular i.e. $f \in C^1\left(\mathbb{R}^d; \mathbb{R}\right)$. Similarly $\gamma'(t) \in \mathbb{R}^d$ so via a slight abuse of notation we write the product notation $\nabla f \cdot \gamma := \langle \nabla f;\, \gamma' \rangle_{\mathbb{R}^d}$ when we are actually dealing with the scalar product of the two vectors ∇f and γ' on $\mathbb{R}^d$ endowed with its structure of the Euclidean space.
Thus if $f \in C^1\left(\mathbb{R}^d; \mathbb{R}\right)$ then $\nabla f : \mathbb{R}^d \to \mathbb{R}^d$ and it is continuous. This consideration warrants the following definition.

Definition 4.3. A function $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$ with U some open set is called a vector field on U. Given a continuous vector field $F \in C\left(\Omega \subset \mathbb{R}^d; \mathbb{R}^d\right)$ and a C^1 path $\gamma : [a, b] \to \Omega \subset \mathbb{R}^d$ we define the integral of F along γ as

$$\int_\gamma F := \int_a^b F(\gamma(t)) \cdot \gamma'(t)dt.$$

An alternative notation is given by $\int_\gamma F \equiv \int F d\gamma$.

It is noteworthy that the definition of a path carries in itself not only the information on its support i.e. the set $\{x \in \mathbb{R}^d \,|\, \exists t \in [a, b] \text{ with } \gamma(t) = x\}$ but also the actual parameterization with the segment $[a, b]$. However two curves differ by a reparameterization and thus in some sense describe the same support have many important properties. The way we approach studying the properties of curves is via their action, or pairing via the integral, on vector fields.
Let $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$ be a C-vector field and let $\gamma : [a, b] \to U$ be a C^1-curve. Let $\phi : [c, d] \to [a, b]$ be an invertible C^1 function with C^1 inverse $\phi^{-1} : [a, b] \to [c, d]$ that we call a reparameterization. We also require

that ϕ be orientation conserving i.e. (monotone) increasing. Given a reparameterization one can associate to it a new curve $\tilde{\gamma} : [c, d] \to U$ given by $\tilde{\gamma}(s) = (\gamma \circ \phi)(s) = \gamma(\phi(s))$. It is clear that the two curves are closely related. In particular we have the following property:

$$\int_a^b F(\gamma(t))\gamma'(t)dt = \int_c^d F(\tilde{\gamma}(s))\tilde{\gamma}'(s)ds$$

That can be written concisely as

$$\int_\gamma F = \int_{\gamma\circ\phi} F = \int_{\tilde{\gamma}} F$$

Notice that this statement holds for any vector field. While the statement is trivial for vector fields $F = \nabla f$ that are gradients of C^1 functions since the above quantities depend only on the beginning and end points of γ and $\tilde{\gamma}$, the invariance with respect to reparameterization is true for general F.
The proof of this fact is based on the properties of change of variables in one dimension. As a matter of fact applying the chain rule to differentiation and the changing variable $t = \phi(s)$ yields

$$\int_c^d F(\tilde{\gamma}(t)) \cdot \tilde{\gamma}'(t)dt = \int_c^d F(\gamma(\phi(s))) \cdot \gamma'(\phi(s))\phi'(s)ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t)dt$$

as required.

Exercise 4.4. *The monotonicity of a reparameterization map ϕ follows directly from the condition on invertibility, however without the condition on ϕ being increasing a change of sign (or orientation) may happen. Consider $\phi : [0, 1] \to [0, 1]$ given by $\phi(s) = 1 - s$ and let $\tilde{\gamma}(s) = \gamma \circ \phi(s) = \gamma(1 - s)$. One can show that in this case $\int_{\tilde{\gamma}} F = -\int_\gamma F$.*

Definition 4.5. A vector field $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$ is called conservative if there exists $f : U \to \mathbb{R} \in C^1$ such that $F = \nabla f$.

Theorem 4.6. *Let $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$ be a conservative vector field and $\gamma, \tilde{\gamma} : [a, b] \to U$ two C^1 curves with the same start and end points: $\gamma(a) = \tilde{\gamma}(a)$, $\gamma(b) = \tilde{\gamma}(b)$. Then*

$$\int_\gamma F = \int_a^b F(\gamma(t)) \cdot \gamma'(t)dt = \int_a^b F(\tilde{\gamma}(t)) \cdot \tilde{\gamma}'(t)dt = \int_{\tilde{\gamma}} F$$

The proof of this statement has been given at the beginning of this lesson and relies on the Fundamental Theorem of Calculus in one dimension. An

example of this can be imagined by considering a $f : \mathbb{R}^2 \to \mathbb{R}$ to be the function that assigns the altitude of the terrain of some map. The gradient field $F = \nabla f$ will thus be a vector that indicates the direction of steepest climb and its norm characterizes the "steepness". A curve $\gamma : [a, b] \to \mathbb{R}^2$ can be thought of as a time-dependent parameterization of a path taken so that $\gamma'(t)$ is the speed at time t and $F(\gamma(t)) \cdot \gamma'(t)$ becomes the rate of climb (up the hill/mountain). At this point the integral

$$\int_\gamma F = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt$$

represent nothing else than the gain in altitude from time a to time b. It is clear that this gain depends only on the starting point $\gamma(a)$ and the end point of the journey $\gamma(b)$ and doesn't actually depend on the path that has been undertaken. Notice that in general the definition of a conservative vector field is non-trivial: there exist non-conservative vector fields.
We now pass to some properties of open sets, also called open domains, that are relevant to the properties of vector fields and paths.

Definition 4.7. An open set $U \subset \mathbb{R}^d$ is said to be path-wise connected if for any two points $x_0, x_1 \in U$ there is a continuous curve $\gamma : [a, b] \to U$ such that $\gamma(a) = x_0$, $\gamma(b) = x_1$.

Definition 4.8. An open set $U \subset \mathbb{R}^d$ is said to be connected if given a splitting $U = V \cup W$ with V and W disjoint open sets i.e. with $V \cap W = \emptyset$, then either $V = \emptyset$ or $W = \emptyset$.

Theorem 4.9. *An open set $U \subset \mathbb{R}^d$ is connected if and only if it is path-wise connected.*

Proof.

$\Leftarrow$ Let us reason by contradiction: let U be path-wise connected and let $U = V \cap W$ with V, W non-empty open sets. Choose $x_0 \in V$, $x_1 \in W$ and let $\gamma : [a, b] \to U$ be given by assumption of path-wise connectedness. Let $t = \sup\{s \in [a, b] \,|\, \gamma(s) \in V\}$. Now we distinguish two cases: either $\gamma(t) \in V$ or $\gamma(t) \in W$, but we will see that both of these situations lead to a contradiction. If $\gamma(t) \in V$ then $t < b$ because $\gamma(b) = x_1 \in W$ and, since V is open, there exists an $\epsilon > 0$ such that $B_\epsilon(\gamma(t)) \subset V$. Since γ is continuous there exists a $\delta > 0$ such that for all $t \leq s \leq t + \delta < b$ we have that $\|\gamma(s) - \gamma(t)\| < \epsilon$ so $\gamma(s) \in V$ and this contradicts the maximality of the choice of t.

If $\gamma(t) \in W$ we reach a contradiction by a similar argument. Clearly $t > a$ since $\gamma(a) \in V$. Since W is open there exists an $\epsilon > 0$ such that $B_\epsilon(\gamma(t)) \subset W$ and by continuity of γ in t we have that there exists a δ such that for all $t - \delta < s < t$ one has $\|\gamma(s) - \gamma(t)\| < \epsilon$ so $\gamma(s) \in W$ for all $t - \delta < s$ and this contradicts that t is the least upper bound.

$\Rightarrow$ Let U be connected non-empty set and let $x_0 \in U$. Define the set V to be the set of points of U reachable from x_0 via a continuous path contained in U: $V = \{x_1 \in U \mid \exists \gamma : [a,b] \to U \text{ with } \gamma(a) = x_0,\ \gamma(b) = x_1\}$. The set V is open: as a matter of fact take $x_1 \in V$; since V U is open there exists a ball $B_\epsilon(x_1)$ of $\mathbb{R}^d$ for some $\epsilon > 0$ such that $B_\epsilon(x_1) \subset U$. for any point $y \in B_\epsilon(x_1)$ one can construct a continuous path $\tilde{\gamma}$ that connects x and y. Suppose that $\gamma : [a,b] \to U$ connects x_0 to x_1 and setting $\tilde{\gamma} : [a, b+1] \to U$ with $\tilde{\gamma}(t) = \gamma(t)$ for $t \in [a,b]$ and $\tilde{\gamma}(t) = (y - x_1)(t - b)$ for $t \in [b, b+1]$. $\tilde{\gamma}$ thus defined is continuous.

However the set $W = U \setminus V$ is also open for a similar reason. By contradiction let x_1 not be reachable by any curve γ starting from x_0 and consider a ball of $\mathbb{R}^d$ such that $B_\epsilon(x_1) \subset U$. If any point of $y \in B_\epsilon(x_1)$ were reachable from x_0 we would have a contradiction similarly to before.

So, since U is connected and V and W are both open and disjoint then $V = \emptyset$ or $W = \emptyset$ but $x_0 \in V$ so $W = \emptyset$ and thus $V = U$ and this is exactly what was required.

□

End of lecture 17. June 2, 2015

We have defined the integral of a vector field $F : \mathbb{R}^d \to \mathbb{R}^d$ along a curve $\gamma : [a,b] \to \mathbb{R}^d$ by

$$\int_\gamma F := \int_a^b F(\gamma(t))\gamma'(t)dt.$$

We mention that one can also write it in the form

$$\begin{aligned}\int_\gamma F &= \int_a^b \sum_{j=1}^d F_j(\gamma(t))\gamma_j'(t)dt \\ &= \int_\gamma \sum_{j=1}^d F_j(x)dx_j \\ &= \int_\gamma F_1(x)dx_1 + F_2(x)dx_2 + \cdots + F_d(x)dx_d, \end{aligned} \tag{20}$$

where the second equality can be seen by substituting $\gamma(t) = x$. The expression (20) is also called *1-form*.

Remark. The *trace* of a curve $\gamma : [a, b] \to \mathbb{R}^d$ is defined by

$$\mathrm{tr}(\gamma) := \{\gamma(t) : t \in [a, b]\}$$

Assume that now $\gamma \in C^1$ and divide the interval $[a, b]$ into $a := a_0 < a_1 < \cdots < a_n := b$. Assume that $\mathrm{tr}(\gamma|_{[a_k,a_{k+1}]})$ is a line segment from x_k to x_{k+1} for $k = 0, \ldots, n-1$.[7] Then we have

$$\begin{aligned}\int_\gamma F &= \sum_{k=0}^{n-1} \int_{a_k}^{a_{k+1}} F(\gamma(t)) \cdot \gamma'(t)dt \\ &= \sum_{k=0}^{n-1} \int_0^1 F(x_k + s(x_{k+1} - x_k)) \cdot (x_{k+1} - x_k)ds \end{aligned} \tag{21}$$

The parametrization in (21) is called the *natural parametrization*. Each line segment is parametrized by its length.

To see (21) we may assume $x_k = 0$ and $x_{k+1} = e_1$. Then $\mathrm{tr}(\gamma|_{[a_k,a_{k+1}]}) = \{\alpha e_1 : \alpha \in [0, 1]\}$ and $\gamma(t) \cdot e_j = 0$ for $j = 2, \ldots, d$. Define $s = \gamma(t) \cdot e_1$. Then

$$\begin{aligned}&\int_{a_k}^{a_{k+1}} F(\gamma(t)) \cdot \gamma'(t)dt \\ =&\int_{a_k}^{a_{k+1}} F(\gamma(t)) \cdot e_1(\gamma'(t) \cdot e_1)dt \\ =&\int_0^1 F(se_1) \cdot e_1 ds\end{aligned}$$

[7] In the previous lecture we saw that there exists a C^1 parametrisation of a piecewise linear curve.

Given a natural parametrisation of a piecewise linear curve, we can thus define a path integral along that curve by defining it on each $[a_k, a_{k+1}]$ separately and then sum the pieces together.

The statement of the following theorem includes Theorem 4.6 from the previous lecture and its converse.

Theorem 4.10. *Let $U \subset \mathbb{R}^d$ be open and connected. Let $F : U \to \mathbb{R}^d$ be a continuous vector field. The following are equivalent:*

1. *F is conservative, that is, there exists $f : U \to \mathbb{R} \in C^1$ such that $F = \nabla f$ (i.e. $F_i = D_i f$ for $i = 1, \dots, d$.)*
2. *If $\gamma : [a, b] \to U$, $\widetilde{\gamma} : [a, b] \to U$ are C^1 with $\gamma(a) = \widetilde{\gamma}(a)$ and $\gamma(b) = \widetilde{\gamma}(b)$, then*
$$\int_\gamma F = \int_{\widetilde{\gamma}} F.$$
3. *If γ is closed ($\gamma(a) = \gamma(b)$), then $\int_\gamma F = 0$.*

In 1., the function f is called the *potential* of F.

Proof. The implication 1. $\Rightarrow$ 2. has already been observed in Lecture 17. We briefly repeat the argument, which relies on the fundamental theorem of calculus: for $F = \nabla f$ we have

$$\begin{aligned}\int_\gamma F &= \int_a^b \nabla f(\gamma(t))\gamma'(t)dt = \int_a^b (f \circ \gamma)'(t)dt \\ &= f \circ \gamma(b) - f \circ \gamma(a) = f \circ \widetilde{\gamma}(b) - f \circ \widetilde{\gamma}(a) \\ &= \cdots = \int_a^b \nabla f(\widetilde{\gamma}(t))\widetilde{\gamma}'(t)dt = \int_{\widetilde{\gamma}} F.\end{aligned}$$

We now prove 2. $\Rightarrow$ 1. Choose $x_0 \in U$. Since U is connected and therefore path-wise connected, for $x_1 \in U$ there is a continuous curve $\eta : [a, b] \to U$ with $\eta(a) = x_0$, $\eta(b) = x_1$. We would like to set $f(x_1) := \int_\eta F$. The problem is that η may not be differentiable, which is required to define $\int_\eta F$.
We shall circumvent this problem by finding a piecewise linear curve between x_0 and x_1, which lies in U. By Lemma 4.2 we may then pick a C^1 parametrisation of the curve.

