Analysis II (notes)

Abhik Pal
ab.pal@jacobs-university.de

Spring 2018, Jacobs University Bremen.

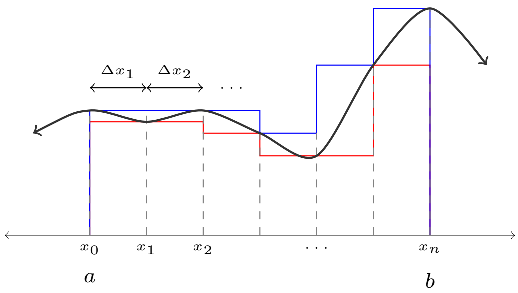
Contents

Chapter 1

Riemann–Stieltjes Integral in $\mathbb{R}$

1.1 The Riemann Integral in $\mathbb{R}$

Integration Integration gives us the "area under the curve." And to actually compute this area for some function $f : [a, b] \to \mathbb{R}$ in the interval $[a, b]$, we divide the interval into smaller sections and approximate the area by taking the sum of the rectangles introduced by these sections:



This approximation gives us the upper Riemann sum (in blue) by taking the $\sup_{x\in\Delta x} f(x)$ for each Δx and the lower Riemann sum (in red) of rectangles by taking the $\inf_{x\in\Delta x} f(x)$ for each Δx.

However, for some functions, this approximation may not be very "nice" and the upper and lower sums might end up disagreeing. So we say that the Riemann integral of a function exists only if the upper Riemann sum and the lower Riemann sum coincide.

Definition 1.1.1 (Partition)**.** For an interval $[a, b] \subset \mathbb{R}$, a *partition* $\mathcal{P}$ is a finite set of points $x_0, x_1, \ldots, x_n$ with $a = x_0 < x_1 < \cdots < x_n = b$. We also set $\Delta x_i = x_i - x_{i-1}$ for $i = 1, \ldots, n$.

We can now define the integral of a bounded function $f : [a, b] \to \mathbb{R}$. Since the function is bounded, we know that $\exists m, M \in \mathbb{R}$ such that $\forall x \in [a, b], m \leq f(x) \leq M$. For the i-th interval in some partition, we can define

$$m_i = \inf_{x\in[x_{i-1},x_i]} f(x), \qquad M_i = \sup_{x\in[x_{i-1},x_i]} f(x).$$

Definition 1.1.2 (Riemann Sums)**.** For some function $f : [a, b] \to \mathbb{R}$ the *upper Riemann sum* $U(f, \mathcal{P})$ on some partition $\mathcal{P}$ of $[a, b]$ is defined as

$$U(f, \mathcal{P}) := \sum_{i=1}^{n} M_i \Delta x_i.$$

Similarly we define the *lower Riemann sum* $L(f, \mathcal{P})$ as

$$L(f, \mathcal{P}) := \sum_{i=1}^{n} m_i \Delta x_i.$$

Definition 1.1.3 (Riemann integrals)**.** We define the *upper Riemann integral* for $f : [a, b] \mapsto \mathbb{R}$ on some partition $\mathcal{P}$ of $[a, b]$ as

$$\overline{\int_a^b} f(x)\,\mathrm{d}x := \inf_{\mathcal{P}} U(f, \mathcal{P}),$$

and the *lower Riemann Integral* as

$$\underline{\int_a^b} f(x)\,\mathrm{d}x := \sup_{\mathcal{P}} L(f, \mathcal{P}).$$

For the upper Riemann integral, the general idea is to take the infimum of all possible upper sums over all possible partitions. Similar for the lower Riemann integral, we consider the supremum of the lower Riemann sums.

Definition 1.1.4 (Riemann integrable function)**.** A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if the lower and the upper Riemann sums coincide and we write:

$$\int_a^b f(x)\,\mathrm{d}x := \overline{\int_a^b} f(x)\,\mathrm{d}x = \underline{\int_a^b} f(x)\,\mathrm{d}x \left(= \int_a^b f\,\mathrm{d}x \right).$$

We also define $\mathcal{R}[a, b]$ as the set of all Riemann integrable functions on $[a, b]$.

1.2 The Riemann-Stieltjes Integral

The Riemann integral uniformly assigns "weights" to different parts of the function to compute their contribution to the final value of the integral. Generalizing this notion of "weight" gives us the Riemann–Stieltjes Integral (or simply, the Stieltjes integral).

Let $\alpha : [a, b] \to \mathbb{R}$ be a monotonically increasing (and thus bounded) function. We now define, for a given partition $\mathcal{P}$, $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$ for $i = 1, \ldots, n$. From monotonicity, it immediately follows that $\Delta\alpha_i \geq 0$. We now redefine the upper and lower Riemann sums to account for this α.

Note. α need not be continuous. Often we would deliberately make α discontinuous to get some desired result (see Homework B.1.2).

Definition 1.2.1 (Upper/Lower Riemann–Stieltjes sums)**.** We define the upper and lower Riemann sums respectively as

$$U(f, \mathcal{P}, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i \text{ and } L(f, \mathcal{P}, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i.$$

Definition 1.2.2 (Upper/Lower Riemann–Stieltjes integrals)**.** The upper and lower Riemann-Stieltjes integrals for $f : [a,b] \to \mathbb{R}$ on some partition $\mathcal{P}$ of $[a,b]$ are respectively given by

$$\overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) := \inf_{\mathcal{P}} U(f,\mathcal{P},\alpha) \text{ and } \underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) := \sup_{\mathcal{P}} L(f,\mathcal{P},\alpha).$$

When the upper and the lower Riemann–Stieltjes sums coincide, we get the Riemann–Stieltjes integral for that particular function.

Definition 1.2.3 (Riemann–Stieltjes Integrable Functions)**.** A function $f : [a,b] \to \mathbb{R}$ is Riemann–Stieltjes integrable with respect to α when the upper and the lower Riemann–Stieltjes sums coincide. That is

$$\int_a^b f(x)\,\mathrm{d}\alpha := \overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) = \underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) \left(= \int_a^b f\,\mathrm{d}\alpha(x) \right).$$

We also define $\mathcal{R}(\alpha)[a,b]$ as the set of all Riemann–Stieltjes integrable functions with respect to α.

Note (Riemann integral). The usual Riemann integral is a special case of the Riemann–Stieltjes integral with $\alpha(x) = x$.

Note (Differentiable α). When $\alpha(x)$ is differentiable, then the integral is "weighted" by $\alpha'(x)$. That is

$$\int f\,\mathrm{d}\alpha = \int f \frac{\mathrm{d}\alpha}{\mathrm{d}x}\,\mathrm{d}x = \int f\alpha'\,\mathrm{d}x.$$

That will be made precise in Section 1.8.

Note (Step function α). Consider

$$\alpha(x) = \begin{cases} 0, \text{ if } x > \xi \\ 1, \text{ if } x \leq \xi \end{cases}.$$

Here $\Delta\alpha_i = 0$ unless $\xi \in (x_{i-1}, x_i]$. Therefore for this fixed i, $U(f,\mathcal{P},\alpha) = M_i$ and $L(f,\mathcal{P},\alpha) = m_i$. Then on $[x_{i-1}, x_i], m_i \leq f(x) \leq M_i$. If, additionally, f is continuous at ξ, we can show that

$$\overline{\int_a^b} f(x)\,\mathrm{d}\alpha = \underline{\int_a^b} f(x)\,\mathrm{d}\alpha = \int_a^b f\,\mathrm{d}\alpha = f(\xi),$$

see Homework B.1.1.

1.3 Refinement of Partitions

Definition 1.3.1 (Refinement)**.** A partition $\mathcal{P}^*$ is *finer* than $\mathcal{P}$ if $\mathcal{P} \subset \mathcal{P}^*$.

Definition 1.3.2 (Common Refinement)**.** We define the *common refinement* $\mathcal{P}^*$ of $\mathcal{P}_1$ and $\mathcal{P}_2$ as $\mathcal{P}^* := \mathcal{P}_1 \cup \mathcal{P}_2$.

Lemma 1.3.1. *If $\mathcal{P}^*$ is a refinement of $\mathcal{P}$, then $U(f,\mathcal{P}^*,\alpha) \leq U(f,\mathcal{P},\alpha)$ and $L(f,\mathcal{P}^*,\alpha) \geq L(f,\mathcal{P},\alpha)$.*

Proof. Partitions are by definition finite, so we only consider the case when $\mathcal{P}^* = \mathcal{P} \cup \{x^*\}$. Then for this point $x^* \in [x_{i-1}, x_i]$ we have: $M_i = \sup_{x\in[x_{i-1},x_i]} f(x)$, $M_i^* = \sup_{x\in[x_{i-1},x^*]} f(x)$, $M_i^{**} = \sup_{x\in[x^*,x_i]} f(x)$.

Now, we get

$$U(f, \mathcal{P}^*, \alpha) = M_i^* \left(\alpha(x^*) - \alpha(x_{i-1})\right) + M_i^{**} \left(\alpha(x_i) - \alpha(x^*)\right) + U^*_{\mathcal{P}^*},$$

where $U^*_{\mathcal{P}^*}$ is the sum unaffected by the refinement. Since $M_i^*, M_i^{**} \leq M_i$,

$$U(f, \mathcal{P}^*, \alpha) \leq M_i \Delta\alpha_i + U^*_{\mathcal{P}^*} = U(f, \mathcal{P}, \alpha)$$

The proof for the lower sum is analogous and uses the the infima m_i. □

Corollary.

$$\underline{\int_a^b} f \,\mathrm{d}\alpha \leq \overline{\int_a^b} f \,\mathrm{d}\alpha$$

Proof. Given $\mathcal{P}_1$ and $\mathcal{P}_2$ we take $\mathcal{P}^* = \mathcal{P}_1 \cup \mathcal{P}_2$ the common refinement. From Lemma 1.3.1 we get [rest of proof missing]. □

1.4 Existence criterion

Theorem 1.4.1. *A function $f \in \mathcal{R}(\alpha)[a, b] \iff$ for all $\varepsilon > 0$, there is some partition $\mathcal{P}$ such that $0 \leq U(f, \mathcal{P}, \alpha) - L(f, \mathcal{P}, \alpha) < \varepsilon$.*

Proof. We prove sufficiency (⇐) and necessity (⇒) separately.

"$\Longleftarrow$" We know that for all $\mathcal{P}$,

$$L(f, \mathcal{P}, \alpha) \leq \underline{\int_a^b} f \,\mathrm{d}\alpha \leq \overline{\int_a^b} f \,\mathrm{d}\alpha \leq U(f, \mathcal{P}, \alpha).$$

Then,

$$0 \leq U(f, \mathcal{P}, \alpha) - L(f, \mathcal{P}, \alpha) \leq \varepsilon \implies 0 \leq \overline{\int_a^b} f \,\mathrm{d}\alpha - \underline{\int_a^b} d \,\mathrm{d}\alpha \leq \varepsilon.$$

Since this holds for all $\varepsilon > 0$, $\overline{\int_a^b} f \,\mathrm{d}\alpha = \underline{\int_a^b} f \,\mathrm{d}\alpha \implies f \in \mathcal{R}(\alpha)[a, b]$.

"$\Longrightarrow$" For some $\varepsilon > 0$ there are $\mathcal{P}_1$ and $\mathcal{P}_2$ such that

$$0 \leq U(f, \mathcal{P}_1, \alpha) - \overline{\int_a^b} f \,\mathrm{d}\alpha \leq \frac{\varepsilon}{2}, \quad 0 \leq L(f, \mathcal{P}_2, \alpha) - \underline{\int_a^b} f \,\mathrm{d}\alpha \leq \frac{\varepsilon}{2}.$$

Since $f \in \mathcal{R}(\alpha)[a, b]$, $\int_a^b f \,\mathrm{d}\alpha = \overline{\int_a^b} f \,\mathrm{d}\alpha = \underline{\int_a^b} f \,\mathrm{d}\alpha$. Taking the common refinement $\mathcal{P}^* = \mathcal{P}_1 \cup \mathcal{P}_2$, we get:

$$\begin{aligned} U(f, \mathcal{P}^*, \alpha) &\leq U(f, \mathcal{P}_1, \alpha) \\ &\leq \frac{\varepsilon}{2} + \int_a^b f \,\mathrm{d}\alpha \\ &\leq L(f, \mathcal{P}_2, \alpha) + \varepsilon \leq L(f, \mathcal{P}^*, \alpha) + \varepsilon \end{aligned}$$

This implies $0 \leq U(f, \mathcal{P}^*, \alpha) - L(f, \mathcal{P}^*, \alpha) \leq \varepsilon$.