Existence of the desired curve between x_0 and x_1 can be seen by the following compactness argument. For simplicity of notation suppose that $[a, b] = [0, 1]$.

Since U is open, for each $t \in [0,1]$ there is ε_t such that $B_{2\varepsilon_t}(\eta(t)) \subset U$. Balls with half the radii cover the trace of η, i.e. $\mathrm{tr}(\eta) \subset \cup_{t\in[0,1]} B_{\varepsilon_t}(\eta(t))$. The trace of η is compact as it is the image of a compact set under a continuous map[8]. Thus, there exist $t_0 := 0 < t_1 < \cdots < t_n := 1$ such that

$$\mathrm{tr}(\eta) \subset \cup_{j=0}^{n} B_{\varepsilon_{t_j}}(\eta(t_j)).$$

By uniform continuity of η there is $N \in \mathbb{N}$ such that for all $k = 0, \ldots, N$,

$$\|\eta(\frac{k+1}{N}) - \eta(\frac{k}{N})\| \leq \min_j \varepsilon_{t_j}.$$

Since $\eta(\frac{k}{N}) \in B_{\varepsilon_{t_j}}(\eta(t_j))$ for some t_j, this implies that the line segment between $\eta(\frac{k}{N})$ and $\eta(\frac{k+1}{N})$ lies in U.

By γ denote (a C^1 parametrization of) the constructed piecewise linear curve between x_0 and x_1. Define $f(x_1) := \int_\gamma F$. Now we show that f is differentiable and $D_j f(x_1) = F_j(x_1)$. Denote by $\widetilde{\gamma}$ the piecewise linear curve between x_0 and $x_1 + se_j$, $s > 0$. We do not know how to compare γ and $\widetilde{\gamma}$. For that we choose another curve $\widetilde{\widetilde{\gamma}}$ between x_0 and $x_1 + se_j$ such that $\mathrm{tr}(\widetilde{\widetilde{\gamma}})$ agrees with $\mathrm{tr}(\gamma)$ between x_0 and x_1. Consider now

$$f(x_1 + se_j) - f(x_1) = \int_{\widetilde{\gamma}} F - \int_\gamma F = \int_{\widetilde{\widetilde{\gamma}}} F - \int_\gamma F,$$

where the last equality follows by the path independence (assumption 2.). Using (21), all terms but one vanish and we obtain

$$\int_0^s F(x_1 + te_j) \cdot e_j dt = \int_0^s F_j(x_1 + te_j) dt.$$

The last display equals

$$F_j(x_1)s + \int_0^s (F_j(x_1 + te_j) - F(x_1)) dt.$$

By continuity of F_j, for every $\varepsilon > 0$ there is $\delta > 0$ such that for $s < \delta$,

$$\Big| \int_0^s (F_j(x_1 + te_j) - F(x_1)) dt \Big| \leq s\varepsilon.$$

Thus,

$$\frac{f(x_1 + se_j) - f(x_1)}{s} = F_j(x_1) + \text{error}$$

[8]We leave the proof of this fact to the reader.

where $|\text{error}| < \varepsilon$. Altogether we obtain

$$\lim_{s\to 0} \frac{f(x_1 + se_j - f(x_1))}{s} = F_j(x_1).$$

This finishes the proof of 2. $\Rightarrow$ 1.
The equivalence 2. $\Leftrightarrow$ 3. is left to the reader. $\square$

We observe that if $F \in C^1$ is conservative, for all $i, j \in \{1, \ldots, d\}$ we have

$$D_i F_j = D_j F_i.$$

This follows from the theorem of Schwartz. Indeed, let $F = \nabla f$. Then $F_i = D_i f$ and $D_i D_j f = D_j D_i f$.

Definition 4.11. A C^1 vector field F is called *irrotational*, if $D_i F_j = D_j F_i$ for all $i, j \in \{1, \ldots, d\}$.

Now we can summarize the preceding discussion as follows.

Theorem 4.12. *If a vector field $F \in C^1$ is conservative, then it is irrotational.*

It is naturally to ask whether the converse of this theorem holds. That is, is every irrotational C^1 vector field conservative? As we proceed we shall see that there is a partial converse. However, in general, the answer is negative. This can be seen by the following example.

Example. Let $U = \mathbb{R}^2 \setminus \{(0,0)\}$ and $F : U \to \mathbb{R}^2$ given by

$$F(x, y) = \Big(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2}\Big)$$

We compute

$$(D_1 F_2 - D_2 F_1)(x, y) = \frac{x^2 + y^2 - 2x^2}{(x^2 + y^2)^2} + \frac{x^2 + y^2 - 2y^2}{(x^2 + y^2)^2} = 0,$$

so F is irrotational.[9]

[9]The quantity $D_1 F_2 - D_2 F_1$ is called the *curl* of F, denoted curl(F). An irrotational vector field is also called *curl free*.

But F is not conservative. This can be seen by considering the curve $\gamma : [0, 2\pi] \to U$, $\gamma(t) = (\cos t, \sin t)$. Then

$$\begin{aligned}\int_\gamma F &= \int_\gamma F_1 dx + F_2 dy \\ &= \int_0^{2\pi} \frac{-\sin t}{\cos^2 t + \sin^2 t}(-\sin t) + \frac{\cos t}{\cos^2 t + \sin^2 t}(\cos t)dt \\ &= \int_0^{2\pi} 1 = 2\pi \neq 0.\end{aligned}$$

If F were conservative, this would contradict 3. from Theorem 4.10.
More generally, if $F = \nabla f$ and $\gamma : [0, s] \to \mathbb{R}^2$, then $\int_\gamma F = f(\gamma(s)) - f(\gamma(0))$.
Let γ be given by $\gamma(t) = (\cos t, \sin t)$. By the same calculation as above,

$$f(\cos s, \sin s) - f(1, 0) = \int_0^s F(\gamma(t)) \cdot \gamma'(t)dt = \int_0^s dt = s$$

Suppose $f(1, 0) = 0$. Then at a point $(\cos s, \sin s)$, the function f describes the angle of the vector of that point with the x-axis. In $\mathbb{R}^2 \setminus \{(0, 0)\}$, the angle cannot be defined continuously - starting at 0 and going around the circle once, the angle approaches 2π.
However, F is conservative if we change the domain U to exclude such paths which wind around the origin. For instance, if we restrict ourselves to the upper half plane

$$\{(x, y) : y > 0\}$$

or if we consider

$$\mathbb{R}^2 \setminus \{(x, 0) : x > 0\}.$$

We shall elaborate on this in the following lecture.
Recall that for $x > 0$ we have defined the polar angle of (x, y) as $\operatorname{atan}(\frac{y}{x})$ (see the example from Lecture 12). One can check that F is the gradient of $\operatorname{atan}(\frac{y}{x})$. In the upper half plane, this is a continuous function.

End of lecture 18. June 25, 2015

We proceed with the question when is an irrotational C^1 vector field conservative. We consider the following example.

Example. Let U be an open neighbourhood of the unit square $Q = [0, 1] \times [0, 1]$. Assume $F : U \to \mathbb{R}^2 \in C^1$ is irrotational. Let $\gamma : [0, 4] \to U$ be given

by

$$\gamma|_{[0,1]}(t) = (t, 0), \qquad \gamma|_{[1,2]}(t) = (1, t-1)$$
$$\gamma|_{[2,3]}(t) = (3-t, 1), \qquad \gamma|_{[3,4]}(t) = (0, 4-t)$$

This parametrisation is piecewise C^1, which suffices to define the integral $\sum_{j=1}^{4} \int_{j-1}^{j} F(\gamma(t)) \cdot \gamma'(t) dt$. (If one closely follows our definition of the path integral, one should first pick a C^1 parametrization $\widetilde{\gamma}$, which can be done as the path is piecewise linear. Then one considers $\int_{\widetilde{\gamma}} F$, restricts it to each of the segments and reparametrizes each segment by the restriction of γ to the appropriate interval. This is exactly what has been done in (21).)
We reparametrize each of the pieces such that (see Figure (22))

$$\int_0^1 F_1(t,0)dt - \int_2^3 F_1(3-t,1)dt + \int_1^2 F_2(1,t-1)dt - \int_3^4 F_2(0,4-t)dt$$
$$= \int_0^1 F_1(t,0)dt - \int_0^1 F_1(t,1)dt + \int_0^1 F_2(1,t)dt - \int_0^1 F_2(0,t)dt$$

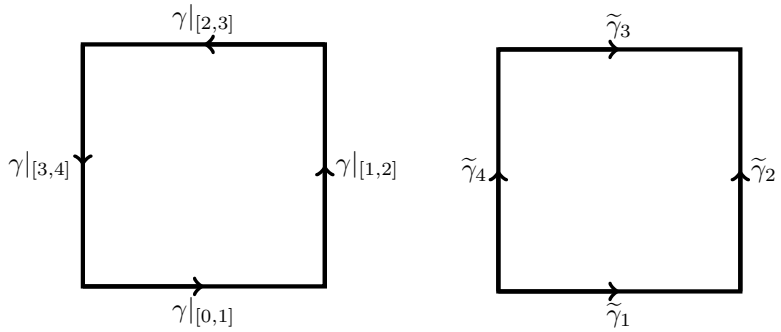


Figure 22: Parametrization γ and piecewise reparametrization.

Then we have

$$\begin{aligned}
&\sum_{j=1}^{4} \int_{j-1}^{j} F(\gamma(t)) \cdot \gamma'(t) dt \\
&= \int_0^1 F_1(t,0)dt - \int_0^1 F_1(t,1)dt + \int_0^1 F_2(1,t)dt - \int_0^1 F_2(0,t)dt \\
&\overset{(1)}{=} -\int_0^1 \int_0^1 D_2F_1(t,s)dsdt + \int_0^1 \int_0^1 D_1F_2(s,t)dsdt \\
&\overset{(2)}{=} \int_0^1 \int_0^1 (D_1F_2 - D_2F_1)(t,s)dsdt \\
&\overset{(3)}{=} 0
\end{aligned}$$

where (1) follows by the fundamental theorem of calculus, in (2) we interchange the integration in s and t, (3) is true by F being irrotational. Thus, the integral along the closed curve γ is 0.
We remark that the derived formula

$$\int_\gamma F = \int_0^1 \int_0^1 (D_1F_2 - D_2F_1)(t,s)dsdt$$

is an instance of the so-called *Stokes theorem*.

We proceed in a more general way. Let $Q = [0,1] \times [0,1]$ and $\varphi : Q \to U \subset \mathbb{R}^d \in C^1$.[10] Let $F : U \to \mathbb{R}^d$ be irrotational and let $\gamma_i : [0,1] \to \mathbb{R}^d$, $i = 1, \ldots, 4$ be given by

$$\begin{aligned}
\gamma_1(t) &= \varphi(t,0), \qquad & \gamma_2(t) &= \varphi(1,t) \\
\gamma_3(t) &= \varphi(t,1), \qquad & \gamma_4(t) &= \varphi(0,t).
\end{aligned}$$

Consider

$$\begin{aligned}
&\int_{\gamma_1} F - \int_{\gamma_3} F + \int_{\gamma_2} F - \int_{\gamma_4} F \\
&= \int_0^1 F(\varphi(t,0)) \cdot D_1\varphi(t,0)dt - \int_0^1 F(\varphi(t,1)) \cdot D_1\varphi(t,1)dt \\
&+ \int_0^1 F(\varphi(1,t)) \cdot D_2\varphi(1,t)dt + \int_0^1 F(\varphi(0,t)) \cdot D_2\varphi(0,t)dt
\end{aligned}$$

[10]The square Q is closed, but one can still define a C^1 map on it - one defines partial derivatives at the boundary such that they point in the perpendicular direction towards the interior of Q.

(which holds as $\gamma_1'(t) = D_1\varphi(t,0)$ etc.) Using

$$F(\varphi(s,t)) \cdot D_j\varphi(s,t) = \sum_{i=1}^{d} F_i(\varphi(s,t))D_j\varphi_i(s,t),$$

the fundamental theorem of calculus and the chain rule, this equals

$$\begin{aligned}
&-\int_0^1\int_0^1 \sum_{i=1}^{d}\sum_{j=1}^{d} D_jF_i(\varphi(t,s))D_2\varphi_j(t,s)D_1\varphi_i(t,s)\\
&\qquad -\sum_{i=1}^{d} F_i(\varphi(t,s))D_2D_1\varphi_i(t,s)dsdt\\
&+\int_0^1\int_0^1 \sum_{i=1}^{d}\sum_{j=1}^{d} D_jF_i(\varphi(s,t))D_1\varphi_j(s,t)D_2\varphi_i(s,t)\\
&\qquad +\sum_{i=1}^{d} F_i(\varphi(s,t))D_1D_2\varphi_i(s,t)dsdt
\end{aligned}$$

By the theorem of Schwarz, the second and the fourth line sum up to zero. In the third line we interchange the order of integration and the order of summation. This gives

$$\int_0^1\int_0^1 \sum_{i=1}^{d}\sum_{j=1}^{d}(D_iF_j - D_jF_i)(\varphi(t,s))D_1\varphi_i(t,s)D_2\varphi_j(t,s)dtds = 0,$$

since F is irrotational and hence $D_iF_j - D_jF_i = 0$ for all i,j.

This shows that there is no such map from $[0,1]\times[0,1] \to \mathbb{R}^2 \setminus \{(0,0)\}$ which would map the boundary of the square to the unit circle. More precisely, we have just shown the following theorem.

Theorem 4.13. *There is no C^1 map $\varphi : [0,1]\times[0,1] \to \mathbb{R}^2 \setminus \{(0,0)\}$ such that for each $t \in [0,1]$,*

$$\begin{aligned}
\varphi(t,0) &= \Big(\cos\frac{\pi t}{2}, \sin\frac{\pi t}{2}\Big)\\
\varphi(1,t) &= \Big(\cos\frac{\pi(t+1)}{2}, \sin\frac{\pi(t+1)}{2}\Big)\\
\varphi(1-t,1) &= \Big(\cos\frac{\pi(t+2)}{2}, \sin\frac{\pi(t+2)}{2}\Big)\\
\varphi(0,1-t) &= \Big(\cos\frac{\pi(t+3)}{2}, \sin\frac{\pi(t+3)}{2}\Big).
\end{aligned}$$

Definition 4.14. Two C^1 paths $\gamma, \widetilde{\gamma} : [0,1] \to U \subset \mathbb{R}^d$ are said to be *homotopic* in U, if $\gamma(0) = \widetilde{\gamma}(0)$, $\gamma(1) = \widetilde{\gamma}(1)$ and there exists a C^1 map (called *homotopy*) $\varphi : [0,1] \times [0,1] \to U$ such that for each $t \in [0,1]$ we have $\varphi(t,0) = \gamma(t), \varphi(t,1) = \widetilde{\gamma}(t)$ and $\varphi(0,t) = \gamma(0), \varphi(1,t) = \gamma(1)$

Intuitively, two maps are homotopic if they have the same starting and ending point and one can be continuously deformed into the other.

Theorem 4.15. *If $F : U \to \mathbb{R}^d$ is irrotational and $\gamma, \widetilde{\gamma} : [0,1] \to U$ are homotopic in U, then*

$$\int_\gamma F = \int_{\widetilde{\gamma}} F.$$

The proof is an immediate consequence of the above calculation.

Theorem 4.16. *If $\gamma, \widetilde{\gamma} : [0,1] \to \mathbb{R}^d$ with $\gamma(0) = \widetilde{\gamma}(0)$, $\gamma(1) = \widetilde{\gamma}(1)$, then γ and $\widetilde{\gamma}$ are homotopic.*

Proof. The map $\varphi : [0,1] \times [0,1] \to \mathbb{R}^d$ given by

$$\varphi(t,s) = s\gamma(t) + (1-s)\widetilde{\gamma}(t)$$

is a homotopy. □

Now we are ready to state a partial converse of Theorem 4.12.

Theorem 4.17. *Let $F : \mathbb{R}^d \to \mathbb{R}^d \in C^1$ be irrotational. Then is F conservative.*

The important part of the theorem is that F maps from $\mathbb{R}^d$. The theorem holds as any two paths $\gamma, \widetilde{\gamma}$ in $\mathbb{R}^d$ are homotopic (by the previous theorem).

More generally, the the same fact holds for convex domains. A subset $U \subset \mathbb{R}^d$ is called *convex*, if for all $x_0, x_1 \in U$ and all $s \in [0,1]$ we have that $sx_0 + (1-s)x_1 \in U$. From the definition of convexity it is clear that it suffices to define a homotopy between any two paths in U. Thus, we have the following stronger version of the previous theorem.[11].

[11]Not part of the lecture: There is an even stronger version. The fact that being irrotational implies being conservative is true on all simply connected domains (pathwise connected domains on which any two continuous paths with the same start- and endpoint are homotopic (with a continuous homotopy)). The class of simply connected domains includes convex domains, but it is larger. For instance, star domains are also examples of simply connected domains. Another instance of such is $\mathbb{R}^d \setminus \{(x,0) : x > 0\}$ from the example from the previous lecture. Intuitively, simply connected domains "have no holes".

Theorem 4.18. *Let $U \subset \mathbb{R}^d$ be convex and $F : U \to \mathbb{R}^d \in C^1$ irrotational. Then F is conservative.*

We conclude by returning to $F(x,y) = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$. Theorem 4.18 implies that F is conservative in the upper half plane $U = \{(x,y) : x > 0\}$.

5 Complex analysis

We consider functions $F : \mathbb{C} \to \mathbb{C}$. We would like to develop differentiation and integration theory for such functions in the same way as for functions mapping $\mathbb{R}$ to $\mathbb{R}$. Recall that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x_0 \in \mathbb{R}$ if there exists a derivative $a := f'(x_0) \in \mathbb{R}$ such that $\forall \varepsilon > 0 \exists \delta > 0 : \forall |h| < \delta$, $|f(x_0 + h) - f(x_0) - ah| \leq \varepsilon |h|$. In the complex case we would like to proceed in the same way: we would like to say that a function F is complex differentiable at $z \in \mathbb{C}$ if there exists $\alpha \in \mathbb{C}$ (which would be called the complex derivative $F'(z)$) such that an analogous condition to the one in the real case holds.
One way would be to define complex differentiability in the same way as for real-valued maps and then look at some properties of complex differentiable functions. However, we rather take the following approach. As $\mathbb{C}$ is identified with $\mathbb{R}^2$, $F : \mathbb{C} \to \mathbb{C}$ is identified with a vector field. We shall restrict our attention to C^1 vector fields. We already know how to differentiate such maps - the differential is a 2×2 matrix. However, there is only a two-dimensional vector subspace of linear maps $\mathbb{R}^2 \to \mathbb{R}^2$ of the form $(x,y) \mapsto (a_1 x - a_2 y, a_1 y + a_2 x)$[12] In other words, only a few complex derivatives would be of the desired form $z \mapsto \alpha z$ for $\alpha \in \mathbb{C}$. Thus, we should impose an additional condition on DF.[13]

Definition 5.1. A map $F : \mathbb{C} \to \mathbb{C} \in C^1$ is called *complex differentiable* at $z \in \mathbb{C}$, if the derivative $DF(z)$ exists and is given via the complex multiplication $DF(z)(\widetilde{z}) = a\widetilde{z}$ for some $a \in \mathbb{C}$.

Recall that complex multiplication is nothing else but the usual multiplication using the identification $i^2 = -1$, so that $(a + ib)(x + iy) = (ax - by) + i(ay + bx)$. We can write this equation in the following matrix form:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix}$$

[12]This corresponds to complex multiplication of (a_1, a_2) and (x,y).

[13]From now on we shall always identify $F : \mathbb{C} \to \mathbb{C}$ with maps $\mathbb{R}^2 \to \mathbb{R}^2$ without further mention. Thus, DF will mean the differential of F as a map $\mathbb{R}^2 \to \mathbb{R}^2$, as well $F \in C^1$ etc.

Since for $F \in C^1$ we have

$$DF = \begin{pmatrix} D_1F_1 & D_2F_1 \\ D_1F_2 & D_2F_2 \end{pmatrix},$$

it follows that $F \in C^1$ is complex differentiable in z if

$$\begin{aligned} D_1F_1(z) &= D_2F_2(z) \\ D_1F_2(z) &= -D_2F_1(z). \end{aligned}$$

This system of equations is also called the *Cauchy-Riemann system* of differential equations. Note that the Cauchy-Riemann system implies that the vector fields $(F_1, -F_2)$ and (F_2, F_1) are irrotational. Thus, if U is convex, they are conservative, i.e. gradient fields. So there exist $f, g \in C^1$ such that

$$\begin{aligned} D_1f &= F_1, \qquad & D_1g &= F_2 \\ D_2f &= -F_2, \qquad & D_2g &= F_1 \end{aligned}$$

In other words, F has a complex antiderivative $(f, g) : U \to \mathbb{C}$ in the sense

$$D(f, g)(z)(\widetilde{z}) = (F_1, F_2)(z)(\widetilde{z}).$$

5.1 Complex path integrals

Let $U \subset \mathbb{C}$, $F : U \to \mathbb{C} \in C^1$, $\gamma : [a, b] \to U$. We define the integral of F along γ by

$$\int_\gamma F := \int_a^b F(\gamma(t))\gamma'(t)dt \tag{22}$$

The multiplication in (22) is now the complex multiplication and not the scalar product in $\mathbb{R}^2$, so this is not the same as the usual path integral. However, we can transfer it to the know path integrals. Indeed, expanding the definition we obtain for (22) the vector of the path integrals

$$\begin{pmatrix} \int_a^b (F_1(\gamma(t)), -F_2(\gamma(t))) \cdot (\gamma_1'(t), \gamma_2'(t))dt \\ \int_a^b (F_2(\gamma(t)), F_1(\gamma(t))) \cdot (\gamma_1'(t), \gamma_2'(t))dt \end{pmatrix}$$

Notice the appearance of the vector fields $(F_1, -F_2)$ and (F_2, F_1), which are by the Cauchy-Riemann system irrotational. Thus, if U is convex, then we have path independence of (22). Equivalently, the integral of F along a closed curve is 0. This fact is also called the *Cauchy integral theorem.*

End of lecture 19. June 29, 2015

6 Rough paths

Recall that a function of bounded variation $f : [a, c] \to \mathbb{R}$ can always be represented as a difference between two monotone functions. Vice versa let $g, h : [a, b] \to \mathbb{R}_{\geq 0}$ be two positive monotone increasing functions and as such, functions of bounded variation. Setting $f = g - h$ we obtain that f is of bounded variation. We thus obtain a precise characterization of functions of bounded variation.

Theorem 6.1. *A real-valued function $f : [a, b] \to \mathbb{R}$ can be represented as a difference of two positive monotone increasing functions if and only if*

$$\|f\|_{V^1} := \sup_{N, a \leq t_0 < t_1 < \cdots < t_N \leq b} \sum_{n=1}^{N} |f(t_n) - f(t_{n-1})| < \infty.$$

Proof.

$\Longrightarrow$: Let $f = g - h$ with g and h monotone increasing positive functions. Then

$$\sum_{n=1}^{N} |f(t_n) - f(t_{n-1})| \leq \sum_{n=1}^{N} |g(t_n) - g(t_{n-1})| + \sum_{n=1}^{N} |h(t_n) - h(t_{n-1})| =$$
$$\sum_{n=1}^{N} g(t_n) - g(t_{n-1}) + h(t_n) - h(t_{n-1})$$
$$= g(t_N) + h(t_N) - g(t_0) - h(t_0) \leq g(b) + h(b) < \infty$$

$\Longleftarrow$: Suppose that $\|f\|_{V^1} < \infty$ and suppose without loss of generality that $f(a) = 0$. Let us define $g : [a, b] \to \mathbb{R}_{\geq 0}$ by setting

$$g(t) = \sup_{N, a \leq t_0 < t_1 < \cdots < t_N \leq t} \sum_{n=1}^{N} |f(t_n) - f(t_{n-1})|$$

i.e. g is the total variation of f up to time t. g is monotone non-decreasing, $g(a) = 0$ and $g(b) = \|f\|_{V^1}$. We claim that $h := g - f$ is also monotone non-decreasing and $h \geq 0$. It is easy to check that $h(a) = 0$ and $h(b) = \|f\|_{V^1} - f(b)$. We must show that given $a \leq t < s \leq b$ we have $h(t) \leq h(s)$. But this is the same as showing that $g(t) - f(t) \leq g(s) - f(s)$ i.e. that $f(s) - f(t) \leq g(s) - g(t)$. We have

that

$$|f(s)-f(t)| \leq \underbrace{\sup_{\substack{a\leq t_0<t_1<\cdots<t_{N-1}<t_N\leq s\\ t_{N-1}=t,\, t_N=s}} \sum_{n=1}^{N} |f(t_n)-f(t_{n-1})|}_{\leq g(s)} -$$

$$\underbrace{\sup_{a\leq t_0<\cdots<t_{N-1}=t} \sum_{n=1}^{N-1} |f(t_n)-f(t_{n-1})|}_{=g(t)} \leq g(t)-g(s)$$

□

The above characterization is relevant for real-valued functions. Only for such functions can one talk about monotonicity and positivity. On the other hand the notion of the variation norm $\|\cdot\|_{V^1}$ is more robust and can be easily extended to vector-valued functions as long as the domain of definition is one dimensional (and thus ordered).

Definition 6.2. A curve $\gamma : [a,b] \to \mathbb{R}^d$ is called rectifiable if

$$\|\gamma\|_{V^1} := \sup_{a\leq t_0<\cdots<t_N\leq b} \sum_{n-1}^{N} \|\gamma(t_n)-\gamma(t_{n-1})\| < \infty$$

Given a rectifiable curve we call $\|\gamma\|_{V^1}$ the length of γ.

Theorem 6.3. *A continuous curve $\gamma \in C\left([a,b];\mathbb{R}^d\right)$ is rectifiable if and only if all the components γ_i for $i \in \{1,\ldots,d\}$ are of bounded variation.*

Proof.

$\Longrightarrow$: Suppose that γ is of bounded variation. For any partition $a \leq t_0 < \cdots < t_N \leq b$ we trivially have

$$\sum_{n=1}^{N} |\gamma_i(t_n)-\gamma_i(t_{n-1})| \leq \sum_{n=1}^{N} \|\gamma(t_n)-\gamma(t_{n-1})\|.$$

$\Longleftarrow$: Using the fact that $\|\gamma(t_n)-\gamma(t_{n-1})\| \leq d\sum_{i=1}^{d} |\gamma_i(t_n)-\gamma_i(t_{n-1})|$ one can conclude the proof. This is left as an exercise.

□

The variation norm V^1 and is a general concept that does not require differentiability of a path. It is useful however to relate this quantity to differential quantities of a path if it happens to be sufficiently smooth.

Theorem 6.4. *Let $\gamma : [a,b] \to \mathbb{R}^d$ be continuously differentiable (C^1), then $\|\gamma\|_{V^1} = \int_a^b \|\gamma'(t)\| dt$.*

To prove this theorem we first introduce the following Lemma that states that when considering the V^1 norm of a path we can restrict ourselves to looking only at time partitions that are very fine i.e. we can impose that $t_i - t_{i-1}$ be arbitrarily small.