□

1.5 Continuous functions

Theorem 1.5.1. *If $f : [a, b] \to \mathbb{R}$ is continuous then $f \in \mathcal{R}(\alpha)[a, b]$ for all monotonically increasing $\alpha(x)$.*

Proof. We fix some $\varepsilon > 0, \eta > 0$ such that

$$[\alpha(b) - \alpha(b)]\, \eta < \varepsilon.$$

Since f is uniformly continuous (see Homework B.1.5), there exists some $\delta > 0$ such that for $x, y \in [a, b]$

$$|x - y| < \delta \implies |f(x) - f(y)| < \eta.$$

We now choose our partition such that $\Delta x_i < \delta$ for all i. This implies in particular that $0 \leq M_i - m_i \leq \eta$ for $i = 1, 2, \ldots, n$, and thus

$$\begin{aligned} 0 \leq U(f, \mathcal{P}, \alpha) - L(f, \mathcal{P}, \alpha) &= \sum_{i=1}^{n} (M_i - m_i) \Delta\alpha_i \\ &\leq \eta \sum_{i=1}^{n} \Delta\alpha_i = \eta\big(\alpha(b) - \alpha(a)\big) < \varepsilon, \end{aligned}$$

so the statement follows from Lemma 1.4.1. □

Theorem 1.5.2. *A bounded function $f : [a, b] \to \mathbb{R}$ with finitely many points of discontinuity $e_1, e_2, \ldots, e_k$ such that α continuous for all e_j is Riemann integrable with respect to α i.e., $f \in \mathcal{R}(\alpha)[a, b]$.*

Note. In particular, $f \in \mathcal{R}[a, b]$.

Note. The theorem also holds for countably infinite many discontinuities.

Proof. [ToDo] □

1.6 Basic properties of integrals

Theorem 1.6.1.
1. *(Linearity) $f_1, f_2 \in \mathcal{R}(\alpha)[a, b] \implies (f_1 + f_2) \in \mathcal{R}(\alpha)[a, b]$ and $\int_a^b (f_1 + f_2)\, \mathrm{d}\alpha = \int_a^b f_1\, \mathrm{d}\alpha + \int_a^b f_2\, \mathrm{d}\alpha$. Further, if $\lambda \in \mathbb{R}$ then $\lambda f_1 \in \mathcal{R}(\alpha)[a, b]$ with $\int_a^b \lambda f_1\, \mathrm{d}\alpha = \lambda \int_a^b f_1\, \mathrm{d}\alpha$.*
2. *(Monotonicity) If $f_1(x) \leq f_2(x)$ for all x then $\int f_1\, \mathrm{d}\alpha \leq \int f_2\, \mathrm{d}\alpha$.*
3. *(Additivity) If $f \in \mathcal{R}(\alpha)[a, b]$ and $c \in (a, b)$ then $f \in \mathcal{R}(\alpha)[a, c]$ and $f \in \mathcal{R}(\alpha)[c, b]$ with $\int_a^c f\, \mathrm{d}\alpha + \int_c^b f\, \mathrm{d}\alpha = \int_a^b f\, \mathrm{d}\alpha$.*
4. *(Standard Estimate) If $f \in \mathcal{R}(\alpha)[a, b]$ and for all $x \in [a, b], f(x) \leq M$, then, $\int_a^b f\, \mathrm{d}\alpha \leq M(\alpha(b) - \alpha(a))$. The same holds in the other direction with some $m \leq f(x)$.*
5. *(Linearity of α) If $f \in \mathcal{R}(\alpha_1)[a, b] \cup \mathcal{R}(\alpha_2)[a, b]$ and $\lambda_1, \lambda_2 \geq 0$, then $\int_a^b f\, \mathrm{d}(\lambda_1\alpha_1 + \lambda_2\alpha_2) = \lambda_1 \int_a^b f\, \mathrm{d}\alpha_1 + \lambda_2 \int_a^b f\, \mathrm{d}\alpha_2$*

Proof. See homework B.2.1. □

1.7 Compositions

1.8 Stieltjes integrals as change of variables

1.9 Change of variables

1.10 Integration and differentiation as inverse operations

1.11 Improper integrals

1.12 Uniform convergence

Chapter 2

Convergence of Series

2.1 Convergence tests

Definition 2.1.1 (Convergence)**.** Consider $(a_k)_{k\in\mathbb{N}}$ a sequence in $\mathbb{C}$ (i.e., each $a_k \in \mathbb{C}$). We say that (a_k) *converges* to a or $a_k \xrightarrow{k\to\infty} a$ if and only if for all $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that for all $n \geq N, |a_n - a| < \varepsilon$. We say that (a_k) *diverges* if it doesn't converge.

Definition 2.1.2 (Cauchy sequence)**.** A sequence $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence if and only if for all $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that for all $n > N, m > N$, $|a_n - a_m| < \varepsilon$.

Note. If (a_k) converges then (a_k) is a Cauchy sequence.

Note. In a complete metric space (e.g., $\mathbb{R}, \mathbb{R}^n, \mathbb{C}$, etc.) all Cauchy sequences converge.

We can generalize the ideas about sequences to series as for some $\sum_{k=0}^{\infty} a_k$, the partial sums $s_n = \sum_{k=0}^{n} a_k$ form a sequence. If the s_n converge to some s, we write

$$\sum_{k=0}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=0}^{n} a_k = \lim_{n\to\infty} s_n = s.$$

Definition 2.1.3 (Absolute convergence)**.** If for some $A = \sum_{k=0}^{\infty} a_k$, the series $\sum_{k=0}^{\infty} |a_k|$ converges, then A converges *absolutely.*

Definition 2.1.4 (Conditional convergence)**.** If $A = \sum_{k=0}^{\infty} a_k$ converges but $\sum_{k=0}^{\infty} |a_k|$ does not, then we say that A converges *conditionally.*

Note. $\sum a_k$ converges if and only if for all $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that for all $m \geq n > N$, $|\sum_{k=n}^{m} a_k| < \varepsilon$ (via Cauchy). The converse, however is not true. Consider with $m = n$ the sum $\sum_{k=0}^{\infty} \frac{1}{k}$. For large k, $\frac{1}{k} \to 0$ but the sum $\sum_{k=0}^{\infty} \frac{1}{k}$ does not converge.

Theorem 2.1.1 (Comparison Test)**.** *Fix some $N \in \mathbb{N}$ then*

1. *If $|a_k| < b_k$ for all $k \geq N$ and if $\sum b_k$ converges, then $\sum |a_k|$ converges as well.*
2. *If $a_k \geq c_k \geq 0$ for all $k \geq N$ and if $\sum c_k$ diverges, then $\sum a_k$ diverges as well.*

Proof. Clear. □

Recall that the *finite geometric series* $\sum_{k=0}^{n} x^k$ equals $\frac{x^{n+1}-1}{x-1}$. Additionally, when $0 \leq x < 1$, the *infinite geometric series* $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ (with $0^0 := 1$). Also note that

$$\limsup_{k\to\infty} a_k := \lim_{k\to\infty} \sup_{n\geq k} a_n, \qquad \liminf_{k\to\infty} a_k := \lim_{k\to\infty} \inf_{n\geq k} a_n.$$

Theorem 2.1.2 (Root test)**.** *For some sequence $(a_k)_{k\in\mathbb{N}}$, let $L = \limsup_{k\to\infty} \sqrt[k]{|a_k|}$.*

1. *If $L < 1$, then $\sum a_k$ converges.*
2. *If $L > 1$, then $\sum a_k$ diverges.*
3. *If $L = 1$, then the root test is inconclusive.*

Proof. We prove each case separately.

1. For $L < 1$ choose M such that $L < M < 1$ and $N \in \mathbb{N}$ large enough such that for all $k \geq N$, $\sqrt[k]{|a_k|} < M \implies \forall k \geq N, a_k \leq |a_k| < M^k$. As $M < 1$, we know that $\sum M^k$ converges. Hence, by the comparison test (2.1.1), $\sum a_k$ converges as well.
2. For $L > 1$, for all $\varepsilon > 0$ there are infinitely many a_k with $\sqrt[k]{|a_k|} > L - \varepsilon$. Now choose ε small enough such that $L - \varepsilon > 1 \implies$ infinitely many $|a_k| > 1$, so the a_k don't converge to 0. Hence $\sum a_k$ can't converge.
3. For $L = 1$ we provide some counter examples:
   a) $a_k = 1 \implies L = 1$ but the sum diverges.
   b) $a_k = (-1)^k/k$, then
   $$\lim_{k\to\infty} \left(\frac{1}{k}\right)^{1/k} = \lim_{x\to 0} x^x = \lim_{x\to 0} e^{x \ln x} = e^0 = 1 \implies L = 1.$$
   However, the sum converges (same applies to $a_k = \frac{1}{k^2}$).

□

Theorem 2.1.3 (Ratio test)**.** *For $(a_k)_{k\in\mathbb{N}}$,*

1. $\limsup_{n\to\infty} \left|\frac{a_{k+1}}{a_k}\right| < 1 \implies \sum_0^\infty |a_k|$ *converges.*
2. *If there exists some $N \in \mathbb{N}$ such that for all $k \geq N$, $\left|\frac{a_{k+1}}{a_k}\right| \geq 1 \implies \sum_0^\infty a_k$ diverges.*

Proof. We prove each part separately.

1. As in the proof the the root test (2.1.2) there is some $M \in \mathbb{R}$ and $N \in \mathbb{N}$ such that for all $k \geq N$,

$$\begin{aligned}\left|\frac{a_{k+1}}{a_k}\right| < M < 1 &\implies |a_{N+1}| < M|a_N| \\ &\implies |a_{N+2}| \leq M|a_{N+1}| \leq M^2|a_N| \\ &\implies \vdots \\ &\implies |a_{N+p}| \leq M^p|a_N| \\ &\implies |a_k| \leq M^{k-N}|a_N|, \forall k \geq N.\end{aligned}$$

The claim follows by the comparison test (2.1.1) and convergence of the geometric series.

2. If there is some N such that for all $k \geq N \in \mathbb{N}$, $|a_{k_1}| \geq |a_k|$, then a_k do not converge to 0. Hence $\sum a_k$ can't converge.

□

Note. The root test (2.1.2) is "better" than the ratio rest (2.1.3) as it looks at the behavior of actual terms in the sequence and *not* the relative behavior of successive terms (think of a sequence where every other term is zero, so we can't even apply the ration test). Hence if the ratio test implies convergence then the root test will imply convergence as well, however, the converse of this is false. That said, the ratio test is significantly easier to apply.

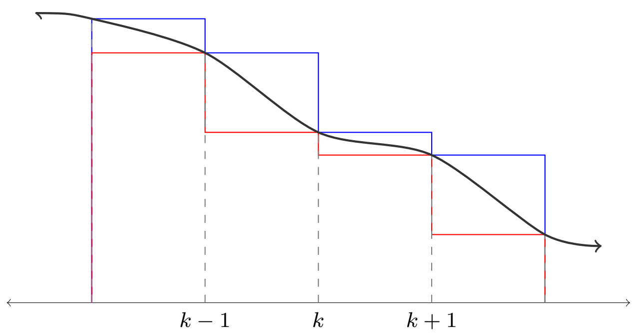
Theorem 2.1.4 (Liebniz condition)**.** *Let $a_k \in \mathbb{R}$ with $a_k \geq 0$ and a_k monotonically decreasing such that $\lim_{k\to\infty} a_k = 0$. Then $\sum_k^\infty (-1)^k a_k$ converges.*

Proof. (advanced calculus) □

Theorem 2.1.5 (Integral test)**.** *Let $f : \mathbb{R}_0^+ \mapsto \mathbb{R}_0^+$ be monotonically decreasing and Riemann integrable with $f \in \mathcal{R}[0, b]$ for all $b \in \mathbb{R}_0^+$. Then,*

$$\int_0^\infty f(x)\,\mathrm{d}x \text{ exists} \iff \sum_{k=0}^\infty f(k) \text{ converges.}$$

Proof. (mainly from an advanced calculus homework)



We see that

$$f(k+1) \leq \int_k^{k+1} f(k)\,\mathrm{d}k \leq f(k)$$

$$\implies \int_1^{N+1} f(k)\,\mathrm{d}k \leq \sum_{k=1}^{N} f(x) \leq \int_0^{N} f(k)\,\mathrm{d}k \leq \sum_{k=0}^{N-1} f(k).$$

From here, taking $\lim_{N\to\infty}$ gives us convergence of series iff convergence of integral. $\square$

Example 2.1.1. $\sum_{k=1}^{\infty} \left(\frac{1}{k}\right)^p$ converges for $p > 1$.