Lemma 6.5. *For any $\epsilon > 0$ we have that*

$$\|\gamma\|_{V^1} = \sup_{\substack{N,\, a \le t_0 < \cdots < t_N \le b \\ \forall n\ |t_n - t_{n-1}| < \epsilon}} \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\|.$$

Proof. Adding the condition that $\forall n \in \{1, \dots, N\}\ t_n - t_{n-1} < \epsilon$ restricts the set of competitors for the sup and thus makes the right hand side smaller so we trivially have

$$\|\gamma\|_{V^1} \ge \sup_{\substack{N,\, a \le t_0 < \cdots < t_N \le b \\ \forall n\ |t_n - t_{n-1}| < \epsilon}} \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\|.$$

Now let us prove that $\|\gamma\|_{V^1} \le \sup_{\substack{N,\, a \le t_0 < \cdots < t_N \le b \\ \forall n\ |t_n - t_{n-1}| < \epsilon}} \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\|$. Consider any sequence of times $a \le t_0 < \cdots < t_N \le b$, we now refine the partition $t_0, \dots, t_n$ by choosing $M \in \mathbb{N}$ and points $a \le s_0 < \cdots < s_M = b$ so that for any $m \le M$ we have $s_m - s_{m-1} < \epsilon$ and $\forall n \le N$ there exists $m_n \le M$ so that $s_{m_n} = t_m$. Then

$$\begin{aligned} \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\| &\le \sum_{n=1}^{N} \sum_{t_{n-1} \le s_m < t_n} \|\gamma(s_n) - \gamma(s_{n-1})\| \\ &= \sum_{m=1}^{M} \|\gamma(s_m) - \gamma(s_{m-1})\| \end{aligned}$$

This shows that

$$\sup_{\substack{N,\, a \le t_0 < \cdots < t_N \le b \\ \forall n\ |t_n - t_{n-1}| < \epsilon}} \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\| \ge \sup_{N,\, a \le t_0 < \cdots < t_N \le b} \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\|$$

as required. □

We now proceed to the proof of the Theorem

Proof. Let us fix some $\delta > 0$ and choose an $\epsilon > 0$ so that $|t - s| < \epsilon$ implies that $\|\gamma'(t) - \gamma'(s)\| \leq \delta$ via uniform continuity of the derivative of γ on the compact interval $[a, b]$.
We will show that for any sequence of times $a \leq t_0 < \cdots < t_N \leq b$ with $|t_n - t_{n-1}| < \epsilon$ for all $n \in \{1, \ldots, n\}$ the following holds.

$$\left| \sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\| - \int_a^b \|\gamma'(t)\| dt \right| \leq 4\delta d|b - a|$$

Since $\delta > 0$ can be chosen to be arbitrarily small we obtain the needed result. To show the above bound it is sufficient to show that for all n we have

$$\left| \|\gamma(t_n) - \gamma(t_{n-1})\| - \int_{t_{n-1}}^{t_n} \|\gamma'(t)\| dt \right| \leq 4\delta d|t_n - t_{n-1}|$$

Since $\gamma(t_n) - \gamma(t_{n-1}) = \int_{t_{n-1}}^{t_n} \gamma'(t)dt$ we need to bound $\left\| \int_{t_{n-1}}^{t_n} \gamma'(t)dt \right\| - \int_{t_{n-1}}^{t_n} \|\gamma'(t)\| dt$ but since for all $t \in [t_{n-1}, t_n)$ we have that $\|\gamma'(t) - \gamma(t_{n-1})\| < \delta$ we obtain by the triangle inequality

$$\left\| \int_{t_{n-1}}^{t_n} \gamma'(t)dt \right\| = |t_n - t_{n-1}| \|\gamma'(t_{n-1})\| + \left\| \int_{t_{n-1}}^{t_n} (\gamma(t) - \gamma(t_{n-1}))\, dt \right\|$$

$$\int_{t_{n-1}}^{t_n} \|\gamma'(t)\|\, dt = |t_n - t_{n-1}| \|\gamma'(t_{n-1})\| + \int_{t_{n-1}}^{t_n} (\|\gamma(t)\| - \|\gamma(t_{n-1})\|)\, dt$$

$$\implies \qquad \left\| \int_{t_{n-1}}^{t_n} \gamma'(t)dt \right\| - \int_{t_{n-1}}^{t_n} \|\gamma'(t)\| dt < 4\delta d|t_n - t_{n-1}|$$

as required. □

Since we have naturally generalized the concept of length from C^1 curves to more general rectifiable ones the next step consists of understanding if one can give a meaning to path integrals for paths of lower regularity. Recall that we have defined the path integral of a vector field F via

$$\int_\gamma F = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt.$$

While the expression on the left hand side contains a derivative that is not necessarily defined for generic rectifiable γ we can avoid this pitfall in a way similar to the one we used to define length.

Theorem 6.6. *Let F be a continuous function on $U \subset \mathbb{R}^d$ and let $\gamma : [a,b] \to U$ be a continuous and rectifiable path. Then $\forall \delta > 0\ \exists \epsilon > 0$ so that for any N, M and time sequences $a \leq t_0 < \cdots < t_N \leq b$ and $a \leq s_0 < \cdots < s_M \leq b$ with $|t_n - t_{n-1}| < \epsilon$ and $|s_m - s_{m-1}| < \epsilon$ we have*

$$\left| \sum_{m=1}^{M} F(\gamma(s_{m-1}))\,(\gamma(s_m - s_{m-1})) - \sum_{n=1}^{N} F(\gamma(t_{n-1}))\,(\gamma(t_n - t_{n-1})) \right| < \delta$$

In particular this means if for every $k \in \mathbb{N}$ we choose a sequence of time partitions $a \leq t_0^{(k)} < t_1^{(k)} < \cdots < t_{N_k-1}^{(k)} < t_{N_k}^{(k)} \leq b$ with $N_k \to \infty$ and $|t_n - t_{n-1}| < \frac{1}{2^k}$ for all $n \in \{1, \ldots, N_k\}$ then the limit

$$\lim_{k\to\infty} \sum_{n=1}^{N_k} F(\gamma(t_{n-1}^{(k)}))\left(\gamma(t_n^{(k)} - t_{n-1}^{(k)})\right)$$

exists and is independent of the sequence of time-partitions $\left(t_n^{(k)}\right)$.

Given the Theorem above, using the same notation we define

$$\int_\gamma F := \lim_{k\to\infty} \sum_{n=1}^{N_k} F(\gamma(t_{n-1}^{(k)}))\left(\gamma(t_n^{(k)} - t_{n-1}^{(k)})\right)$$

for rectifiable paths γ. We will later show that this definition coincides with the one we have given for C^1 paths if γ happens to be also C^1. First however we proceed to the proof of the above Theorem.

Proof. Without loss of generality we can suppose that the partition (s_m) is a refinement of the partition t_n i.e. suppose that $\forall n\, \exists m$ such that $s_m = t_n$. This is true because given any two partitions (t_n) and $(t'_{n'})$ with $|t_n - t_{n-1}| < \epsilon$ and $|t'_{n'} - t'_{n'-1}| < \epsilon$ we can find a common refinement (s_m) of both and then the statement would follow by a triangle inequality.
$F \circ \gamma$ is continuous and thus uniformly continuous on the interval $[a,b]$. For any fixed $\delta > 0$ let us choose $\epsilon > 0$ so that $|x - y| < \epsilon$ implies $\|F(\gamma(y)) - F(\gamma(x))\| < \delta$. Let t_n be such that for any n we have that $t_n - t_{n-1} < \epsilon$ then

for any $t_{n-1} < s_m \leq t_n$ we also have that $s_m - t_{n-1} < \epsilon$ so

$$\left| \sum_{t_{n-1}<s_m\leq t_n} F(\gamma(s_{m-1})) \cdot (\gamma(s_m) - \gamma(s_{m-1})) - F(\gamma(t_{n-1})) \cdot (\gamma(t_n) - \gamma(t_{n-1})) \right| < \delta \|\gamma(t_n) - \gamma(t_{n-1})\|$$

$$\Rightarrow \left| \sum_{m=1}^{M} F(\gamma(s_{m-1})) \cdot (\gamma(s_m) - \gamma(s_{m-1})) - \sum_{n=1}^{N} F(\gamma(t_{n-1})) \cdot (\gamma(t_n) - \gamma(t_{n-1})) \right| < \sum_{n}^{N} \delta \|\gamma(t_n) - \gamma(t_{n-1})\| \leq \delta \|\gamma\|_{V^1}.$$

Applying this statement with $\delta' = \frac{\delta}{\|\gamma\|_{V^1}}$ yields the required result. The second part of the Theorem is left as an exercise and follows by noticing that $\lim_{k\to\infty} \sum_{n=1}^{N_k} F(\gamma(t_{n-1}^{(k)})) \left(\gamma(t_n^{(k)} - t_{n-1}^{(k)})\right)$ is a Cauchy sequence in $\mathbb{R}$. □

The candidate for the definition of a path integral along non C^1 paths of bounded variation effectively extends the notion we introduced for C^1 paths. This is due to the following theorem that states that the two notions coincide if γ is a more regular path.

Theorem 6.7. *Let $\gamma \in C^1([a,b]; U)$ and $F \in C\left(U; \mathbb{R}^d\right)$ with $U \subset \mathbb{R}^d$ then $\forall \delta > 0$ there exists $\epsilon > 0$ so that if for a sequence of times $a \leq t_0 < t_1 < \cdots < t_N \leq b$ one has $|t_n - t_{n-1}| < \epsilon$ for all $n \in \{1, \ldots, N\}$ then*

$$\left| \overbrace{\int F(\gamma(t))\gamma'(t)dt}^{\int_\gamma F :=} - \underbrace{\sum_{n=1}^{N} F(\gamma(t_{n-1}))(\gamma(t_n) - \gamma(t_{n-1}))}_{\text{defined since } \|\gamma\|_{V^1} < \infty} \right| < \delta.$$

Proof. We leave the proof of this Theorem as an exercise as it is similar to the proof that the length $\|\gamma\|_{V^1}$ of a C^1 path is given by $\int_a^b \|\gamma'(t)\|dt$. □

Our interest now is to further extend the definition a path integrals to allow for even less regular paths. This will be done by requiring possibly greater regularity on the vector field F. Let us recall the definition of variation norms for $p \in [1, \infty)$.

Definition 6.8. Let $\gamma : [a,b] \to U \subset \mathbb{R}^d$ be a continuous path. We say that γ is a V^p path if its V^p norm given by

$$\|\gamma\|_{V^p} := \sup_{N,\, a\leq t_0<\cdots<t_N\leq b} \left(\sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\|^p \right)^{\frac{1}{p}}$$

is finite.

Let us consider as an example the path $\gamma : [0, 2\pi] \to \mathbb{R}^2$ that winds M times around the origin. Set

$$\gamma(t) = A \begin{pmatrix} \cos(Mt) \\ \sin(Mt) \end{pmatrix}$$

with some $M \in \mathbb{N}$ that we consider large and $A \in \mathbb{R}^+$ that will be chosen small. Let us estimate the V^p norm of the path γ. First we estimate $\|\gamma\|_{V^p}$ from above. Let $0 \leq t_0 < \cdots < t_N \leq 2\pi$ be some sequence of times so we can subdivide them based on which "turn" around $0 \in \mathbb{R}^2$ we are making. We have

$$\sum_{n=1}^{N} \|\gamma(t_n) - \gamma(t_{n-1})\|^p =$$

$$\sum_{m=1}^{M} \left[\sum_{2\pi\frac{m-1}{M} \leq t_{n-1} < t_n < 2\pi\frac{m}{M}} \|\gamma(t_n) - \gamma(t_{n-1})\|^p + \sum_{t_{n-1} < 2\pi\frac{m}{M} \leq t_n} \|\gamma(t_n) - \gamma(t_{n-1})\|^p \right] \leq$$

$$\sum_{m=1}^{M} \left[\left(\sum_{2\pi\frac{m-1}{M} \leq t_{n-1} < t_n < 2\pi\frac{m}{M}} \|\gamma(t_n) - \gamma(t_{n-1})\| (2A)^{p-1} \right) + (2A)^p \right] \leq$$

$$\sum_{m=1}^{M} \left[\left(\sum_{2\pi\frac{m-1}{M} \leq t_{n-1} < t_n < 2\pi\frac{m}{M}} \|\gamma'\|_\infty \left(t_n - t_{n-1}\right) (2A)^{p-1} \right) + (2A)^p \right] \leq$$

$$\sum_{m=1}^{M} \left(\frac{2\pi}{M} \|\gamma'\|_\infty (2A)^{p-1} + (2A)^p \right)$$

and since $\|\gamma'\| = AM$ we have by taking the power $\frac{1}{p}$ of the above expression that $\|\gamma\|_{V^p} \leq CM^{\frac{1}{p}}A$. On the other hand we have a lower bound given by

$$\sum_{m=1}^{2M} \left\| \gamma\left(\frac{m}{2M}\right) - \gamma\left(\frac{m-1}{2M}\right) \right\|^p = \sum_{m=1}^{2M} (2A)^p = 2M(2A)^p \leq \|\gamma\|_{V^p}$$

so $\frac{1}{C}M^{\frac{1}{p}}A \leq \|\gamma\|_{V^p} \leq CM^{\frac{1}{p}}A$ for some constant $C > 1$. Let us set $A := M^{-\frac{1}{r}}$ so that $\|\gamma\|_{V^p} \approx M^{\frac{1}{p}-\frac{1}{r}}$. We have that

$$\|\gamma\|_{V^p} \to \begin{cases} \infty & \text{if } p < r \\ 0 & \text{if } p > r \end{cases} \qquad \text{as } M \to \infty$$

while if $p = r$ the $\|\gamma\|_{V^p}$ remains bounded. Now let us take the vector field on $\mathbb{R}^2$ given by $F(x,y) = \begin{pmatrix} -y \\ x \end{pmatrix}$ so that

$$\int_\gamma F = \int_0^{2\pi} F(\gamma(t))\gamma'(t)dt =$$

$$\int_0^{2\pi} A\begin{pmatrix} -\sin(Mt) \\ \cos(Mt) \end{pmatrix} \cdot AM\begin{pmatrix} -\sin(Mt) \\ \cos(Mt) \end{pmatrix} dt = \int_0^{2\pi} A^2 M dt = 2\pi A^2 M$$

With our previous choice of normalization $A = M^{-\frac{1}{r}}$ we have that

$$\int_\gamma F = 2\pi M^{-\frac{2}{r}} M \to \begin{cases} 0 & \text{if } r < 2 \\ 2\pi & \text{if } r = 2 \\ \infty & \text{if } r > 2 \end{cases} \qquad \text{as } M \to \infty$$

Geometrically the graph of the path γ converges to the trivial path $\gamma(t) = 0$ so we would expect that for sufficiently regular paths the integral $\int_\gamma F \to 0$. The heuristic is that if $\gamma \to 0$ in V^p norm then we should expect that $\int_\gamma F \to 0$
From the above example it becomes clear that if $p > 2$ we can choose $2 < r < p$ and we would have tat $\gamma \to 0$ in V^p but $\int_\gamma F \not\to 0$. Thus the "critical" regularity is that of V^2.
First, however we would want to illustrate a property about counting th jumps of a V^p function.

Lemma 6.9. *Let $\gamma : [a, b] \to U \subset \mathbb{R}^d$ be continuous and let $\|\gamma\|_{V^p} < \infty$. For $\epsilon > 0$ let us define $t_0 = a$ and $t_n = \inf\{t > t_{n-1} \mid \|\gamma(t_n) - \gamma(t_{n-1})\| \geq \epsilon\}$ if such a t exists and $t_n = b$ otherwise and set $N = n - 1$ for the first n for which $t_n = b$. Since γ is continuous we have that $\|\gamma(t_n) - \gamma(t_{n-1})\| = \epsilon$ so $\sum_{n=1}^N \|\gamma(t_n) - \gamma(t_{n-1})\|^p = N\epsilon^p \leq \|\gamma\|_{V^p}^p$ and thus $N \leq \frac{\|\gamma\|_{V^p}^p}{\epsilon^p}$.*

Theorem 6.10. *Define*

$$I_\epsilon := \sum_{n=1}^N F\left(\gamma(t_{n-1})\right) \cdot \left(\gamma(t_n) - \gamma(t_{n-1})\right)$$

with $a = t_0 < t_1 < \cdots < t_N \leq b$ a sequence of times chosen as in the previous Lemma.
Let $F \in C^1(U; \mathbb{R}^d)$ and let $\gamma : [a, b] \to U \subset \mathbb{R}^d$ with $\|\gamma\|_{V^p} < \infty$ for some $p < 2$ then the limit $\lim_{\epsilon\to 0} I_\epsilon$ exists. This limit defines the expression

$$\int_\gamma F = \int_a^b F(\gamma(t)) \cdot \gamma'(t)dt := \lim_{\epsilon\to 0} I_\epsilon.$$

Proof. Let ϵ_0 be sufficiently small so that $\forall x, y \in \mathrm{im}(\gamma)$ with $\|x-y\| < \epsilon_0$ we have that

$$\left\| F(y) - F(x) - \sum_{j=1}^{d} D_j F(x)\,(y_j - x_j) \right\| \leq \|x-y\|$$
$$\implies \|F(y) - F(x)\| \leq (1 + \|\nabla F\|_\infty)\,\|x-y\|.$$

Such an ϵ_0 exists because $\mathrm{im}(\gamma)$ is compact. Let $\epsilon < \epsilon_0$ and $\frac{\epsilon}{2} < \epsilon' < \epsilon$ and let $t_0, \ldots, t_N$, $N \in \mathbb{N}$ and $t'_0, \ldots, t'_{N'}$, $N' \in \mathbb{N}$ be two sequences of times corresponding to ϵ and ϵ' as in the previous Lemma.
Let us refine the two partitions to obtain a third partition $a = s_0 < s_1 < \cdots < s_M \leq b$ with $\{s_m\} = \{t_n\} \cup \{t'_n\}$ and $M \leq N + N' \leq C\left(\frac{1}{\epsilon^p} + \frac{1}{(\epsilon')^p}\right) \leq C' \frac{1}{(\epsilon')^{-p}}$. The two last inequalities come from the consideration in the previous Lemma.

$$\sum_{n=1}^{N} F\left(\gamma(t_{n-1})\right)\left(\gamma(t_n) - \gamma(t_{n-1})\right) = \sum_{n=1}^{N} \sum_{t_{n-1} < s_m \leq t_n} F\left(\gamma(t_{n-1})\right)\left(\gamma(s_m) - \gamma(s_{m-1})\right)$$

so

$$\sum_{n=1}^{N} F\left(\gamma(t_{n-1})\right)\left(\gamma(t_n) - \gamma(t_{n-1})\right) - \sum_{m=1}^{M} F\left(\gamma(s_{m-1})\right)\left(\gamma(s_m) - \gamma(s_{m-1})\right) =$$
$$\sum_{n=1}^{N} \sum_{t_{n-1} < s_m \leq t_n} F\left(\gamma(t_{n-1})\right)\left(\gamma(s_m) - \gamma(s_{m-1})\right) - \sum_{m=1}^{M} F\left(\gamma(s_{m-1})\right)\left(\gamma(s_m) - \gamma(s_{m-1})\right) =$$
$$\sum_{n=1}^{N} \sum_{t_{n-1} < s_m \leq t_n} \left[F\left(\gamma(t_{n-1})\right) - F\left(\gamma(s_{m-1})\right)\right]\left(\gamma(s_m) - \gamma(s_{m-1})\right)$$

and thus

$$\left\| \sum_{n=1}^{N} F\left(\gamma(t_{n-1})\right)\left(\gamma(t_n) - \gamma(t_{n-1})\right) - \sum_{m=1}^{M} F\left(\gamma(s_{m-1})\right)\left(\gamma(s_m) - \gamma(s_{m-1})\right) \right\| \leq$$
$$\sum_{n=1}^{N} \sum_{t_{n-1} < s_m \leq t_n} \left(\|\nabla F\|_\infty + 1\right) \|\gamma(t_{n-1}) - \gamma(s_{m-1})\|\,\|\gamma(s_m) - \gamma(s_{m-1})\| \leq$$
$$\sum_{n=1}^{N} \sum_{t_{n-1} < s_m \leq t_n} \left(\|\nabla F\|_\infty + 1\right) \epsilon^2 \leq C_F M \epsilon^2 \leq C_F \epsilon^{2-p}.$$

Applying the same procedure to the sequence (t'_n) we obtain via the triangle inequality that

$$|I_\epsilon - I_{\epsilon'}| \leq \tilde{C}_F \epsilon^{2-p} \qquad \frac{\epsilon}{2} \leq \epsilon' \leq \epsilon.$$

By a geometric series trick this implies that we can prove the same for closeness result for any any $\epsilon' < \epsilon < \epsilon_0$. As a matter of fact let us write down the geometric sequence $\epsilon, \frac{\epsilon}{2}, \frac{\epsilon}{4} \dots \frac{\epsilon}{2^{k_0-1}} > \epsilon' \geq \frac{\epsilon}{2^{k_0}}$. The expression $|I_\epsilon - I_{\epsilon'}|$ can be rewritten as a telescoping sum to obtain

$$|I_\epsilon - I_{\epsilon'}| \leq \sum_{k=1}^{k_0} \left| I_{\frac{\epsilon}{2^{k-1}}} - I_{\frac{\epsilon}{2^k}} \right| + \left| I_{\frac{\epsilon}{2^{k_0}}} - I_{\epsilon'} \right| \leq \sum_{k=1}^{k_0} \tilde{C}_F \epsilon^{2-p} 2^{-k(2-p)} \leq C_p \tilde{C}_F \epsilon^{2-p}$$

as required. □

The above proof relies on using an approximation of increments of F with its derivative to "compensate" the fact that the path γ is not regular enough, i.e. that $\gamma \in V^p$ for some $p \in [1, 2)$ but γ is not necessarily in V^1. To be able to offer a similar definition of path integrals for $\gamma \in V^p$ with $p \in [2, 3)$ or even higher one requires that F be even more regular. To prove a Theorem similar to the one above we would need to expand F to the second or further orders. However a second order Taylor expansion of F would involve a term that looks like $\sum_{i,j=1}^d D^2_{i,j} F(t_{n-1}) (\gamma_i(t_n) - \gamma_i(t_{n-1})) (\gamma_j(t_n) - \gamma_j(t_{n-1}))$. Taking progressively finer partitions (t_n) for $\gamma \in V^p$ with $p \geq 2$ we would encounter a problem accounting for the bilinear quantity $(\gamma_i(t_n) - \gamma_i(t_{n-1})) (\gamma_j(t_n) - \gamma_j(t_{n-1}))$. Let us try to understand what this quantity looks and what algebraic properties it possesses for smooth paths γ. Let $\gamma \in C^1\left([a, b]; \mathbb{R}^d\right)$ and let

$$A_{i,j}(c, d) = \int_c^d (\gamma_j(t) - \gamma_j(c)) \gamma'_i(t) dt.$$

We have that

$$A_{i,j}(a, d) = A_{i,j}(a, c) + A_{i,j}(c, d) + \int_c^d (\gamma_j(c) - \gamma_j(a)) \gamma'_i(t) dt.$$

This is not a linear quantity in time but has the above non-trivial algebraic structure. This consideration warrants the following definition.

Definition 6.11. Let $2 \leq p < 3$, a rough path is a pair (γ, A) consisting of a path $\gamma : [a, b] \to \mathbb{R}^d$ with $\|\gamma\|_{V^p} < \infty$ and and a matrix-value function

$A : [a,b] \times [a,b] \to \mathbb{R}^d \times \mathbb{R}^d$ i.e. $A_{i,j} : [a,b] \times [a,b] \to \mathbb{R}$ for all $i,j \in \{1,\dots,d\}$ that satisfies

$$A_{i,j}(a,d) = A_{i,j}(a,c) + A_{i,j}(c,d) + \int_c^d (\gamma_j(c) - \gamma_j(a))\, \gamma_i'(t) dt.$$

and

$$\sup_{N,\, a \le t_0 < \dots, t_N = b} \left(\sum |A_{i,j}(t_{n-1}, t_n)|^{\frac{p}{2}} \right)^{\frac{2}{p}} < \infty.$$

We define

$$I_\epsilon := \sum_{n=1}^{N} F(\gamma(t_{n-1}))\, (\gamma(t_n) - \gamma(t_{n-1})) + \sum_{i,j=1}^{d} D_j F_j(t_{n-1}) A_{i,j}(t_{n-1}, t_n)\, (\gamma_i(t_n) - \gamma_i(t_{n-1}))$$

Theorem 6.12. *If F is a C^2 vector field then $\lim_{\epsilon \to 0} I_\epsilon$ exists and defines $\int_\gamma F$ for a rough path γ.*

End of lecture 20 & 21. July 02, 2015 & July 06, 2015

7 Hairy ball theorem

Definition 7.1. The unit sphere in $\mathbb{R}^{n+1}$ is the set

$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}.$$

A vector field on $\mathbb{S}^n$ is a continuous map $\phi : \mathbb{S}^n \to \mathbb{R}^{n+1}$ such that $\phi(x)$ is tangent to $\mathbb{S}^n$ for each $x \in \mathbb{S}^n$, i.e. $\langle \phi(x), x \rangle = 0$ for every $x \in \mathbb{S}^n$.

Note that $\mathbb{S}^n$ is bounded. It is also closed, as it is the boundary of the unit ball $\mathbb{S}^n = \partial B(0,1) \subset \mathbb{R}^{n+1}$. Thus, $\mathbb{S}^n$ is compact.

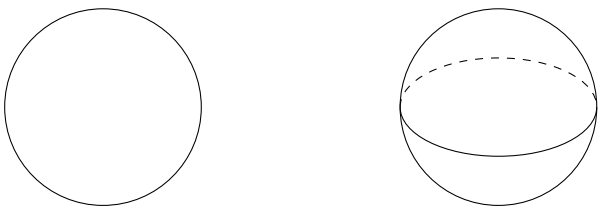


Figure 23: $\mathbb{S}^1$ and $\mathbb{S}^2$.

Theorem 7.2 (Hairy ball theorem[14])**.** *If n is even, there does not exist a vector field on $\mathbb{S}^n$ which is everywhere non-zero.*

This theorem can be interpreted in the following way: one cannot comb a hairy ball flat without creating a cowlick. Alternatively, at all times there must be at least one point on the surface of the Earth where there is no wind at all.

Remark. It is easily seen that the theorem fails for odd n. Consider first the case $n = 1$ and the vector field $F : \mathbb{S}^1 \to \mathbb{R}^2$ given by

$$F(x, y) = (-y, x).$$

Note that this is the map $(x, y) \mapsto (-\frac{y}{r^2}, \frac{x}{r^2})$, $r^2 = x^2 + y^2$, restricted to the unit sphere $\mathbb{S}^1$ on which $r = 1$. We have already met this vector field in Lecture 18. Clearly, the vector field F has no zeroes.

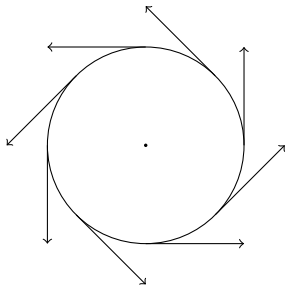


Figure 24: Vector field $F(x, y) = (-y, x)$ on $\mathbb{S}^1$.

For general odd n we define $\phi : \mathbb{S}^n \to \mathbb{R}^{n+1}$ by

$$\phi(x_1, x_2, \ldots, x_n, x_{n+1}) = (-x_2, x_1, \ldots, -x_{n+1}, x_n)$$

Note that since $n + 1$ is even, we are able to group x_1, x_2 and x_3, x_4, $\ldots$, x_n, x_{n+1}, and for each pair we reproduce the example from the case $n = 1$. The map ϕ is continuous (it is even C^∞) and $\langle \phi(x), x \rangle = x_1(-x_2) + x_2 x_1 + \cdots + x_n(-x_{n+1}) + x_{n+1}x_n = 0$. Moreover, ϕ is an isometry: $\|\phi(x)\| = \|x\|$ for all $x \in \mathbb{S}^n$. In particular, $\|\phi(x)\| = 1$ for all $x \in \mathbb{S}^n$, and so ϕ has no zeroes.