Example 2.1.2. $\sum_{k=0}^{\infty} 1/(k \ln k)$ diverges as $\int \frac{1}{k \ln k}\,\mathrm{d}k = \ln \ln k$ diverges.

Example 2.1.3. $\sum_{k=2}^{\infty} 1/(k \ln^p k)$ converges for $p > 1$.

2.2 Rearrangements

Definition 2.2.1 (Rearrangement)**.** If $\sum_0^\infty a_n$ is an infinite sum and there is some bijection $k : \mathbb{N} \mapsto \mathbb{N}$ (the sequence $(k_n)_{n\in\mathbb{N}}$ includes every non-negative integer), then we call $\sum_0^\infty a_{k(n)}$ a *rearrangement* of $\sum_0^\infty a_n$.

Recall. Conditional and absolute convergence (see beginning of section 2.1).

Theorem 2.2.1. *If $\sum a_i$ for $a_i \in \mathbb{C}$ converges* absolutely*, then all rearrangements also converge and have the same limit.*

Proof. Since $\sum a_i$ converges absolutely, we have that for all $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ such that for all $m \geq n \geq N$, $\sum_n^m |a_i| < \varepsilon$.

Now choose M large enough such that $1, 2, 3, \ldots, N$ are all included in rearrangement $k(1), k(2), k(3), \ldots, k(N)$. For $n > M$ and some large enough m' (with $m' = \sup_{1\leq i\leq m} k_i$),

$$\left| \sum_1^n a_i - \sum_1^n a_{k(i)} \right| \leq \sum_{i=N+1}^{m'} |a_i| < \varepsilon.$$

This implies that $\sum_i a_{k(i)}$ converges to the same limit as $\sum a_i$. $\square$

Theorem 2.2.2. *If $\sum a_n$ for $a_n \in \mathbb{C}$ converges* conditionally*, then for all $A \in \mathbb{R} \cup \{\pm\infty\}$ there exists a bijection $k : \mathbb{N} \mapsto \mathbb{N}$ such that $\sum_0^\infty a_{k(n)} = A$.*

Proof. The proof is a bit involved; we don't prove it here. $\square$

Example 2.2.1 ("Magic trick" from Adv. Calc)**.** $\sum \frac{(-1)^n}{n+1}$ conditionally converges to $\ln 2$ (why?). However, we can rearrange it so that it converges to

$\frac{1}{2}\ln 2$:

$$\begin{aligned}\sum \frac{(-1)^n}{n+1} &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \dots \\ &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{4} \pm \dots \\ &= \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} \pm \dots \\ &= \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} \pm \dots\right) \\ &= \frac{1}{2}\ln 2.\end{aligned}$$

2.3 Cauchy Product

For products of infinite sums, we have different "strategies" for multiplying them. For instance:

$$\begin{aligned}\left(\sum^{\infty} a_k\right)\left(\sum^{\infty} b_k\right) &= a_0 \sum^{\infty} b_k + a_1 \sum^{\infty} b_k + \dots \\ &\overset{\text{or}}{=} a_0 b_0 + (a_1 b_0 + a_0 b_1) + (a_2 b_0 + a_1 b_1 + a_0 b_2) + \dots\end{aligned}$$

Notice that the second one becomes particularly nice when we're dealing with power series (§ 2.4) as

$$\begin{aligned}\left(\sum^{\infty} a_k z^k\right)\left(\sum^{\infty} b_l z^l\right) &= a_0 b_0 + (a_1 b_0 + a_0 b_1) z + (a_2 b_0 + a_1 b_1 + a_0 b_2) z^2 + \dots \\ &= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right) z^n\end{aligned}$$

Definition 2.3.1 (Cauchy product)**.** The Cauchy product of two sums $\sum^{\infty} a_k$ and $\sum^{\infty} b_k$ is given by $\sum_{n=0}^{\infty} c_n$ with $c_n = \sum_{k=0}^{n} a_k b_{n-k}$.

2.4 Power series

Chapter 3

Curves in $\mathbb{R}^n$ and Differential Equations

3.1 Curves in $\mathbb{R}^n$

Definition 3.1.1 (Curve)**.** A *curve* in a metric space X is a continuous mapping $\gamma : I \to X$ where $I \subset \mathbb{R}$ is an interval.

With $X = \mathbb{R}^n$ a curve γ is a vector valued map

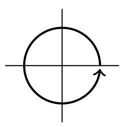
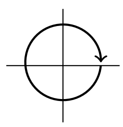
$$\gamma(t) = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \vdots \\ \gamma_n(t) \end{pmatrix} \in \mathbb{R}^n$$

where all $\gamma_i(t) : I \mapsto \mathbb{R}$ are continuous.

Note that a curve refers to the *whole* map and not just the *image* produced by the map and that two different curves can have the same image. Take, for instance, the curves

$$\gamma_1, \gamma_2 : [0, 2\pi] \mapsto \mathbb{R}^2 \text{ with } \gamma_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \text{ and } \gamma_2(t) = \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix}.$$

Both of these curves are different but give us the same image (the unit circle) as γ_1 and γ_2 "traverse" the circle in the anti-clockwise and clockwise direction respectively.

γ_1 :  γ_2 : 

Recall (Dot product, Norm, Cauchy–Schwarz)*.* The *dot product* of two vectors $x, y \in \mathbb{R}^n$ is is given by

$$x \cdot y := \sum_{i=1}^{n} x_i y_i.$$

Also denoted by $\langle x, y \rangle$.

The *norm* of a vector $x \in \mathbb{R}^n$ gives us its "length" with

$$\|x\| = \sqrt{x \cdot x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

The *Cauchy–Schwarz inequality* relates the dot product of two vectors $x, y \in \mathbb{R}^n$ to their norms:

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

Definition 3.1.2 (Differentiability of γ)**.** A curve $\gamma : I \to \mathbb{R}^n$ is differentiable at $t \in I$ if

$$\gamma'(t) = \lim_{\varepsilon \to 0} \frac{\gamma(t+\varepsilon) - \gamma(t)}{\varepsilon}$$

exists in $\mathbb{R}^n$. We call γ' the *derivative* of γ.

Note.

- Above

  $$\frac{\gamma(t+\varepsilon) - \gamma(t)}{\varepsilon} := \begin{pmatrix} \frac{1}{\varepsilon}[\gamma_1(t+\varepsilon) - \gamma(t)] \\ \vdots \\ \frac{1}{\varepsilon}[\gamma_n(t+\varepsilon) - \gamma(t)] \end{pmatrix}.$$

  Hence, γ is differentiable if and only if all γ_i are differentiable.

- When I is not open we consider the one-sided limits at the end points.

- γ differentiable $\implies$ γ continuous

We can also express the derivative of γ in terms of an approximation by an affine linear function. So if γ is differentiable at $t \in I$ then there exists $T_n : \mathbb{R} \mapsto \mathbb{R}^n$ such that

$$T_t(s) = \gamma(t) + (s-t)\gamma'(t)$$

is a unique straight line with $T_t(t) = \gamma(t)$ and $T_t'(t) = \gamma'(t)$. T_t is also the best approximation of γ at t. Take $s = t + \varepsilon$, then

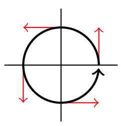
$$\begin{aligned} \|\gamma(s) - T_t(s)\| &= \|\gamma(s) - \gamma(t) - [T_t(s) - \gamma(t)]\| \\ &= \left\| \varepsilon \frac{\gamma(t+\varepsilon) - \gamma(t)}{\varepsilon} - (s-t)\gamma'(t) \right\| \\ &= |\varepsilon| \left\| \frac{\gamma(t+\varepsilon) - \gamma(t)}{\varepsilon} - \gamma'(t) \right\| \end{aligned}$$

So, $\lim_{\varepsilon \to 0} \frac{1}{|\varepsilon|}\|\gamma(s) - T_t(s)\| = 0$.

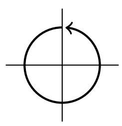
Hence, γ is differentiable at t if $\gamma(s) = T_t(s) + R_t(s)$ where $R_t(s)$ is the remainder term with $\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}|R_t(t+\varepsilon)| = 0$.

Example 3.1.1.

$$\gamma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} :$$



$$\gamma'(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix} :$$



Lemma 3.1.1 ("Speed limit")**.** *Let $\gamma : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b) with $\|\gamma'(t)\| < c$ for all $t \in (a,b)$, then*

$$\|\gamma(b) - \gamma(a)\| \leq c|b-a|.$$

Proof. Let $\Delta = \gamma(b) - \gamma(a)$ and $f(t) = \Delta\gamma(t) = \sum \Delta\gamma_i(t) \implies f(t)$ is differentiable and continuous as $\gamma(t)$ is, and $f'(t) = \Delta\gamma'(t)$.

From the mean value theorem, we know that there exists $x \in [a,b]$ such that $f(b) - f(a) = (b-a)f(x) = (b-a)\Delta\gamma'(x)$.

Now,

$$\begin{aligned} 0 \leq \underbrace{\Delta^2}_{=\|\Delta\|^2} &= \Delta(\gamma(b) - \gamma(a)) = f(b) - f(a) \\ &= (b-a)\Delta\gamma'(x) \leq (b-a)\|\Delta\| \underbrace{\|\gamma'(x)\|}_{<c} \end{aligned}$$

When $\Delta = 0$ the inequality is clear. Otherwise, $\|\gamma(b) - \gamma(a)\| = \|\Delta\| \leq (b-a)c$. □

Definition 3.1.3 (Integration of curves)**.** Let $\gamma : [a,b] \mapsto \mathbb{R}^n$, with

$$\gamma(t) = \begin{pmatrix} \gamma_1(t) \\ \vdots \\ \gamma_n(t) \end{pmatrix}$$

and $\alpha : [a,b] \mapsto \mathbb{R}$ be monotonically increasing. Then $\gamma \in \mathcal{R}(\alpha)[a,b]$ if and only if for all $i, \gamma_i \in \mathcal{R}(\alpha)[a,b]$ and

$$\int_a^b \gamma \, \mathrm{d}\alpha = \begin{pmatrix} \int_a^b \gamma_1 \, \mathrm{d}\alpha \\ \vdots \\ \int_a^b \gamma_n \, \mathrm{d}\alpha \end{pmatrix}$$

Note. Most properties from $n = 1$ also hold here. For instance, linearity of α, γ; $\int \gamma \, \mathrm{d}\alpha = \int \gamma\alpha' \, \mathrm{d}t$ for differentiable α and integrable α'; fundamental theorem; etc.

Lemma 3.1.2. *Let $\gamma : [a,b] \to \mathbb{R}^n$ with $\gamma \in \mathcal{R}(\alpha)[a,b]$. Then $\|\gamma\| \in \mathcal{R}(\alpha)[a,b]$ and $\left\|\int_a^b \gamma \, \mathrm{d}\alpha\right\| \leq \int_a^b \|\gamma\| \, \mathrm{d}\alpha$.*

Proof. (analogous to proof of Lemma 3.1.1.) First note that $\|\gamma\| = \sqrt{\sum \gamma_i^2} \in \mathcal{R}(\alpha)[a,b]$ as sums, products, and square roots of integrable functions are integrable.

Next, define $\Delta := \int_a^b \gamma \,\mathrm{d}\alpha$. Then

$$\begin{aligned}
0 \leq \|\Delta\|^2 &= \sum_1^n \Delta_i \cdot \Delta_i \\
&= \sum_1^n \Delta_i \int_a^b \gamma_i \,\mathrm{d}\alpha \\
&= \int_a^b \Delta\gamma \,\mathrm{d}\alpha \\
&\leq \int_a^b \|\Delta\|\|\gamma\| \,\mathrm{d}\alpha = \|\Delta\| \int_a^b \|\gamma\| \,\mathrm{d}\alpha,
\end{aligned}$$

where the inequality follows from applying Cauchy-Schwarz. For $\Delta = 0$ the proof is clear. Otherwise, for $\Delta \neq 0$, we have

$$\|\Delta\| = \left\| \int_a^b \gamma \,\mathrm{d}\alpha \right\| \leq \int_a^b \|\gamma\| \,\mathrm{d}\alpha.$$

□

3.2 Rectifiablity and curve length

The general idea behind computing the length of some curve $\gamma : [a, b] \to \mathbb{R}^n$ is to first select some partition $\mathcal{P} = \{x_0, \ldots, x_n\}$ with $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$ and then make the partitions finer and finer so that the sum of the straight line segments

$$\lambda(\gamma, \mathcal{P}) = \sum_{i=1}^n \|\gamma(x_i) - \gamma(x_{i-1})\|$$

approaches the length of the curve.

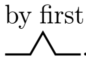
Note. If $\mathcal{P}^* \supset \mathcal{P}$ is a refinement of $\mathcal{P}$, then

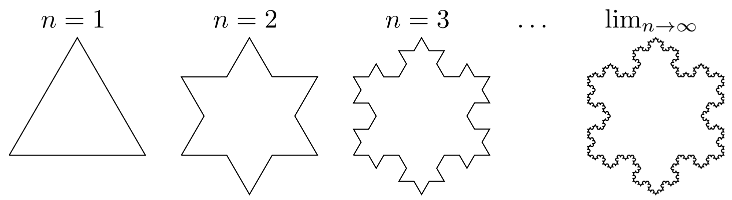
$$\lambda(\gamma, \mathcal{P}^*) \geq \lambda(\gamma, \mathcal{P}).$$

This follows directly from the triangle inequality.

Definition 3.2.1 (Rectifiable curves)**.** For $\Lambda(\gamma) = \sup_{\mathcal{P}} \lambda(\gamma, \mathcal{P})$ the curve γ is *rectifiable* if $\Lambda(\gamma) < \infty$. If this is indeed the case, we call $\Lambda(\gamma)$ the length of the curve γ.

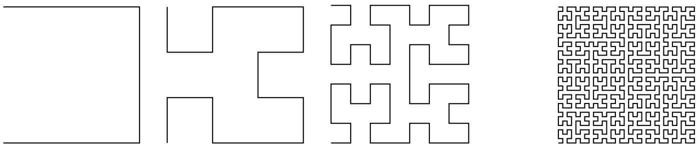
Note. There exist many non-rectifiable curves.

Example 3.2.1 (The Koch snowflake)**.** The Koch snowflake is obtained by first taking a triangle and then replacing each straight line segment into a . Iteratively doing this on all line segments produces the Koch snowflake



Note that the total length at the end of the n-th iteration is given by $3\left(\frac{4}{3}\right)^n$ Since $\lim_{n\to\infty} 3\left(\frac{4}{3}\right)^n = \infty$, the Koch snowflake has infinite length. See Homework B.6.2 for more details.

Example 3.2.2 (Space filling curves)**.** A *space filling curve* is a curve $\gamma : [0,1] \mapsto [0,1]^2$ that is continuous and surjective (these are usually limits of repeated patterns.)



See Homework B.6.2 (bonus) for more details.

Theorem 3.2.1. *If $\gamma : [a,b] \mapsto \mathbb{R}^n$ is continuous differentiable, then γ is rectifiable and $\Lambda(\gamma) = \int_a^b \|\gamma'(t)\|\,\mathrm{d}t$.*

Proof.

$$\begin{aligned}
\lambda(\gamma, \mathcal{P}) &= \sum_1^n \|\gamma(x_i) - \gamma(x_{i-1})\| = \sum_1^n \left\| \int_{x_{x-i}}^{x_i} \gamma'(t)\,\mathrm{d}t \right\| \\
&\le \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \|\gamma'(t)\|\,\mathrm{d}t = \int_a^b \|\gamma'(t)\|\,\mathrm{d}t \\
&\implies \Lambda(\gamma, \mathcal{P}) \le \int_a^b \|\gamma'(t)\|\,\mathrm{d}t < \infty \\
&\implies \gamma \text{ is rectifiable.}
\end{aligned}$$

Next, we find $\mathcal{P}$ such that $\Lambda(\gamma, \mathcal{P}) \ge \int_a^b \|\gamma'(t)\|\,\mathrm{d}t - \varepsilon$ for all $\varepsilon > 0$. Using uniform continuity of γ on $[a,b]$ for fixed $\varepsilon > 0$ there exists a δ such that for al $s, t \in [a,b]$, $|s-t| \implies \|\gamma'(s) - \gamma'(t)\| < \varepsilon$. If we take $\mathcal{P}$ with $|x_i - x_{i-1}| < \delta$ for all $i = 1, \ldots, n$ then

$$\begin{aligned}
\int_{x_{i-1}}^{x_i} \|\gamma'(t)\|\,\mathrm{d}t &= \int_{x_{i-1}}^{x_i} \Big(\|\gamma(x_{i-1})\| + \varepsilon \Big)\,\mathrm{d}t \\
&= \|\gamma'(x_i)\|\Delta x_i + \varepsilon \Delta x_i \\
&= \left\| \int_{x_{i-1}}^{x_i} \Big(\gamma'(t) + \gamma'(x_{i-1}) - \gamma'(t) \Big)\,\mathrm{d}t \right\| + \varepsilon \Delta x_i \\
&\le \left\| \int_{x_{i-1}}^{x_i} \gamma'(t)\,\mathrm{d}t \right\| + \left\| \int_{x_{i-1}}^{x_i} \Big(\gamma'(x_{i-1}) - \gamma'(t) \Big)\,\mathrm{d}t \right\| + \varepsilon \Delta x_i \\
&\le \int_a^b \|\gamma'(t)\|\,\mathrm{d}t + \varepsilon \Delta x_i + \varepsilon \Delta x_i \\
&\overset{\sum_1^n}{\implies} \int_a^b \|\gamma'(t)\|\,\mathrm{d}t \le \Lambda(\gamma, \mathcal{P}) + 2(b-a)\varepsilon.
\end{aligned}$$

□

We saw that the length of a curve is given by

$$\Lambda(\gamma) = \int_a^b \|\gamma'(t)\| \,\mathrm{d}t.$$

A more general statement is also true: For any curve $\gamma : [a, b] \mapsto \mathbb{R}^n$ that is differentiable on (a, b) with the set $\{\|\gamma'(t)\| : a \leq t \leq b\}$ bounded and integrable $t \mapsto \|\gamma'(t)\|$. Then $\Lambda(\gamma) = \int_a^b \|\gamma'(t)\| \,\mathrm{d}t$.

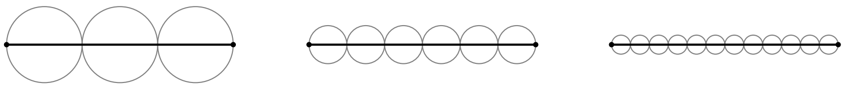
We denote the length of γ from a to x by $\Lambda(\gamma, a, x)$. Note that

$$\frac{\mathrm{d}\Lambda(\gamma, a, x)}{\mathrm{d}x} = \|\gamma'(x)\|$$

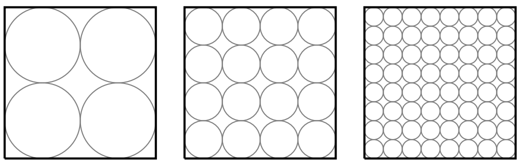
is true by the fundamental theorem if γ' is continuous.

Fractals When the *fractal dimension* of a curve is greater than one then the curve is non-rectifiable. We get this fractal dimension by looking at the growth in the number of circles required to cover the entire figure as we make these circles smaller and smaller. The fractal dimension is then just the ratio at which the number of circles required grows.

For a straight line the total number of circles required is proportional to the size. Hence, its fractal dimension is one.



For a square, the fractal dimension is two as the number of circles required grows quadratically.



For fractals, e.g., the Koch snowflake from the example above, the fractal dimension is non-integer.

Lemma 3.2.2. *If a function $f : [a, b] \mapsto \mathbb{R}$ is continuous and monotone, then the graph of $f(t)$ given by $\Gamma(t) = \begin{pmatrix} t \\ f(t) \end{pmatrix}$ is rectifiable.*

Proof. For any partition $\mathcal{P}$ and $x_i \in \mathcal{P}$, we have

$$\|\Gamma(x_i) - \Gamma(x_{i-1})\| = \left\| \begin{pmatrix} x_i - x_{i-1} \\ f(x_i) - f(x_{i-1}) \end{pmatrix} \right\| \leq (x_i - x_{i-1}) + (f(x_i) - f(x_{i-1}))$$

Since the function is monotone, summing up for all i gives us

$$\Lambda(\Gamma, \mathcal{P}) \leq (b - a) + |f(b) - f(a)|$$

which is bounded for all $\mathcal{P}$ hence, $\sup_{\mathcal{P}} \Lambda(\Gamma, \mathcal{P})$ exists and Γ is rectifiable. □

Definition 3.2.2 (Reparametrization)**.** For a curve $\gamma : [a, b] \to \mathbb{R}^n$ and a function $\varphi : [a', b'] \mapsto [a, b]$, the *reparametrization* of γ is given by $(\gamma \circ \varphi)(t) = \gamma(\varphi(t))$.

Example 3.2.3. For $\gamma : [0, 2\pi] \to \mathbb{R}^2$ with $\gamma(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$ and $\varphi : [0, 1] \mapsto [0, 2\pi]$ with $\varphi(t) = 2\pi t^2$ the reparametrization of γ with respect to φ is

$$(\gamma \circ \varphi)(t) = \gamma(\varphi(t)) = \begin{pmatrix} \cos 2\pi t^2 \\ \sin 2\pi t^2 \end{pmatrix}.$$

Lemma 3.2.3. *If $\gamma : [a, b] \to \mathbb{R}^n$ is continuously differentiable and $\varphi : [c, d] \to [a, b]$ is a diffeomorphism (a bijection with differentiable φ and φ^{-1}) then $\Lambda(\gamma) = \Lambda(\gamma \circ \varphi)$.*

Proof.

$$\Lambda(\gamma) = \int_a^b \|\gamma'(t)\| \,\mathrm{d}t = \int_{\varphi^{-1}(a)=c}^{\varphi^{-1}(b)=d} \|\gamma'(\varphi(t))\| \varphi'(t) \,\mathrm{d}t,$$
$$\Lambda(\gamma \circ \varphi) = \int_c^d \left\| \frac{\mathrm{d}\gamma(\varphi(t))}{\mathrm{d}t} \right\| \mathrm{d}t = \int_c^d \|\gamma'(\varphi(t))\| \varphi'(t) \,\mathrm{d}t.$$

□

3.3 Differential equations

Motivation A differential equation relates a function to its derivative. Consider the following examples:

Example 3.3.1. If $x'(t) = \lambda x(t)$ for $\lambda \in \mathbb{R}$ then the solution is $x(t) = a\mathrm{e}^{\lambda t}$ with $a \in \mathbb{R}$.

Example 3.3.2. $x'(t) = tx(t) \implies x(t) = a\mathrm{e}^{t^2/2}, a \in \mathbb{R}$.

Example 3.3.3. $x'(t) = \frac{\lambda x(t)}{t}$ with $\lambda \in \mathbb{R}$ then $x(t) = at^\lambda$ for $a \in \mathbb{R}$.

Example 3.3.4. $x'(t) = -x(t)^2 \implies x(t) = \frac{1}{t-a}, a \in \mathbb{R}$.

There are only few examples of differential equations were we can explicitly find the solution. In most interesting cases we cannot find it explicitly. However, these simple cases are important for more general results (e.g., think of the Gronwall lemma, see Homework B.7.4). On the other hand, the theory of differential equations is more concerned with the questions of existence and uniqueness of solutions to general classes of equations, and with the properties of such solutions — Are they global or local? What is their long time behavior? etc. Here we only concern ourselves with first order autonomous (time-independent) differential equations.

Definition 3.3.1 (Vector field, Flow line)**.** A *vector field* in $\mathbb{R}^n$ is a map $v : \mathbb{R}^n \to \mathbb{R}^n$. A *flow line* in this vector field v is a continuously differentiable map (a $\mathcal{C}^1$ map) $\gamma : I \to \mathbb{R}^n$ with $\gamma'(t) = v(\gamma(t))$ where I is some interval.