Definition 7.3. A vector field $\phi : \mathbb{S}^n \to \mathbb{R}^{n+1}$ is called *continuously differentiable* (C^1) if there exists a neighborhood $U \subset \mathbb{R}^{n+1}$ of $\mathbb{S}^n$ and a C^1 function $f : U \to \mathbb{R}^{n+1}$ such that $f|_{\mathbb{S}^n} = \phi$.

[14]dt. Satz vom Igel.

7.1 Proof of Theorem 7.2

The proof is by contradiction. Let n be even and assume that ϕ is an everywhere non-zero vector field on $\mathbb{S}^n$. We proceed in three steps.

(1) Reduction to the smooth case
The purpose of this step is to show that it suffices to prove the theorem for $\phi \in C^1$. The reader may skip this technical part at first and return to it at the end of the lecture.
Let $\phi : \mathbb{S}^n \to \mathbb{R}^n$ be a continuous nowhere vanishing vector field. Let

$$m := \inf_{x \in \mathbb{S}^n} \|\phi(x)\| = \min_{x \in \mathbb{S}^n} \phi(x).$$

By a generalization of the Stone-Weierstrass theorem, we can find a polynomial mapping $P : \mathbb{S}^n \to \mathbb{R}^{n+1}$ such that

$$\|P(x) - \phi(x)\| < \frac{m}{2}$$

for every $x \in \mathbb{S}^n$. (Alternatively, one could approximate ϕ with smooth functions by convolving it with the elements of a Dirac sequence.) Define $w : \mathbb{S}^n \to \mathbb{R}^{n+1}$ by

$$w(x) = P(x) - \langle P(x), x \rangle x.$$

This is a C^1 vector field on $\mathbb{S}^n$, since

$$\begin{aligned} \langle w(x), x \rangle =& \langle \langle P(x), x \rangle x, x \rangle \\ =& \langle P(x), x \rangle - \langle P(x), x \rangle \underbrace{\langle x, x \rangle}_{\|x\|^2 = 1} = 0 \end{aligned}$$

for all $x \in \mathbb{S}^n$. We claim that w has no zeroes. To see this we first compute

$$\begin{aligned} \|w(x) - P(x)\| =& |\langle P(x), x \rangle| \underbrace{\|x\|}_{=1} \\ =& |\langle P(x), x \rangle - \langle \phi(x), x \rangle| \\ =& |\langle P(x) - \phi(x), x \rangle| \\ \overset{(*)}{\leq}& \|P(x) - \phi(x)\| < \frac{m}{2}, \end{aligned}$$

where in $(*)$ we used the Cauchy-Schwarz inequality. Assume that $w(x_0) = 0$. Then $\|P(x_0)\| < \frac{m}{2}$. Also, the triangle inequality implies

$$\|\phi(x_0)\| - \|P(x_0)\| \leq \|P(x_0) - w(x_0)\| < \frac{m}{2}.$$

Hence

$$\|\phi(x_0)\| < \frac{m}{2} + \|P(x_0)\| < m,$$

which is in contradiction to the definition of m.
The map $\widetilde{\phi} : \mathbb{S}^n \to \mathbb{R}^{n+1}$, $\widetilde{\phi} = \frac{w(x)}{\|w(x)\|}$, is then a C^1 (even C^∞) vector field on $\mathbb{S}^n$ without zeroes.
Thus it suffices to proceed with the proof assuming $\phi \in C^1$.

(2) Two preparatory lemmata
Let $V \subset \mathbb{R}^{n+1}$ be open and $K \subset V$ compact. (For instance, we might take $K = \mathbb{S}^n$ and $V = \{x \in \mathbb{R}^{n+1} : a < \|x\| < b\}$ for $0 < a < 1 < b$.) Let $F : V \to \mathbb{R}^{n+1}$ be C^1. For $t > 0$ we define a C^1-perturbation of the identity $\phi_t : V \to \mathbb{R}^{n+1}$ which is given by $\phi_t(x) = x + tF(x)$. We have $\phi_t \in C^1(V; \mathbb{R}^{n+1})$ and $D\phi_t(x) = (I_{n+1} + tDF)(x)$, where I_{n+1} denotes the $(x+1) \times (n+1)$ identity matrix. As an example, for $n = 1$ we have

$$D\phi_t(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + t \begin{pmatrix} \partial_x F_1 & \partial_y F_1 \\ \partial_x F_2 & \partial_y F_2 \end{pmatrix} = \begin{pmatrix} 1 + t\partial_x F_1 & t\partial_y F_1 \\ t\partial_x F_2 & 1 + t\partial_y F_2 \end{pmatrix}$$

The Jacobian determinant is[15]

$$\det(D\phi_t(x)) = 1 + \underbrace{(\partial_x F_1 + \partial_y F_2)}_{\mathrm{div(F)}}\, t + \underbrace{((\partial_x F_1)(\partial_y F_2) - (\partial_y F_1)(\partial_x F_2))}_{\det(DF)}\, t^2,$$

which is a polynomial in t with the constant term 1. For general n we obtain a polynomial

$$\det(D\phi_t(x)) = \underbrace{p_0(x)}_{=1} + p_1(x)t + \cdots + p_{n+1}(x)t^{n+1}$$

for certain coefficients $p_j(x)$, $j = 1, \ldots, n+1$. Note that, for sufficiently small t, one has $\det(D\phi_t(x)) > 0$.

Lemma 7.4. *There exists $\delta > 0$ such that for $0 < t < \delta$*

1. *The map $\phi_t : K \to \phi_t(K)$ is bijective and its inverse is C^1.*
2. *The map $t \mapsto \mathrm{vol}(\phi_t(K))$ is a polynomial in t.*

Proof. 1. Since $F \in C^1(V; \mathbb{R}^{n+1})$, the mapping DF restricted to the compact set K is bounded. This implies that there exists $L > 0$ such that $\|F(x) -$

[15] $\mathrm{div}(F) = \nabla \cdot F = \frac{\partial F_1}{\partial x_1} + \cdots + \frac{\partial F_n}{\partial x_n}$ is called the *divergence* of a vector field $F : \mathbb{R}^n \to \mathbb{R}^n$.

$F(y)\| \leq L\|x-y\|$ for all $x, y \in K$. The reader may verify this claim as an exercise. Then, for $x, y \in K$, we have that

$$\begin{aligned}\|\phi_t(x) - \phi_t(y)\| =& \|(x-y) + t(F(x) - F(y))\| \\ \geq& \|x-y\| - t\|F(x) - F(y)\| \\ \geq& \|x-y\| - tL\|x-y\| \\ =& (1-tL)\|x-y\|.\end{aligned}$$

This shows that for $t < \frac{1}{L}$, $\|\phi_t(x) - \phi_t(y)\| = 0$ implies $\|x-y\| = 0$, i.e. ϕ_t is injective. Continuous differentiability of the inverse map follows from the inverse mapping theorem, since $\det(D\phi_t(x)) > 0$ for t small enough, i.e., the Jacobi matrix is invertible at each point.

2. By the change of variables formula, we have

$$\mathrm{vol}(\phi_{\mathrm{t}}(\mathrm{K})) = \int_K |\det(D\phi_t(x))| dx.$$

For $t > 0$ sufficiently small, it then follows that

$$\begin{aligned}\mathrm{vol}(\phi_t(K)) =& \int_K \det(D\phi_t(x)) dx \\ =& \int_K (1 + p_1(x)t + \cdots + p_{n+1}(x)t^{n+1})\, dx \\ =& \mathrm{vol}(K) + \Big(\int_K p_1(x)dx\Big)t + \cdots + \Big(\int_K p_{n+1}(x)dx\Big)t^{n+1}.\end{aligned}$$

□

For $r > 0$, we denote by $r \cdot \mathbb{S}^n$ the dilation of $\mathbb{S}^n$ by r, i.e., set of all vectors in $x \in \mathbb{R}^{n+1}$ with $\|x\| = r$.

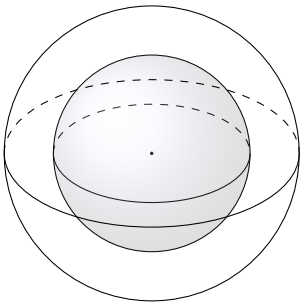


Figure 25: The unit sphere $\mathbb{S}^2$ and the sphere $r \cdot \mathbb{S}^2$ for some $r > 1$.

Lemma 7.5. *Let $\phi : \mathbb{S}^n \to \mathbb{R}^{n+1}$ be a C^1 vector field with $\|\phi(x)\| = 1$ for every $x \in \mathbb{S}^n$. The map*

$$w(x) = \|x\|\phi\Big(\frac{x}{\|x\|}\Big)$$

extends ϕ to a vector field on $S(a,b) := \{x \in \mathbb{R}^{n+1} : a < \|x\| < b\}$ for $0 < a < 1 < b$. Then, for small $t > 0$ and $a < r < b$, the map $\phi_t(x) := x + tw(x)$ maps the sphere $r \cdot \mathbb{S}^n$ onto the sphere $r\sqrt{1+t^2} \cdot \mathbb{S}^n$ bijectively.

Proof. The function $w : S(a,b) \to \mathbb{R}^{n+1}$ is continuously differentiable, since it is the product of the C^1 maps $x \mapsto \phi(\frac{x}{\|x\|})$ and $x \mapsto \|x\|$ (note that $\|x\| \neq 0$ and $\nabla\|x\| = \frac{x}{\|x\|}$ for $x \neq 0$). For $x \in S(a,b)$, $w(x)$ is orthogonal to x. Also, $\|w(x)\| = \|x\|$ by our assumption on ϕ. Thus

$$\begin{aligned}\|\phi_t(x)\|^2 =& \langle x + tw(x), x + tw(x)\rangle \\ =& \|x\|^2 + 2t\langle x, w(x)\rangle + t^2\|w(x)\|^2 \\ =& \|x\|^2(1+t^2),\end{aligned}$$

i.e. $\phi_t(x) = \|x\|\sqrt{1+t^2}$. This implies that $\phi_t(r \cdot \mathbb{S}^n) \subseteq r\sqrt{1+t^2} \cdot \mathbb{S}^n$.
It remains to show that the map $\phi_t : r \cdot \mathbb{S}^n \to r\sqrt{1+t^2} \cdot \mathbb{S}^n$ is surjective. One possible approach relies on the Banach fixed point theorem, and we leave it to the reader as a nice exercise. We shall instead prove surjectivity using connectedness of the sphere. We know that $D\phi_t = I + tDw$ is invertible on $S(a,b)$ for sufficiently small t. From the inverse function theorem, it follows that $\phi_t(S(a,b))$ is an open subset of $\mathbb{R}^{n+1}$. Indeed, for simplicity of notation, denote $U = S(a,b)$, $V = \phi_t(S(a,b))$ and $H := \phi_t^{-1} : V \to U$. We know that H exists, is continuous on V and $\phi_t(U) = H^{-1}(U)$. It is known that the pre-image of an open set under a continuous map is open, and so $\phi_t(S(a,b)) \subset \mathbb{R}^{n+1}$ is open. Thus,

$$\phi_t(r \cdot \mathbb{S}^n) = \phi_t(S(a,b)) \cap r\sqrt{1+t^2} \cdot \mathbb{S}^n$$

is an open subset of $r\sqrt{1+t^2} \cdot \mathbb{S}^n$. But $\phi_t(r \cdot \mathbb{S}^n)$ is also closed, since it is compact as the image of a compact set under a continuous map. The sphere $r\sqrt{1+t^2} \cdot \mathbb{S}^n$ is pathwise connected and therefore connected. Since $\phi_t(r \cdot \mathbb{S}^n)$ is non-empty, we must have[16] $\phi_t(r \cdot \mathbb{S}^n) = r\sqrt{1+t^2} \cdot \mathbb{S}^n$. This establishes surjectivity. □

(3) Concluding the proof
Let n be even and $\phi : \mathbb{S}^n \to \mathbb{R}^{n+1}$ a nowhere vanishing C^1 vector field. Then

[16] In a connected metric space, a non-empty subset which is both closed and open must equal the whole space.

$x \mapsto \|\phi(x)\|$ is a C^1 map on $\mathbb{S}^n$ and $\widetilde{\phi} : \mathbb{S}^n \to \mathbb{R}^{n+1}$, $\widetilde{\phi}(x) = \frac{\phi(x)}{\|\phi(x)\|}$, is a C^1 vector field with norm 1. Thus, by replacing ϕ with $\widetilde{\phi}$, we may assume that $\|\phi(x)\| = 1$ for every $x \in \mathbb{S}^n$.

Let $0 < a < 1 < b < \infty$ and

$$K = K(a, b) := \{x \in \mathbb{R}^{n+1} : a \leq \|x\| \leq b\},$$

which is closed and bounded and therefore compact. Let t be so small that Lemma 7.4 and Lemma 7.5 hold for $F : K \to \mathbb{R}^{n+1}$ given by $F(x) = \|x\|\phi(\frac{x}{\|x\|})$ and $\phi_t(x) = x + tF(x)$. By Lemma 7.5 we have

$$\phi_t(K) = \sqrt{1 + t^2} \cdot K.$$

Then

$$\mathrm{vol}(\phi_t(K)) = \mathrm{vol}(\sqrt{1 + t^2} \cdot K) = (1 + t^2)^{\frac{n+1}{2}} \mathrm{vol}(K).$$

The last identity holds due to the fact that $\mathrm{vol}(\lambda \cdot A) = \lambda^d \mathrm{vol}(A)$ for $\lambda > 0$ and $A \subset \mathbb{R}^d$, where $\lambda \cdot A := \{\lambda x : x \in A\}$.
Now, since n is even, $(1+t^2)^{\frac{n+1}{2}}$ is not a polynomial in t. This can be seen by noting that none of the derivatives of the function $t \mapsto (1 + t^2)^{\frac{n+1}{2}}$ vanishes. However, in view of Lemma 7.4, $\mathrm{vol}(\phi_t(K))$ is a polynomial in t. This is a contradiction, which finishes the proof of the hairy ball theorem.