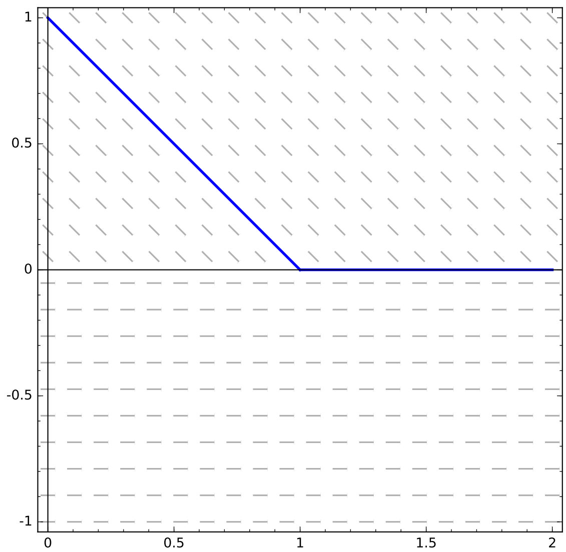
Definition 3.3.2 (Initial value problem)**.** Given a vector field v, initial time $t_0 \in \mathbb{R}$, initial position $x_0 \in \mathbb{R}^n$, is there a flow line $\gamma : I \to \mathbb{R}^n$ such that $\gamma'(t) = v(\gamma(t))$ with $\gamma(t_0) = x_0$? Is so, is the curve unique?

Note. The existence/uniqueness can be *local* i.e., we get solutions only on some interval $I \in \mathbb{R}$ with $t_0 \in I$. A *global* solution works for all $t \in \mathbb{R}$.

Example 3.3.5. Consider the function

$$v(x) = \begin{cases} -1, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

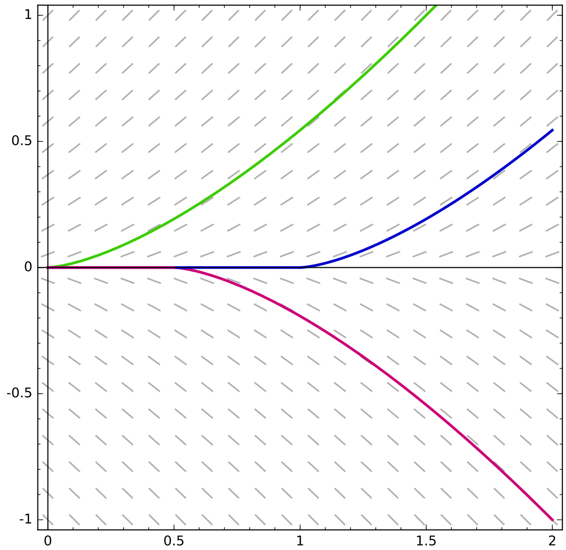
Heuristically, any solution will have a point where it is not differentiable (for instance, the point $(1, 0)$ in the graph below)[1].



Hence, the differential equation $\gamma'(t) = v(\gamma(t))$ doesn't have a global solution.

[1]The vector field $v(x)$ is shown in gray and the solution curve γ in blue.

Example 3.3.6. Now consider the function $v(x) = x^{\frac{1}{3}}$ for some t_0 and $x_0 = 0$.



For $v(x)$ we can find many $\mathcal{C}^1$ solutions. For instance $\gamma_0(t) = 0, \gamma_1(t) = \left(\frac{2t}{3}\right)^{\frac{3}{2}}, \gamma_3 = -\left(\frac{2t}{3}\right)^{\frac{3}{2}}$, or even

$$\gamma_\tau(t) = \begin{cases} 0, & t < \tau \\ \pm\left(\frac{2(t-\tau)}{3}\right)^{\frac{3}{2}}, & t \geq \tau \end{cases}.$$

Hence, solutions to $v(x)$ in this case aren't unique.

Definition 3.3.3 (Lipschitz condition)**.** A vector field v satisfies the *global Lipschitz condition* (or is Lipschitz continuous) with some Lipschitz constant $L > 0$ if for all $x_0 \in \mathbb{R}^n, x_1 \in \mathbb{R}^n$

$$\|v(x_1) - v(x_0)\| \leq L\|x_1 - x_0\|.$$

A vector field v satisfies the *local* Lipschitz condition if for all $x \in \mathbb{R}^n$ there is a neighborhood U_x such that for all $x_0, x_1 \in U_x$ the Lipschitz condition holds with some Lipschitz condition L_x.

Note. Local Lipschitz condition implies continuity and the global Lipschitz condition implies uniform continuity.

Example 3.3.7. If $v(x) = \lambda x$ then $\|v(x_0) - v(x_1)\| = \lambda\|x_0 - x_1\|$. Hence, $v(x)$ satisfies the global Lipschitz condition. Note that $\|a\mathrm{e}^{\lambda t} - b\mathrm{e}^{\lambda t}\| = \mathrm{e}^{\lambda t}\|a - b\|$, which is true as an inequality for a large class of differential equations.

Theorem 3.3.1 (Picard–Lindelöf)**.** *Let $v : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a vector field that satisfies the global Lipschitz condition with Lipschitz constant L. Further let $t_0 \in \mathbb{R}$ and $x_0 \in \mathbb{R}$. Then there exists a unique continuously differentiable (i.e., $\mathcal{C}^1$) curve $\gamma : \mathbb{R} \mapsto \mathbb{R}^n$ with $\gamma'(t) = v(\gamma(t))$.*

Furthermore, two solution curves γ_1 and γ_2 each with different initial data (t_0, x_0) satisfy

$$\|\gamma_1(t) - \gamma_2(t)\| \leq \mathrm{e}^{L|t-t_0|}\|\gamma_1(t_0) - \gamma_2(t_0)\|.$$

Note.

- Only continuity implies local existence of solutions but not necessarily uniqueness.
- Only (local) Lipschitz continuity implies local existence and uniqueness.
- Non-autonomous (time-dependent) vector fields i.e., of the form $v(x,t)$ imply existence and uniqueness of solutions if the field satisfies the Lipschitz condition in x and is continuous in t.

Proof sketch (3.3.1). The general idea is to fist integrate locally ($\int_{t_0}^{s} v(\gamma(t))\,\mathrm{d}t$) and then show the existence of $\gamma = \int_{t_0}^{s} v(\gamma(t))\,\mathrm{d}t$ and then extend this local solution to all of $\mathbb{R}$.

With $\varepsilon = \frac{1}{2}L$, first define a space of curves

$$X = \{\gamma : [t_0, t_0 + \varepsilon] \mapsto \mathbb{R}^n \text{ with } \gamma(t_0) = x_0\}$$

Note that the metric space (X, d) with

$$d(\gamma_1, \gamma_2) = \|\gamma_1 - \gamma_2\|_{\sup} = \sup_{t\in[t_0,t_0+\varepsilon]} \|\gamma_1(t) - \gamma_2(t)\|$$

is a complete metric space.

Now define a map $P : X \mapsto X$ such that for $s \in [t_0, t_0 + \varepsilon]$

$$P_\gamma(s) = \int_{t_0}^{s} v(\gamma(t))\,\mathrm{d}t + x_0 \implies P_\gamma \in X.$$

We are interested in a fixed point of P i.e., a curve $\gamma^* : [t_0, t_0 + \varepsilon] \mapsto \mathbb{R}^n$ and $\gamma^*(t_0) = x_0$ with $\gamma^* = P_{\gamma^*}$. Because then

$$\frac{\mathrm{d}\gamma^*}{\mathrm{d}s} = \frac{\mathrm{d}P_{\gamma^*}(s)}{\mathrm{d}s} = v(\gamma^*(s)).$$

That is, γ^* is a solution on $[t_0, t_0 + \varepsilon]$.

Existence and uniqueness is provided by the Banach fixed point theorem (see 3.3.2) (also called the contraction mapping principle).

We need to show that P_γ is indeed a contraction i.e.

$$\|P_{\gamma_1}(s) - P_{\gamma_2}(s)\| \leq \cdots = L \cdot \varepsilon \cdot d(\gamma_1, \gamma_2) = \frac{1}{2} \cdot d(\gamma_1, \gamma_2).$$

Then, this gives us a unique solution on $[t_0, t_0 + \varepsilon]$. Repeating this for other intervals $[t_0 + \varepsilon, t_0 + 2\varepsilon], \ldots$ gives us the global solution.

The inequality (from the theorem) is proven by Grönwall's inequality (see 3.3.3). □

Theorem 3.3.2 (Banach fixed point theorem)**.** *If (X, d) is a complete metric space and $f : X \mapsto X$ a contraction (i.e., there exists $r \in (0, 1)$ such that for all $x, x' \in X, d(f(x), f(x')) \leq rd(x, x')$) then f has a unique fixed point $x^* \in X$, meaning that $f(x^*) = x^*$.*

Proof idea. For $x_0 \in X$ define a sequence $x_n = f(x_{n-1})$. Then one can show that $x_n \to x^*$. □

Lemma 3.3.3 (Grönwall's inequality)**.** *If f is differentiable, then for some $c \in \mathbb{R}$,*

$$f'(t) \leq cf(t) \implies f(t) \leq \mathrm{e}^{ct} f(0).$$

Proof. Was done in some bonus problem from Adv. Calc. □

Note. For higher order ODEs $x^{(n)} = F(x, x', x'', \ldots, x^{(n-1)})$ we can introduce $x_1 = x, x_2 = x', x_3 = x'', \ldots, x_n = x^{(n-1)}$ and then solve for

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \\ \vdots \\ x_n' \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ x_4 \\ \vdots \\ F(x_1, \ldots, x_n) \end{pmatrix} = V(x_1, \ldots, x_n).$$

Theorem 3.3.4. *If I, J are open intervals and $f : I \mapsto \mathbb{R}$, $g : J \mapsto \mathbb{R}$ are continuous with $g(x) \neq 0$ for all $x \in J$, then for all initial conditions $(t_0, x_0) \in I \times J$ there is an open interval $I_0 \subset I$ containing t_0 and a $\mathcal{C}^1$ function $\gamma : I_0 \mapsto \mathbb{R}$ with $\gamma(t_0) = x_0$ that solves the differential equation $\gamma'(t) = f(t) \cdot g(\gamma(t))$.*

Proof. See Homework B.7.3. □

Chapter 4

Basic Topology

4.1 Topology and continuity

Motivation We would like to study continuity and convergence form a more abstract viewpoint. We do so by first generalizing concepts about open sets.

Example 4.1.1 (ε-balls in $\mathbb{R}$)**.** An ε-ball around a point x is defined as

$$B_\varepsilon(x) = \{x' \in \mathbb{R} : |x - x'| < \varepsilon\}\,.$$

These satisfy the following properties:

- Arbitrary (even infinite) unions of open balls are open.
- Finite intersections are open (but infinite intersections might produce closed intervals.).

Definition 4.1.1 (Topology)**.** A *topology* on a set X is a collection $\tau = \{U_i\}_{i\in I}$ of subsets $U_i \subset X$ where I is any index set such that

- $\varnothing, X \in \tau$.
- For any $I' \in I$, also $\bigcup_{i\in I'} U_i \in \tau$ (arbitrary union).
- For any *finite* $I' \in I$, $\bigcap_{i\in I'} U_i \in \tau$.

Elements of τ are called *open sets* and their complements are called *closed sets*. The original set together with the topology (X, τ) is called a *topological space*.

Example 4.1.2 (Standard topology)**.** The standard topology on a metric space (X, d) can be *generated* using:

- ε-Balls $B_\varepsilon(x) = \{x' \in \mathbb{R} : |x - x'| < \varepsilon\}$ are defined as open sets.
- All open sets are arbitrary unions of $B_\varepsilon(x)$ (with possibly different ε and x).

Example 4.1.3 (Trivial (in-discrete) topology)**.** $\tau = \{\varnothing, X\}$ is the *coarsest* possible topology on X.

Example 4.1.4 (Discrete topology)**.** $\tau = \mathcal{P}(X)$ (the power set of X) is the *finest* possible topology on X (here, all possible subsets of X are defined to be open).

Example 4.1.5. For $X = \{1, 2, 3\}$, $\tau = \{\varnothing, \{1\}, \{2\}, \{1, 2, 3\}\}$ is not a topology as $\{1\} \cup \{2\} = \{1, 2\} \notin \tau$. But $\tau = \{\varnothing, \{1\}, \{1, 2, 3\}\}$ is a topology.

Definition 4.1.2. Let (X, τ_X) be a topological space and $Y \subset X$, then the *subspace topology* on Y *inherited/induced* from X is

$$\tau_Y = \{Z \subset Y \text{ such that } Z = U_i \cap Y, U_i \in \tau_X\}.$$

Example 4.1.6. The standard topology on circles, etc.

Note. In general $\tau_Y \subset \tau_X$ is not true.

Definition 4.1.3. Let (X, τ) be a topological space, then

- An *open neighborhood* of $p \in X$ is any open set containing p.
- A *neighborhood* of $p \in X$ is a set that contains an open neighborhood of p.

Definition 4.1.4 (Hausdorff space)**.** A topological space (X, τ) is called a *Hausdorff space* if for any $x, y \in X$ with $x \neq y$ there are open sets $U \ni x$ and $V \ni y$ such that U, V are disjoint.

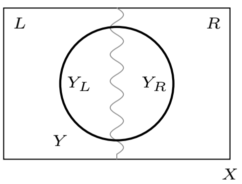
Note. Any metric space (X, d) with the standard topology is Hausdorff. For x, y take $U = B_\varepsilon(x), V = B_\varepsilon(y)$ where $\varepsilon = d(x, y)/2$.

Example 4.1.7 (Zariski topology)**.** The *Zariski topology* on an infinite set X is defined such that a set $U \subset X$ is open if and only if $U = \varnothing$ or $X \setminus U$ is finite. The Zariski topology is not Hausdorff.

Example 4.1.8. The trivial topology $(X, \{\varnothing, X\})$ is not Hausdorff provided $|X| \geq 2$.

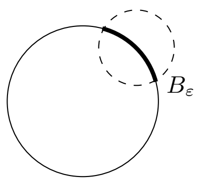
Example 4.1.9. The discrete topology is Hausdorff.

Example 4.1.10 ("Funny" induced topology)**.**



$$\tau_X = \{\varnothing, X, L, R\}$$
$$\tau_Y = \{\varnothing = Y \cap \varnothing, Y = Y \cap X, Y_L = L \cap Y, Y_R = Y \cap R\}$$

Example 4.1.11 ($\mathbb{R}^2$ with std. topology)**.** The induced topology on the circle is just the open intervals *on* the circle (the bold part below). We find these by looking at intersections of open balls B_ε in $\mathbb{R}^2$ with the circle.



Note. If we talk about subsets of a topological subspace, we always assume the induced topology.

Definition 4.1.5 (Topological continuity)**.** Let (X, τ_X) and (Y, τ_Y) be topological spaces, then a map $f : X \to Y$ is called continuous if for every open set $U \subset Y$, the pre-image $f^{-1}(U) = \{x \in X : f(x) \in U\}$ is open.

Note. The definition above is often shortened as "pre-images of open sets are open" or, alternatively, "pre-images of closed sets are closed." On metric spaces this definition is equivalent to the ε–δ definition of continuity (proved last semester).

4.2 Compact sets

Motivation Compact sets allows us to conclude "finiteness" properties for infinite sets. For instance, why do some functions, say $\frac{1}{x}$ on $(0, 1)$, not have a maximum? Why is it not uniformly continuous? On the other hand, why do bounded functions like $1 - \mathrm{e}^{-x}$ for $x \geq 0$ not have a maximum? With compactness we can generalize notions of bounded and closed sets on $\mathbb{R}$.

Definition 4.2.1 (Cover of a topological space)**.** A *cover* of a topological space (X, τ_X) is a collection of sets $\{V_i\}$ with $V \subset X$ such that $\bigcup_i V_i = X$. A cover is open if and only if all V_i are open.

Definition 4.2.2 (Compactness)**.** A topological space (X, τ_X) is called *compact* if every open cover has a finite sub-cover.

Note. In general, on metric spaces compactness = sequential compactness, i.e., every sequence in X has a sub-sequence that converges in X.

Theorem 4.2.1. *If a topological space (X, τ_X) is compact, then every continuous function $f : X \to \mathbb{R}$ assumes a maximum and minimum.*

Proof. We first show that f is bounded by contradiction. Assuming f is unbounded, for all $n \in \mathbb{Z}$ construct

$$U_n = \{x \in X : |f(x) - n| < 2\}.$$

This implies that all U_n are open as they are pre-images of open sets and f is continuous. Then, $\{U_n : n \in \mathbb{Z}\}$ is an open cover of X. But if f is unbounded, then no finite sub-cover of $\{U_n\}$ will cover X ↯ Since we get a contradiction to compactness, f has to be bounded.

Next, we set $m = \sup_{x \in X} f(x) \in \mathbb{R}$ to show that there is some x_m such that $f(x_m) = m$. We show this by contradiction by assuming that no such maximum exists. Then for $n \in \mathbb{Z}$ set

$$U_n = \left\{x \in X : f(x) \in (m - 2^{-(n-1)}, m - 2^{-(n+1)})\right\}.$$

Then $\{U_n, n \in \mathbb{Z}\}$ are an open cover (all $f(x) < m$ by assumption). But, there is no finite sub-cover as $f(x)$ can be arbitrarily close to m (since m is the supremum) ↯ We again get a contradiction to compactness, i.e., such an x_m exists.

Hence, f assumes it's maximum (for the minimum, we just need to prove the result for $-f$). □

Alternative proof (using Theorems 4.2.2 and 4.2.3). X is compact and $f : X \mapsto \mathbb{R}$ continuous then the image of $f(x)$ is compact (bounded and closed) hence f assumes its maximum and minimum. □

Theorem 4.2.2 (Heine–Borel)**.** *A subset $X \in \mathbb{R}$ is compact if and only if X is bounded and closed.*

Proof. " $\Longrightarrow$ ". We first assume that f is not bounded and define

$$U_n := \{x \in X \colon |x - n| < 1\}.$$

These are open sets of X (intersections of X with open sets in $\mathbb{R}$; note that they are not necessarily open sets in $\mathbb{R}$). Then U_n for $n \in \mathbb{Z}$ are open covers of X but as X is not bounded, no finite cover exists ↯ Contradiction to compactness. Next assume that X is bounded but not closed. Then X not closed $\Longrightarrow$ there is a limit point p of a sequence in X such that $p \in X^c$. Then define

$$U_n = \{x \in X \colon |x - p| \in (2^{-(n-1)}, 2^{-(n+1)})\}.$$

These U_n-s are an open cover but no finite sub-cover exists ↯.

"$\Longleftarrow$". Let $\{U_i\}$ be an open cover of X and define the sequence

$$X_m = X \cap [m, m+1].$$

Assuming one of the X_m-s cannot be covered by finitely may U_i, inductively define sets

$$Y_n = \left[a_n - 2^{-(n+1)}, a_n + 2^{-(n+1)}\right] \cap X$$

such that Y_n cannot be covered by finitely many U_i-s. Here, the sequence a_n is Cauchy; say $\lim_{n\to\infty} a_n = a \in \mathbb{R}$. Since X is closed, we also have $a \in X$. Since all U_i-s are open covers, there is some U_k with $a \in U_k$. Since this selected U_k is also open, there is some $N \in \mathbb{N}$ such that for all $n > N$, $Y_n \subset U_k$ i.e., $Y_1 \supset Y_2 \supset \dots$ Hence all Y_n-s have a finite sub-cover ↯ (contradiction to our definition of Y_n). □

Theorem 4.2.3. *The continuous images of any compact set is compact.*

Proof. See Homework B.8.2. □

Theorem 4.2.4. *If a map $f : X \mapsto Y$ where (X, d_X) and (Y, d_Y) are metric spaces with X compact is continuous, then it is also uniformly continuous.*

Proof. See Homework B.9.2. □

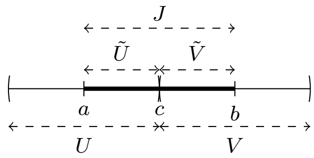
4.3 Connected sets

Definition 4.3.1 (Connectedness)**.** A topological space (X, τ) is *connected* if the only clopen (closed and open) subsets of X are X and $\varnothing$. A subset is connected if it is connected in the subspace topology.

Note. A topological space (X, τ) is not connected if there is a *separation / decomposition* of $X = U \cup V$ where U, V are non-empty, disjoint, and open (and thus also closed).

Lemma 4.3.1. *An interval $I \subset \mathbb{R}$ is connected if and only if it is an interval or a point, i.e., for $x, y \in I$ and $x < z < y$ for some z then $z \in I$ as well.*

Proof. "$\Longleftarrow$". Consider $I = [a, b]$ and suppose that it is not connected. Then there are non-empty open sets U, V (mutually disjoint) such that $I = U \cup V$.



Let $a \in U, b \in V$ and without loss of generality take $a < b$ (if not, then rename U, V). Take $J = [a, b]$. Then J can be separated with $J = \tilde{U} \cup \tilde{V}$ with $\tilde{U} = U \cap J$ and $\tilde{V} = V \cap J$. Set $c = \sup \tilde{U}$. Now, if $c \in \tilde{U}$ then also $[c, c + \varepsilon] \in \tilde{U}$ (for some small enough ε as $\tilde{U}$ is open in the subspace topology on J) $\implies$ c cannot be the supremum ↯. On the other hand, if $c \in \tilde{V}$ then also $[c - \varepsilon, c] \in \tilde{V}$ for small enough ε. ↯.

" $\implies$ ". Take some $X \subset \mathbb{R}$. If X is an interval, then for all $a, b \in X$ and $a < c < b$, $c \in X$ as well. Suppose X is not an interval and there exists some $c \in (a, b)$ such that $c \notin X$. Then define

$$X_- = \{x \in X : x < c\}, \quad X_+ = \{x \in X : x > c\}.$$

In $\mathbb{R}$, both X_- and X_+ are open as they are, respectively, the intersections of the open sets $(-\infty, c)$ and (c, ∞) with X. Hence they are open in X as well. They are also disjoint and non-empty. As $c \notin X$, $X = X_- \cup X_+ \implies X$ is disconnected ↯. □

Lemma 4.3.2. *The continuous image of a connected set is connected.*

Proof. Let $f : X \to Y$ be continuous, X connected, and $Y = f(X)$ (the image of X under f). By definition f is surjective. Now, suppose that Y is not connected, then there is a clopen set V such that $V \neq \varnothing$ and $V \neq Y$. Then the pre-image $f^{-1}(V)$ is clopen $\implies$ X is not connected ↯. □

Theorem 4.3.3 (Generalized intermediate value theorem)**.** *Let X be a connected topological space and $f : X \to \mathbb{R}$ a continuous function such that there are $a, b \in X$ with $f(a) < 0 < f(b)$. Then there exists some $c \in X$ with $f(c) = 0$.*

Proof. Since X is connected and f continuous, from Lemma 4.3.2, the image $f(X)$ has to be connected. Now, from Lemma 4.3.1 the image $f(X)$ is an interval. Hence, there must be $c \in X$ such that $f(c) \in (f(a), f(b))$, in particular there is a c such that $f(c) = 0$. □

Definition 4.3.2 (Path connectedness)**.** A topological space (X, τ) is path connected if for any $x, y \in X$ can be connected by a continuous path $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

Lemma 4.3.4. *Every path connected topological space is connected.*

Proof. See Homework B.9.1 □

Note. The converse to Lemma 4.3.4 is not true in general. For instance, the "topologists' sine curve" defined by

$$f(x) = \begin{cases} \sin \frac{1}{x}, & x > 0 \\ 0, & x = 0 \end{cases}$$

has a graph that is connected but not path connected. See Homework B.8.4.

Chapter 5

Differentiation in $\mathbb{R}^n$

5.1 Definition of the derivative

Definition 5.1.1 (Differentiability in $\mathbb{R}^n$)**.** Let $U \in \mathbb{R}^n$ be open, then $f : U \to \mathbb{R}^m$ is called differentiable at $p \in U$ if there is an affine map $T_p : \mathbb{R}^n \mapsto \mathbb{R}^m$ such that

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

If f is differentiable for all $p \in U$, then f is simply called differentiable. If we write the affine function as $T_p(x) = A(x-p) + f(p)$ then

$$A = Df|_p = Df(p) = f'(p)$$

is called the derivative of f at p or the total derivative of f at p.