End of lecture 22. July 9, 2015

8 Ordinary Differential Equations

Let $\phi : [a, b] \to \mathbb{R}^d$, $\phi' : [a, b] \to \mathbb{R}^d$. The expression $\phi' \circ \phi^{-1}$ defines a vector field on the image of the path ϕ. One can interpret a differential equation as, given a vector field $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$, the problem of finding a path ϕ so that on $\mathrm{im}\,\phi$ the vector field F coincides with the tangent vector field $\phi' \circ \phi^{-1}$.

Definition 8.1. A path $\phi : [a, b] \to U$, $\phi \in C^1([a, b]; U)$ is called an integral curve of the vector field $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$ if for all times $x \in [a, b]$ one has $\phi'(x) = F(\phi(x))$.
The integral curve solves the initial value problem with initial value $\phi(x_0) = y_0$ with some given $x_0 \in [a, b]$ and $y_0 \in U$.

Such a construction motivates the definition of a flow of a vector field.

Definition 8.2. Given a vector field $F : U \subset \mathbb{R}^d \to \mathbb{R}^d$ we say $\Phi : [a,b] \times V \subset \mathbb{R} \times U \to U$ is its flow if

$$\begin{aligned} &\Phi(x_0, y_0) = y_0 && \forall y_0 \in V \\ &D_1\Phi(x, y_0) = F\left(\Phi(x, y_0)\right) && \forall x \in [a,b] \end{aligned}$$

We can reformulate the above setting for time dependent vector fields. Let $F : [a,b] \times U \subset \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$. An integral curve of F is a path $\phi : [a,b] \to U$ such that

$$\phi'(x) = F(x, \phi(x)).$$

We call the above equation a *explicit system of ordinary differential equations of first order*. The term differential equation is due to the fact that we are trying to find a path ϕ that satisfies an equality with derivatives falling on the unknown ϕ. The equation is actually a system of equations since we have an equation for each coordinate:

$$\begin{aligned} \phi_1'(x) &= F_1(x, \phi(x)) \\ &\vdots \\ \phi_1'(x) &= F_1(x, \phi(x)). \end{aligned}$$

The differential equation is ordinary (as opposed to being a partial differential equation) is related to the fact that ϕ depends only on one real variable $x \in [a,b]$ thus there exists only one directional derivative. A partial differential equation would involve a function $\tilde{\phi} : \Omega \subset \mathbb{R}^N \to \mathbb{R}^d$ and would be given by an expression involving partial derivatives of $\tilde{\phi}$: $D_1\tilde{\phi}, \ldots, D_N\tilde{\phi}$.
The property of being first order refers the fact that there are only derivatives of first order (there are no second or further derivatives). An n^{th} order equation would be given by

$$\phi^{(n)}(x) = F\left(x,\, \phi(x), \phi^{(2)}(x), \ldots, \phi^{(n-1)}(x)\right)$$

with $\phi : [a,b] \to U$, the derivatives $\phi^{(k)} : [a,b] \to \mathbb{R}^d$ for all $k \in \{1, \ldots, n\}$ and $F : [a,b] \times \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{n \text{ times}} \to \mathbb{R}^d$. A n^{th} order ODE (ordinary differential equation) can be transformed naturally into a first order equation by setting

$$\begin{aligned} &\psi : [a,b] \to U \times \mathbb{R}^{(n-1)d} \\ &\psi(x) := \begin{pmatrix} \phi(x) \\ \phi'(x) \\ \vdots \\ \phi^{(n-1)}(x) \end{pmatrix} \end{aligned}$$

so that ψ satisfies the system of equations

$$\begin{aligned}
\psi_0'(x) &= \psi_1(x) \\
\psi_1'(x) &= \psi_2(x) \\
&\vdots \\
\psi_{n-2}'(x) &= \psi_{n-1}(x) \\
\psi_{n-1}'(x) &= F\left(x, \psi_0(x), \ldots, \psi_{n-1}(x)\right)
\end{aligned}$$

where each $\psi_k : [a,b] \to \mathbb{R}^d$ is the k^{th} component of ψ and it is the k^{th} derivative of ϕ. Notice that the to formulate the corresponding IVP we have to specify initial data $\psi(x_0) = y_0 \in U \times \mathbb{R}^{(n-1)d}$ this means we are specifying the first initial values of ϕ and of its first $n-1$ derivatives.
Finally we say that the ODE is explicit because the equation expresses dependence of the highest order derivative term explicitly in terms of all the remaining ones:

$$\phi'(x) = F\left(x, \phi(x)\right).$$

An non-explicit ODE would have the form

$$G(x, \phi(x), \phi'(x)) = 0$$

where one would require some non-degeneracy of G in its third argument to be able to reasonably expect such a problem to have a solution.
When solving an ODE

$$y'(x) = F\left(x, y(x)\right)$$

it is important to specify in what "set" or space we are looking for solutions. As a first approach it is reasonable to require that $y \in C^1\left([a,b]; \mathbb{R}^d\right)$ so that the expression $y'(x)$ is well defined everywhere on $[a,b]$ and is a continuous function. We also require that F be a continuous functions. For such a setting it is necessary that F also be defined on $[a,b] \times \mathbb{R}^d$ so that $F(x, y(x))$ is necessarily defined for all admissible y and all times $x \in [a,b]$.
As an example we can consider the ODE on $\mathbb{R}^d$ defined by the function $F(x,y) = y$ so that equation becomes

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

Two solutions to this equation are given by $y(x) = \pm e^x$.

Theorem 8.3. *Now let $F : [a,b] \times U \to \mathbb{R}^d$ be a continuous function and let $\phi \in C^1\left([a,b]; U\right)$ be a solution to the initial value problem (IVP)*

$$\begin{cases} \phi'(x) = F(x, \phi(x)) \\ \phi(x_0) = y_0 \end{cases}.$$

Consider also the integral equation (INT)

$$\phi(x) = y_0 + \int_{x_0}^{x} F(t, \phi(t))\, dt \qquad x \in [a, b].$$

A $C^1([a, b]; U)$ function ϕ solves (IVP) if and only if it satisfies (INT).

Proof.

(IVP) $\implies$ (INT) We use the fact that $\phi(x) - \phi(x_0) = \int_{x_0}^{x} \phi'(t)dt$ (or $-\int_{x}^{x_0} \phi(t)dt$ if $x < x_0$) for all $x \in [a, b]$. Substituting (IVP) we have that

$$\begin{aligned} \phi(x_0) &= 0 \\ \phi'(t) &= F(t, \phi(t)) \end{aligned}$$

and this yields the required result.

(INT) $\implies$ (IVP) If (INT) holds then just setting $x = x_0$ we get that $\phi(x_0) = y_0$. Since F is a continuous function on $[a, b] \times U$ and ϕ is also continuous from $[a, b] \to U$ we have that $[a, b] \ni t \to F(t, \phi(t))$ is a continuous function and so we can apply the Fundamental Theorem of Calculus to get that

$$\frac{d}{dx} \int_{x_0}^{x} F(t, \phi(t))dt = F(x, \phi(x)).$$

We can derive the right and left sides of equation (INT) (since ϕ is C^1 and F is continuous) to obtain

$$\phi'(x) = \frac{d}{dx} \int_{x_0}^{x} F(t, \phi(t))dt = F(x, \phi(x))$$

as required.

□

8.1 Picard-Lindelöf theorem

We will now use the above crucial consideration of the equivalence of finding a solution to (IVP) and a function satisfying the integral equation (INT) to state a theorem that allows us to construct a solution via a Banach fixed point argument.

Theorem 8.4 (Picard-Lindelöf)**.** *Let $F : [a,b] \times U \subset \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function and suppose that it is uniformly Lipschitz in the second variable i.e. there exists $L > 0$ such that $\forall y_1, y_2 \in U$ one has $\|F(x,y_1) - F(x,y_2)\| \leq L\|y_1 - y_2\|$ for all times $x \in [a,b]$. The first part of the condition is called Lipschitz continuity with constant $L > 0$ while uniformity is given by the fact that a common $L > 0$ can be chosen for all points $x \in [a,b]$.*
If the above conditions on F hold then for $0 < \epsilon < \frac{1}{L}$ the map $T : C([x_0, x_0+\epsilon]) \to C([x_0, x_0+\epsilon])$ given by

$$T(\phi)(x) = y_0 + \int_{x_0}^{x} F(t, \phi(t))dt$$

is a strict contraction map. The space $C\left([x_0, x_0+\epsilon]; \mathbb{R}^d\right)$ is endowed with its natural $\|\cdot\|_\infty$ norm.

Proof. T is a strict contraction map if

$$\|T(\phi) - T(\psi)\|_\infty = \sup_{x \in [x_0, x_0+\epsilon]} \|T(\phi)(x) - T(\psi)(x)\| < C\|\phi - \psi\|_\infty$$

for some $C < 1$ (notice the strict inequality). Substituting the definition of T we have

$$\|T(\phi) - T(\psi)\|_\infty = \sup_{x \in [x_0, x_0+\epsilon]} \left\| \int_{x_0}^{x} (F(t,\phi(t)) - F(t,\psi(t)))\, dt \right\| \overbrace{\leq}^{\text{total variation bound}}$$

$$\sup_{x \in [x_0, x_0+\epsilon]} \int_{x_0}^{x} \|F(t,\phi(t)) - F(t,\psi(t))\|\, dt \underbrace{\leq}_{\text{Lip. condition}}$$

$$\sup_{x \in [x_0, x_0+\epsilon]} \int_{x_0}^{x} L\|\phi(t) - \psi(t)\|dt \leq \sup_{x \in [x_0, x_0+\epsilon]} \int_{x_0}^{x} L\|\phi - \psi\|_\infty dt \leq \underbrace{L\epsilon}_{<1} \|\phi - \psi\|_\infty$$

and this concludes the proof. □

We can now apply the Banach Fixed Point Theorem to this problem: there exists a unique $\phi \in C([x_0, x_0+\epsilon]; U)$ that is a fixed point of the map T i.e. $T(\phi) = \phi$. We can rewrite the fixed point condition as

$$\phi(x) = T(\phi)(x) = y_0 + \int_{x_0}^{x} F(t, \phi(t))dt$$

so that the fixed point satisfies the integral equation (INT). Furthermore, since ϕ and F are continuous so that $F(t, \phi(t))$ is also continuous we have that $T(\phi) = y_0 + \int_{x_0}^{x} F(t, \phi(t))dt$ is continuously differentiable via the Fundamental Theorem of Calculus, thus $\phi \in C^1([a,b]; U)$ and thus, using the equivalence between (IVP) and (INT) we can conclude that ϕ solves (IVP).

Theorem 8.5 (Picard-Lindelöf)**.** *Let $F : [a,b] \times U \subset \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function that is uniformly Lipschitz in the second variable with constant L. Let $[x_0, x_0+\epsilon] \subset [a,b]$ and let $y_0 \in \mathbb{R}^d$ then there exists a unique map $\phi : [x_0, x_0+\epsilon] \to \mathbb{R}^d$ that solves the IVP*

$$\begin{cases} \phi'(x) = F(x, \phi(x)) \quad \forall x \in [x_0, \epsilon] \\ \phi(x_0) = y_0 \end{cases}$$

Let us study some examples. A particularly important class of ODEs is given by linear homogeneous and inhomogeneous equations:

$$\begin{aligned} y'(x) &= A(x)y(x) && \text{homogeneous} \\ y'(x) &= A(x)y(x) + b(x) && \text{inhomogeneous.} \end{aligned}$$

Both these equations satisfy the Lipschitz condition given reasonable assumptions on A. We have that $F(x,y) = A(x)y + b(x)$ so that

$$\|F(x,y_1) - F(x,y_2)\| = \|A(x)y_1 - A(x)y_2\| \leq \|A(x)\|\|y_1 - y_2\|$$

so we require that $\|A(x)\|_\infty$ be finite.

End of lecture 23. July 13, 2015

Let everything be as in the previous theorem. We consider the IVP with the initial data y_0

$$\begin{cases} \phi'(x) = F(x, \phi(x)) \\ \phi(x_0) = y_0 \end{cases} \tag{23}$$

and another IVP with the initial data y_1

$$\begin{cases} \widetilde{\phi}'(x) = F(x, \widetilde{\phi}(x)) \\ \widetilde{\phi}(x_0) = y_1. \end{cases} \tag{24}$$

We would like to show that if $\|y_0 - y_1\|$ is small, the solutions of the considered initial value problems are close to each other. In other words, we want to show that solutions of (23) depend continuously on the initial data y_0. The reader may make the notions "small" and "close" precise.
Let $T, \widetilde{T}$ be given by

$$\begin{aligned} T\phi(x) &= y_0 + \int_{x_0}^{x} F(t, \phi(t))dt \\ \widetilde{T}\widetilde{\phi}(x) &= y_1 + \int_{x_0}^{x} F(t, \widetilde{\phi}(t))dt. \end{aligned}$$

In the previous lecture we showed that $T, \widetilde{T}$ are strict contractions from $C([x_0, x_0+\varepsilon], \mathbb{R}^d)$ to itself. Denote by ϕ the unique fixed points of T, which is the unique solution of the IVP (23). To obtain the solution of (24) we start the Banach iteration with ϕ. We have that $\|\widetilde{T}\phi - \phi\|$ is small since

$$\|\widetilde{T}\phi - \phi\| = \|\widetilde{T}\phi - T\phi\| = \|y_1 - y_0\|.$$

For $k \in N$ denote $\widetilde{T}^{(k)}\phi := \underbrace{\widetilde{T}(\widetilde{T}(\ldots(\widetilde{T}}_{k-\text{times}}\phi)$. Then, for any $k \in \mathbb{N}$ we have

$$\|\widetilde{T}^{(k)}\phi - \phi\| \leq \sum_{j=1}^{k} \|\widetilde{T}^{(j)}\phi - \widetilde{T}^{(j-1)}\phi\| \tag{25}$$

$$\leq \Big(\sum_{j=1}^{k} q^{j-1}\Big)\|\widetilde{T}\phi - \phi\| \leq C\|\widetilde{T}\phi - \phi\|. \tag{26}$$

By $\widetilde{\phi}$ we denote $\widetilde{\phi} = \lim_{k\to\infty} \widetilde{T}^{(k)}\phi$, which is the unique fixed point of $\widetilde{T}$ given by the Banach iteration. The calculation (25) shows us that

$$\|\widetilde{\phi} - \phi\|$$

is small.