Note. We can alternatively write $f(x) = A(x-p) + f(p) + r_p(x)$, then f is differentiable at p if and only if

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

Lemma 5.1.1. *If $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is differentiable at p then it is also continuous at p.*

Proof. Clearly, $f(p)$ and $A(x-p)$ are continuous and

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

So also $\lim_{x \to p} \|r_p(x)\| = 0$, i.e., also $r_p(x)$ is continuous at p. □

Lemma 5.1.2. *If $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is differentiable at p, then $Df|_p$ is unique.*

Proof. (by contradiction)

Suppose the derivative is not unique and there are two distinct derivatives with

$$T_p^1(x) = A_1(x-p) + f(p), \quad T_p^2(x) = A_2(x-p) + f(p).$$

Then for some $h = x - p \in \mathbb{R}^n$ we have

$$\begin{aligned}(A_1 - A_2)(h) &= \|(f(p+h) - f(p) - r_1(h+p)) - (f(p+h) - f(p) - r_2(h+p))\| \\ &= \|r_2(h+p) - r_1(h+p)\| \\ &\leq \|r_1(h+p)\| + \|r_2(h+p)\|.\end{aligned}$$

Then $\lim_{h\to 0} \frac{\|(A_1-A_2)(h)\|}{\|h\|} = 0$ if and only if f is differentiable at p. Choose $u \in \mathbb{R}^n$ ($u \neq 0$) and set $h = tu$ for some non-zero $t \in \mathbb{R}$, then

$$\frac{\|(A_1 - A_2)h\|}{\|h\|} = \frac{\|(A_1 - A_2)(tu)\|}{\|tu\|} = \frac{\|(A_1 - A_2)(u)\|}{\|u\|}$$

should go to zero as t tends to zero. But that can only be true for

$$A_1 u = A_2 u \implies A_1 = A_2.$$

□

5.2 Directional and partial derivatives

Definition 5.2.1 (Directional derivative)**.** A function $f : U \mapsto \mathbb{R}^m$ is differentiable in at $p \in U \subset \mathbb{R}^n$ in the direction of the unit vector $u \in \mathbb{R}^n$ if the limit

$$\lim_{t\to 0, t>0} \frac{f_i(p+tu) - f_i(p)}{t}$$

exists for each component f_i of f. The limit is called the *directional derivative* and is denoted by $D_u f(p)$.

Definition 5.2.2 (Partial derivative)**.** The *partial derivative* of some function f differentiable at p is the directional derivative in the direction of the standard basis vectors e_j. The j-th partial derivative is denoted by

$$D_{e_j} f(p) = D_j f(p) = \frac{\partial f(p)}{\partial x_j} = \begin{pmatrix} \frac{\partial f_1(p)}{\partial x_j} \\ \vdots \\ \frac{\partial f_n(p)}{\partial x_j} \end{pmatrix}.$$

Example 5.2.1. Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$ defined by:

$$f(x_1, x_2) = \begin{pmatrix} x_1^2 + x_1 x_2 \\ 2x_1 - x_2^2 \end{pmatrix}.$$

The the partial derivatives are

$$\frac{\partial f}{\partial x_1} = \begin{pmatrix} 2x_1 + x_2 \\ 2 \end{pmatrix}, \quad \frac{\partial f}{\partial x_2} = \begin{pmatrix} x_1 \\ -2x_2 \end{pmatrix}.$$

Theorem 5.2.1. *If $f : U \mapsto \mathbb{R}^m$ for $U \subset \mathbb{R}^n$ is differentiable at $p \in U$ then all directional derivatives exist. And for the unit vector $u \in \mathbb{R}^n$ the directional derivative is given by*

$$\underbrace{Df|_p}_{\in L(\mathbb{R}^n, \mathbb{R}^m)} \cdot \underbrace{u}_{\in \mathbb{R}^n}.$$

In particular, $\frac{\partial f_i}{\partial x_j} = (Df)_{i,j}$ at p.

Example 5.2.2. For the function from Example 5.2.1,

$$Df = \begin{pmatrix} 2x_1 + x_2 & x_1 \\ 2 & -2x_2 \end{pmatrix}.$$

Proof. (Theorem 5.2.1)

Since f is differentiable at p,

$$\begin{aligned} &\lim_{h\to 0} \frac{\|f(p+h) - f(p) - Df|_p \cdot h\|}{\|h\|} = 0 \\ \implies &\lim_{t\to 0, t>0} \frac{\|f(p+tu) - f(p) - Df|_p \cdot tu\|}{\|tu\|} = 0 \\ \implies &\lim_{t\to 0, t>0} \left\| \frac{f(p+tu) - f(p)}{t} - Df|_p \cdot u \right\| = 0 \\ \implies &\lim_{t\to 0, t>0} \frac{f(p+tu) - f(p)}{t} = Df|_p \cdot u. \end{aligned}$$

□

Note. f partially differentiable at x with continuous partial derivatives $\iff$ f totally differentiable and continuous $\implies$ f totally differentiable $\implies$ f differentiable in all directions around x $\implies$ f partially differentiable at x.

Theorem 5.2.2. *Let $f : U \to \mathbb{R}^m$ where $U \subset \mathbb{R}^n$ is open. Then f is partially differentiable at all $p \in U$ with $\frac{\partial f(p)}{\partial x_j}$ continuous for all $1 \le j \le n$ if and only if f is totally differentiable at all p with $Df|_p$ continuous.*

Proof. "$\Longleftarrow$". Fix $p \in U$ and suppose f is differentiable at p with $Df|_p$ continuous i.e., for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\|p - q\| < \delta \implies \|Df|_p - Df|_q\| < \varepsilon$. Then from Theorem 5.2.1 all partial derivatives exist and

$$\begin{aligned} \left| \frac{\partial f_i}{\partial x_j}(p) - \frac{\partial f_i}{\partial x_j}(q) \right| &= \|(Df|_p)_{i,j} - (Df|_q)_{i,j}\| \\ &= \|e_i(Df|_p - Df|_q)e_j\| \\ &\le \|Df|_p - Df|_q\| \le \varepsilon. \end{aligned}$$

Hence the partial derivatives $\frac{\partial f}{\partial x_j}$ are continuous at p.

" $\implies$ ". We show that f is totally differentiable. This implies that $(Df)_{i,j} = \frac{\partial f_i}{\partial x_j}$. Further continuity of $(Df)_{i,j}$ for all i, j implies continuity of (Df).

Now consider a component $f_i : U \mapsto \mathbb{R}$ and fix $p \in U, \varepsilon > 0$. Then continuity of $\frac{\partial f_i}{\partial x_j}$ implies that there is some $\delta > 0$ such that for $q \in U$ and

$$q \in B_\delta(p) \implies \left| \frac{\partial f_i}{\partial x_j}(p) - \frac{\partial f_i}{\partial x_j}(q) \right| < \frac{\varepsilon}{mn}$$

for all $1 \le i \le n, 1 \le j \le m$.

Now, let $h \in \mathbb{R}^m$ with $\|h\| < \delta$ and $h = \sum_1^m h_j e_j$. Define $v_0 = 0, v_1 = h_1 e_1, v_2 = v1 + h_2 e_2, \ldots$ i.e, $v_k = \sum_1^k h_j e_j$. Then

$$f(p+h) - f(p) = \sum_1^m \left(f_i(p + v_j) - f_i(p - v_{j-1}) \right).$$

Note that $0 \leq \|v_j\| \leq \|h\| > \delta$ so the line between in $(p+v_j)$ and $(p+v_{j-1})$ is still in $B_\delta(p)$. Using the one dimensional mean value theorem on the sum,

$$\begin{aligned} f_i(p+v_j) - f_i(p+v_{j-1}) = f_i(p+v_{j-1}+h_j e_j) &= f_i(p+v_{j-1}) \\ &= h_j \frac{\partial f_i}{\partial x_j} \underbrace{(p+v_{j-1}+c_j h_j e_j)}_{\in B_\delta(p)} \end{aligned}$$

for some $c_j \in [0,1]$. Plugging this back into the initial inequality,

$$\begin{aligned} &\left| f_i(p+v_j) - f_i(p+v_{j-1}) - h_j \frac{\partial f_i}{\partial x_j}(p) \right| \\ &\quad = |h_j| \left| \frac{\partial f_i}{\partial x_j}(p+v_{j-1}+c_j h_j e_j) - \frac{\partial f_i}{\partial x_j}(p) \right| \\ &\quad \leq |h_j| \frac{\varepsilon}{mn} \\ \Longrightarrow &\left| f_i(p+v_j) - f_i(p+v_{j-1}) - \sum_1^m h_j \frac{\partial f_i}{\partial x_j}(p) \right| \\ &\quad \leq \frac{\varepsilon}{nm} \sum_1^n \|h\| \leq \frac{\varepsilon}{m} \|h\| \\ \Longrightarrow &\frac{\|f(p+h) - f(p) - Df(p) \cdot h\|}{\|h\|} \\ &\quad \leq \sum_1^m \frac{\left| f_i(p+h) - f_i(p) - \sum_1^n \frac{\partial f_i}{\partial x_j}(p) \cdot h_j \right|}{\|h\|} \\ &\quad \leq \sum_1^m \frac{\varepsilon}{m} = \varepsilon. \end{aligned}$$

Hence f is differentiable at p with $(Df)_{i,j}(p) = \frac{\partial f_i}{\partial x_j}(p)$. □

5.3 The gradient

Definition 5.3.1. Let $f : U \mapsto \mathbb{R}$ where $U \subset \mathbb{R}^n$ is open and f differentiable, then

$$\nabla f := Df = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

is called the *gradient* (∇f the gradient or *nabla* f)[1].

Recall (Maxima/Minima).

- For some function f, if there exists a neighborhood V around p such that $f(p) \geq f(q)$ for all $q \in V$ then p is called a *local maxima* of f (it's the local minima when $f(x) \leq f(q)$ for all $q \in V$).

[1] Since this came up in class: ∇ is called "nabla" as its shape resembles that of a Phonecian harp called (surprise!) "Nabla". The Modern Greek term is *anadelta* as ∇ is just an inverted ("*ana–*", opposite) delta.

- If $f(p) \geq f(q)$ for all q on which the function is defined (the domain of f) then p is the global maximum (it's the global minimum when $f(p) \leq f(q)$).

Theorem 5.3.1. *Let $f : U \to \mathbb{R}$ with $U \subset \mathbb{R}^n$ open and f differentiable. If f has local maxima / minima at $p \in U$ then $\nabla f = 0$. If $\nabla f \neq 0$ then $\frac{\nabla f(p)}{\|\nabla f(p)\|}$ is the unit vector along which f has the largest directional derivative with value $\|\nabla f(p)\|$.*

Proof. Consider $\nabla f \neq 0$ then the directional derivative in the direction of unit vector u is

$$Df \cdot u = \nabla f \cdot u = \|\nabla f\| \|u\| \cos \varphi = \|\nabla f\| \cos \varphi.$$

where $\varphi = \sphericalangle(\nabla f, u)$ the angle between ∇f and u. Clearly, $\nabla f \cdot u$ is maximum when $\cos \varphi = 1 \implies \varphi = 0$. Then $\nabla f \cdot u = \|\nabla f\|$ and $u = \frac{\nabla f}{\|\nabla f\|}$. Also, when $\nabla f \neq 0$ there is a direction along which f increases and opposite to which f decreases. Hence f cannot have an extremum at that point. □

5.4 Higher order derivatives

Definition 5.4.1. A function $f : U \to \mathbb{R}^n$ is called *of class* $\mathcal{C}^k$ (or $f \in \mathcal{C}^k$ or $f \in \mathcal{C}^k(U \to \mathbb{R}^n)$) if all combinations of k-th partial derivatives exist and are continuous.

Note. We write $\frac{\partial}{\partial x_j} \cdot \frac{\partial f}{\partial x_i}$ as $\frac{\partial^2 f}{\partial x_j \partial x_i}$. In general, the partial derivatives are *not* commutative! Generally, $\frac{\partial^2 f}{\partial x_j \partial x_i} \neq \frac{\partial^2 f}{\partial x_i \partial x_j}$, see Homework B.11.2.