8.2 Cauchy-Kovalevskaya theorem

In this section we discuss a local existence and uniqueness theorem which applies to initial value problems with analytic coefficients. For now we restrict our attention to $d = 1$.

Suppose that $F : [a, b] \times \mathbb{R}^1 \to \mathbb{R}^1$ is $n-$times differentiable and assume we are given a differentiable function ϕ satisfying

$$\phi'(x) = F(x, \phi(x)). \tag{27}$$

We differentiate the right hand-side, which gives

$$D_1F(x, \phi(x)) + D_2F(x, \phi(x))\phi'(x).$$

This implies that the left hand-side of (31) is differentiable and

$$\phi''(x) = D_1F(x, \phi(x)) + D_2F(x, \phi(x))\phi'(x).$$

Since now we know that ϕ' is differentiable, we can derive the right hand-side of the last expression to obtain

$$\begin{aligned} &D_1^2F(x,\phi(x)) + D_2D_1F(x,\phi(x))\phi'(x) \\ &+D_1D_2F(x,\phi(x))\phi'(x) + D_2^2F(x,\phi(x))(\phi'(x))^2 + D_2F(x,\phi(x))\phi''(x) \end{aligned}$$

As above we conclude that the last display equals $\phi'''(x)$. Iterating we see that if F is n-times differentiable, then ϕ is $(n+1)$-times differentiable and its $(n+1)-$th derivative is given by

$$D^n[F(x,\phi(x))] = \sum_{0\leq\alpha_1,\alpha_2,\beta_1,\ldots,\beta_n,\gamma_1,\gamma_n\leq n} C_{\alpha_1,\ldots,\gamma_n} D_1^{\alpha_1}D_2^{\alpha_2}F(x,\phi(x)) \prod_{j=1}^{n}(\phi^{(\beta_j)}(x))^{\gamma_j} \qquad (28)$$

for some non-negative constants $C_{\alpha_1,\alpha_2,\beta_1,\ldots,\beta_n,\gamma_1,\gamma_n}$. One can prove this last fact by induction.

The idea of the existence theorem presented in this chapter is that in order to solve[17]

$$\phi'(x) = F(x,\phi(x)), \quad \phi(0) = 0,$$

we use (28) to determine all the derivatives of the solution ϕ at 0. Then we construct the Taylor series of the solution. To assure convergence of the Taylor series around the origin, it is required that F is real analytic around $(0,0)$.

In Analysis I we have already met real analytic function on $\mathbb{R}$. Recall that a C^∞ function is not necessarily real analytic. An example is the function which equals $e^{-1/x}$ for $x > 0$ and is 0 for $x \leq 0$. The following definition generalizes the definition of real analyticity from Analysis 1 to two dimensions.

Definition 8.6. A function $F : \mathbb{R}^2 \to \mathbb{R}$ is called *real analytic around* $(0,0)$ if there exists $r > 0$ and a_{nm} with $|a_{nm}| < Cr^{-n-m}$, such that for all $|x| < r, |y| < r$ we have

$$F(x,y) = \sum_{n,m=0}^{\infty} a_{nm}x^n y^m. \qquad (29)$$

[17]For simplicity of notation we consider only initial data $\phi(0) = 0$. One proceeds similarly if $\phi(x_0) = y_0$ for $(x_0,y_0) \neq (0,0)$.

The condition $|a_{nm}| < Cr^{-n-m}$ implies that the series (29) converges absolutely as for $|x|, |y| < r$ we have

$$\sum_{n,m=0}^{N} |a_{nm}x^n y^m| \leq C \sum_{n,m=0}^{N} \left|\left(\frac{x}{r}\right)^n \left(\frac{y}{r}\right)^m\right| \leq C \sum_{n=0}^{N} \left|\frac{x}{r}\right|^n \sum_{m=0}^{N} \left|\frac{y}{r}\right|^m \leq C'$$

independently of N.
A two-dimensional analytic[18] function is infinitely differentiable and in particular continuous. We deduce this fact from the one-dimensional theory as follows. For a fixed y, $F(x, y)$ is real analytic around 0. Indeed, since we know that the series for F converges absolutely, we can rewrite it as

$$\sum_n \Big(\sum_m a_{nm} y^m\Big) x^n \leq C \sum_n \Big(\sum_m r^{-n} \Big(\frac{y}{r}\Big)^m\Big) x^n \leq C' \sum_n r^{-n} x^n.$$

The expression $\sum_n r^{-n} x^n$ defines another power series convergent for $|x| < r$ and thus it is real analytic around 0. Therefore, for a fixed y, $F(x, y)$ is infinitely differentiable in a neighbourhood of 0. Its first derivative is given by the first formal derivative of the power series

$$D_1 F(x, y) = \sum_{n,m} a_{nm} n x^{n-1} y^m$$

and similarly for higher derivatives.[19] In particular, all partial derivatives by the first variable are continuous. A similar conclusion can be derived with the roles of x and y interchanged. Thus, for each $n \in \mathbb{N}$, n-th partial derivatives of $F(x, y)$ exist and are continuous. Hence n-th total derivative of F exists and is given by the n-th formal derivative of the power series (29). Moreover, for real analytic functions, (29) equals its Taylor series.

Claim. If F is real analytic around $(0, 0)$ and ϕ is real analytic around 0 with $\phi(0) = 0$, then $F(x, \phi(x))$ is real analytic around 0.

Proof. First note that it suffices to consider the case when the convergence radii of φ and F are 1. Otherwise we rescale, i.e. replace $\phi(x)$ by $\widetilde{\phi}(x) = \phi(rx)$ and $F(x, \phi(x))$ by $\widetilde{F}(x, \widetilde{\phi}(x)) := rF(rx, r\phi(rx))$. Note that

$$F(x, y) \leq \sum_{m,n} C|x^m y^n| \text{ and } \phi(x) \leq \sum_n C'|x^n|.$$

[18] Sometimes we shall shorten "real analytic" to "analytic". We do not discuss complex analyticity here.

[19] Recall that "deriving under the $\sum$ sign" is allowed due to uniform convergence of the series on each $[-\rho, \rho]$ for $0 < \rho < r$. For higher derivatives one needs to note that for any $\varepsilon > 0$ we have $nr^{-n} \leq C_\varepsilon (r - \varepsilon)^{-n}$, so $D_1 F(x, y)$ is another real analytic function at 0.

This gives a hint that one should be able to rather work with the functions

$$G(x,y) := \sum_{m,n=0}^{\infty} Cx^m y^n \text{ and } \psi(x) := \sum_{n=1}^{\infty} C' x^n.$$

They are real analytic and easy to handle as they can be explicitly computed. The idea is now to compare the functions F and ϕ with G and ψ, respectively, and use analyticity of the composition $G(x,\psi(x))$ to conclude the claim. Observe that $\psi(0) = 0$ and

$$G(x,y) = C\frac{1}{1-x}\frac{1}{1-y}, \quad \psi(x) = C'\Big(\frac{1}{1-x} - 1\Big).$$

The composition

$$G(x,\psi(x)) = C\frac{1}{1-x}\frac{1}{1-C'(\frac{1}{1-x}-1)}$$

is a rational function without a pole at 0 and is thus real analytic around 0. Write $f(x) := F(x,\phi(x))$ and $g(x) := G(x,\psi(x))$. To deduce analyticity of f at 0 it suffices to show that for each $x \in [-1/2, 1/2]$ and all $n \in \mathbb{N}$ [20]

$$|D^n f(x)| \leq D^n g(x).$$

If this is true, then real analyticity of g around 0 implies real analyticity of f around zero.
We show the required bound for $x = 0$, the general case is left to the reader. Since $F(x,y)$ equals its Taylor series around $(0,0)$, we have

$$|D^{\alpha_1} D^{\alpha_2} F(0,0)| = |a_{nm}|\alpha_1!\alpha_2! \leq C\alpha_1!\alpha_2! = D^{\alpha_1} D^{\alpha_2} G(0,0).$$

Note that $D^{\alpha_1} D^{\alpha_2} G(0,0) \geq 0$. Similarly one sees that $|\phi^{(\beta_j)}(0)| \leq \psi^{(\beta_j)}(0)$ and $\psi^{(\beta_j)}(0) \geq 0$. We use (28) and estimate

$$\begin{aligned}
|D^n f(0)| &\leq \sum_{\alpha_1,\ldots,\gamma_n} \underbrace{C_{\alpha_1,\ldots,\gamma_n}}_{\geq 0} |D_1^{\alpha_1} D_2^{\alpha_2} F(0,0) \prod_{j=1}^{n} (\phi^{(\beta_j)}(0))^{\gamma_j}| \\
&\leq \sum_{\alpha_1,\ldots,\gamma_n} C_{\alpha_1,\ldots,\gamma_n} D_1^{\alpha_1} D_2^{\alpha_2} G(0,0)) \prod_{j=1}^{n} (\psi^{(\beta_j)}(0))^{\gamma_j} \\
&= D^n g(0).
\end{aligned}$$

[20] We keep the notation $D^n f$ although f is now a one-dimensional function, so that we remember that it equals $D^n[F(x,\phi(x))]$.

To deduce the last line it was crucial that all the quantities are non-negative, so that it was allowed to omit the absolute values. This is the required estimate at $x = 0$. □

Now we return to the initial value problem

$$\begin{cases} \phi'(x) = F(x, \phi(x)) \\ \phi(0) = 0 \end{cases}$$

for F analytic around $(0, 0)$. We are set to construct a local analytic solution to this problem. Inductively define the sequence ϕ_n by

$$\begin{aligned} \phi_0 &:= 0 \\ \phi_{n+1} &:= \sum_{\alpha_1,\ldots,\gamma_n} C_{\alpha_1,\ldots,\gamma_n} C D_1^{\alpha_1} D_2^{\alpha_2} F(0,0) \prod_{j=1}^{n} (\phi_{\beta_j})^{\gamma_j}. \end{aligned} \tag{30}$$

Now we again consider the function

$$G(x, y) = \sum_{n,m=1}^{\infty} C x^n y^m = C \frac{1}{1-x} \frac{1}{1-y}$$

and the IVP

$$\begin{cases} \psi'(x) = G(x, \phi(x)) \\ \psi(0) = 0. \end{cases}$$

This last IVP can easily be solved explicitly. If we write $y := y(x) := \psi(x)$, then we can rewrite the ODE as

$$y' = C \frac{1}{1-x} \frac{1}{1-y}. \tag{31}$$

Observe the simple form of this equation: the variables x and y on the right hand-side "split". We can solve this ODE by the so-called *separation of variables*. Multiplying the equation by $(1 - y)$ yields

$$y'(1-y) = C \frac{1}{1-x}$$

The left hand-side of the last display equals $-\frac{1}{2}((1-y)^2 - 1)'$. Integrating in x and using $y(0) = 0$ we obtain for x near 0

$$\begin{aligned} -\frac{1}{2}((1-y)^2 - 1) &= -C \log(1-x) \\ \Leftrightarrow\ \psi(x) = y &= 1 - \sqrt{1 - 2C \log(1-x)}. \end{aligned}$$

Note that all the derivatives of ψ at 0 are positive, which can be seen by induction using (30). Inductively we also see that for each $n \in \mathbb{N}$

$$|\phi_n| \leq \psi^{(n)}(0).$$

The function ψ is real analytic around zero. Therefore, the series

$$\phi(x) := \sum_{n=0}^{\infty} \frac{1}{n!} \phi_n x^n$$

converges in a positive radius around zero. By construction,

$$\phi'(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \phi_{n+1} x^n = \sum_{n=0}^{\infty} \frac{1}{n!} D^n f(0) x^n = F(x, \phi(x)).$$

The last equality follows by our claim as $f(x) = F(x, \phi(x))$ is real analytic. Thus, ϕ is a solution to the IVP.
We summarize this discussion in the following theorem, which is a special case of a more general theorem by Cauchy and Kovalevskaja.

Theorem 8.7 (Cauchy-Kovalevskaya)**.** *Let $F : [-a, a] \times \mathbb{R}^d \to \mathbb{R}^d$ be real analytic around $(0, 0)$. Then there exists $\varepsilon > 0$ and a unique real analytic function $\phi : [-\varepsilon, \varepsilon] \to \mathbb{R}^d$ with*

$$\begin{aligned} D\phi(x) &= F(x, \phi(x)) \\ \phi(0) &= (0, \ldots, 0) \end{aligned} \tag{32}$$

for all $x \in [-\varepsilon, \varepsilon]$.

Existence for $d = 1$ has just been shown. The IVP (32) uniquely determines derivatives of all orders at 0 of the solution and hence the Taylor series of the solution. This implies uniqueness.

We briefly comment on the case $d > 1$. Analiticity in $\mathbb{R}^{d+1}$ is defined analogously as in $\mathbb{R}^2$. In (28), the products need to be replaced by matrix products. Instead of 31 one now considers the system of differential equations

$$y_j' = G_j(x, y_1, \ldots, y_d) = C \frac{1}{1-x} \frac{1}{1-y_1} \cdots \frac{1}{1-y_d}, \quad y_j(0) = 0$$

for $j = 1, \ldots, d$. Since the system is completely symmetric with respect to $y_1, \ldots, y_d$, we make the following ansatz for the solution

$$y_1 = y_2 = \cdots = y_d =: y.$$

This reduces the system to only one equation

$$y' = C\frac{1}{1-x}\frac{1}{1-y^d}, \quad y(0) = 0,$$

which can again be solved by a separation of variables. Its solution is

$$y = 1 - \sqrt[d+1]{1 - C\log(1-x)}.$$

The rest of the proof proceeds analogously as in the case $d = 1$.

End of lecture 24. July 16, 2015