Theorem 5.4.1 (Schwarz/Clairaut's theorem[2]). *If $f : U \mapsto \mathbb{R}^m$ with $U \subset \mathbb{R}^n$ open and $f \in \mathcal{C}^2(U \to \mathbb{R}^m)$ then*

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

This is used to prove the more general result

Theorem 5.4.2. *If $f : U \to \mathbb{R}^m$ with $U \subset \mathbb{R}^n$ open and $f \in \mathcal{C}^k(U \to \mathbb{R}^m)$ then, all partial derivatives up-to the k-th order commute.*

For the proof of Theorem 5.4.1 we need the following lemma.

Lemma 5.4.3. *Let $f : U \to \mathbb{R}$ and U open in $\mathbb{R}^2$ such that $\frac{\partial f}{\partial x}$, $\frac{\partial^2 f}{\partial y \partial x}$ exist in U.*

Let $Q \subset U$ be a rectangle with sides parallel to the coordinate axes with its opposite corners given by (p, q) and $(p + h, q + k)$ $(k, h \neq 0)$. Set

$$\Delta(f, Q) = f(p + h, q + k) - f(p, q + k) - f(p + h, q) + f(p, q).$$

Then there exists an interior point of Q with

$$\Delta(f, Q) = hk \frac{\partial^2 f}{\partial y \partial x}(x, y)$$

[2] more precisely, *Clairaut's theorem on equality of mixed partials*

Proof. Define $u(t) := f(t, q+k) - f(t, q)$ for $t \in [p, p+h]$ given $h > 0$. Then, for some $p < x < p+h$ it follows from the intermediate value theorem that

$$\begin{aligned}\Delta(f, Q) &= u(p+h) - u(p)\\ &= hu'(x)\\ &= h\left(\frac{\partial f}{\partial x}(x, q+k) - \frac{\partial f}{\partial x}(x, q)\right)\\ &= hk\left(\frac{\partial}{\partial y}\frac{\partial}{\partial x}f(x, y)\right) \qquad \text{for } q < y < q+k\end{aligned}$$

where we obtain the last step by applying the intermediate value on the second variable. □

This allows us to prove Theorem 5.4.1.

Proof (Clairaut; Theorem 5.4.1). We consider the case when $m = 1$ (otherwise, take each component separately and $n = 2$ (or, keep all but two variables fixed)). Choose $p, q \in U$ and set $A := \frac{\partial^2 f}{\partial y \partial x}(p, q)$.

For some (arbitrary) fixed $\varepsilon > 0$ choose Q as in Lemma 5.4.3. Then with $h, k > 0$ chosen small enough, we get from continuity that for all $(x, y) \in Q$

$$\begin{aligned}&\left|A - \frac{\partial^2 f}{\partial y \partial x}(x, y)\right| < \varepsilon\\ \overset{5.4.3}{\Longrightarrow} &\left|A - \frac{\Delta(f, Q)}{hk}\right| < \varepsilon\\ \overset{k \to 0}{\Longrightarrow} &\left|A - \frac{\frac{\partial f}{\partial y}(p+h, q) - \frac{\partial f}{\partial y}(p, q)}{h}\right| < \varepsilon\\ \overset{h \to 0}{\Longrightarrow} &\left|A - \frac{\partial^2 f}{\partial x \partial y}(p, q)\right| < \varepsilon.\end{aligned}$$

□

Definition 5.4.2 (Hessian Matrix). Let $f : U \mapsto \mathbb{R}$ and $U \subset \mathbb{R}^n$ open and $f \in \mathcal{C}^2$. The *Hessian Matrix* (or just the Hessian) is the matrix H of second derivatives whose terms are given by $H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

Note. Theorem 5.4.1 implies that the Hessian is symmetric.

Theorem 5.4.4 (Second order Taylor). *Let $f : U \to \mathbb{R}$ with $U \subset \mathbb{R}^n$ open and $f \in \mathcal{C}^2$. Let $p \in U$ and $h \in \mathbb{R}^n$ such that $p + th \in U$ for all $t \in [0, 1]$. If we write*

$$f(p+h) = f(p) + (Df)(p) \cdot h + \underbrace{\frac{1}{2}h^T \cdot H_f(p) \cdot h}_{\frac{1}{2}\sum_{i,j} h_i \cdot (H_f(p))_{i,j} \cdot h_j} + r_p(h)$$

then $\lim_{h \to 0} \frac{|r_p(h)|}{\|h\|^2} = 0$.

Proof. Define $g : [0,1] \to \mathbb{R}$ such that $g(t) = f(p+th)$ since $f \in \mathcal{C}^2$ also $g \in \mathcal{C}^2$. Then, from one-dimensional Taylor expansion around 0, we have

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(\tau) \qquad \text{for some } \tau \in [0,1].$$

Now from the definition of $g(t)$ we have

$$\begin{aligned}
g(1) &= f(p+h), \\
g(0) &= f(p), \\
g'(t) &= (Df)(p+th) \cdot h = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p+th)h_i, \\
g''(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p+th)h_i = \sum_{i=1}^{n} h_i \left(D\frac{\partial f}{\partial x_i} \right)(p+th) \cdot h \\
&= \sum_{i,j} h_i \frac{\partial^2 f}{\partial x_i \partial x_j}(p+th)h_j = h^T \cdot H_f(p+th) \cdot h.
\end{aligned}$$

Plugging these into the expansion of $g(1)$,

$$f(p+h) = f(p) + (Df)(p+th) + \frac{1}{2}h^T \cdot H_f(p+\tau h) \cdot h.$$

Hence, for the required rest term $r_p(h)$,

$$\begin{aligned}
r_p(h) &= f(p+h) - f(p) - (Df)(p) \cdot h - \frac{1}{2}h^T \cdot H_f(p) \cdot h \\
&= \frac{1}{2}h^T \cdot H_f(p+\tau h) \cdot h - \frac{1}{2}h^T \cdot H_f(p) \cdot h \\
&= \frac{1}{2}h^T \cdot (H_f(p+\tau h) - H_f(p)) \cdot h \\
\implies |r_p(h)| &\leq \frac{1}{2}\|h\|^2 \|H_f(p+\tau h) - H_f(p)\|.
\end{aligned}$$

Since $f \in \mathcal{C}^2$, H_f is continuous. That is for given ε and $\|h\| < \delta$,

$$\frac{|r_p(h)|}{\|h\|^2} = \frac{1}{2}\|H_f(p+\tau h) - H_f(p)\| < \frac{\varepsilon}{2}.$$

Hence, $\lim_{h\to 0} \frac{|r_p(h)|}{\|h\|^2} = 0$ as stated. □

Theorem 5.4.5. *Let $f : U \to \mathbb{R}^n$ with U open be a $\mathcal{C}^2$ function with $\nabla f(p) = Df(p) = 0$ for some $p \in U$. Then*

- *if $H_f(p)$ is positive definite i.e., for all $h \in \mathbb{R}^n$ with $h \neq 0, h^T \cdot H_f(p) \cdot h > 0$, then f has a (local) minimum at p.*
- *if $H_f(p)$ is negative definite i.e., for all $h \in \mathbb{R}^n$ with $h \neq 0, h^T \cdot H_f(p) \cdot h < 0$, then f has a (local) maximum at p.*

Proof. See Homework B.12.2. □

Note. When a matrix A is symmetric then it is positive definite if and only if all of its eigenvalues $\lambda_k > 0$ as

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \implies h^T \cdot A \cdot h = \sum_1^n \lambda_k |h_k|^2.$$

5.5 Inverse function theorem

Motivation We're interested in conditions under which some function $f : W \to \mathbb{R}^n$ where $W \subseteq \mathbb{R}^n$ is invertible. We know that for $n = 1$, if the continuously differentiable function f with some $f'(p) \neq 0$ then $f'(p)$ is invertible in some neighborhood of p. This implies that f is invertible in the neighborhood of p and f^{-1} is continuously differentiable. And hence,

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

holds. Invertibility of $f'(p)$ is the crucial condition here. If possible, we would like to generalize this to $\mathbb{R}^n$.

Theorem 5.5.1 (Inverse function theorem)**.** *Let $f : W \to \mathbb{R}^n$ be a $\mathcal{C}^1$ function with $W \subseteq \mathbb{R}^n$ open. If the $n \times n$ matrix $Df(p)$ is invertible for some $p \in W$ then*

1. *There are open neighborhoods U of p and V of $f(p)$ such that the restriction of the function to $U, f|_U : U \to V$ is bijective and thus $f|_U$ invertible.*
2. *If g is the inverse of $f|_U$ i.e., $g(f|_U(x)) = x$ for all $x \in U$ then $g \in \mathcal{C}^1$.*

Put differently, theorem 5.5.1 states that $y = f(x_1, \dots, x_n)$ can be solved for $x_1, \dots, x_n$ in terms of $y_1, \dots, y_n$ if x and y lie in small enough neighborhoods U and V respectively.

Note (Diffeomorphism). If a $\mathcal{C}^k$ function $f : W \to V$ has an inverse f^{-1} with $f^{-1} \in \mathcal{C}^k$, then f is called a $\mathcal{C}^k$-diffeomorphism. Further if $p \in W$ has a neighborhood U such that the restriction $f|_U$ is a $\mathcal{C}^k$-diffeomorphism, then f is called a *local* $\mathcal{C}^k$-diffeomorphism.

Note (Convexity). A set U is convex if and only if all points on the straight line between any $p, q \in U$ are also in U.

A complete proof of Theorem 5.5.1 requires two smaller lemmas which we now prove.

Lemma 5.5.2. *Let $f : U \to \mathbb{R}^n$ with $U \subseteq \mathbb{R}^n$ open and convex. If f is differentiable and $\|Df(x)\| \leq M$ for some $M > 0$ and $x \in U$, then*

$$\|f(p) - f(q)\| \leq M\|p - q\|$$

for all $p, q \in U$.

Proof. Since U is convex, some curve $\gamma : [0,1] \to \mathbb{R}^n$ defined by $\gamma(t) = tq + (1-t)p$ is in U whenever $p, q \in U$. Let $g(t) = f(\gamma(t))$. Then,

$$\begin{aligned} f(q) - f(p) &= f(\gamma(1)) - f(\gamma(0)) \\ &= g(1) - g(0) \\ &= \int_0^1 g'(t)\,\mathrm{d}t \qquad \text{(integrated component wise);} \\ g'(t) &= f'(\gamma(t)) \cdot \gamma'(t) = Df(\gamma(t)) \cdot (q-p) \\ \implies \|f(q) - f(p)\| &\le \int_0^1 \|g'(t)\|\,\mathrm{d}t \\ &\le \|q-p\| \int_0^1 \underbrace{\|Df(\gamma(t))\|}_{\le M}\,\mathrm{d}t \\ &\le M\|q-p\| \int_0^1 \mathrm{d}t = M\|q-p\|. \end{aligned}$$

□

Lemma 5.5.3. 1. *If $A, B : \mathbb{R}^n \to \mathbb{R}^n$ are linear maps where A is invertible and $\|A - B\| < \frac{1}{\|A^{-1}\|}$, then B is invertible as well (i.e., the set of all invertible maps is open).*

2. *Taking the inverse map is a continuous operation i.e., $\lim_{B\to A} B^{-1} = A^{-1}$.*

Proof (heuristics). We would like to use a geometric series to show that B is invertible. In particular, we compute the inverse using[3]

$$\begin{aligned} B^{-1} &= \frac{1}{A - A + B} \\ &= \frac{1}{A(I - (I - A^{-1}B))} \\ &= \left(\sum_0^\infty (I - A^{-1}B)^n\right) A^{-1} \end{aligned}$$

□

Proof. 1. Using the given bound $\|A - B\| < \frac{1}{\|A^{-1}\|}$, we get

$$\|1 - A^{-1}B\| = \|A^{-1}(A - B)\| \le \|A^{-1}\|\|A - B\| < 1.$$

Hence, the sum

$$\sum_{k=0}^N (1 - A^{-1}B)^k$$

is a Cauchy sequence and converges as $N \to \infty$.

[3] Note that $I = \mathbf{1} = \mathbb{1}_{\mathbb{R}^n\times\mathbb{R}^n} = \mathrm{id}_{n\times n}$ all denote the identity matrix of size n.

Then,

$$\begin{aligned}
&\sum_{k=0}^{N}(I-A^{-1}B)^k \underbrace{\Big(I-(I-A^{-1}B)\Big)}_{A^{-1}B} \\
&\qquad = \sum_{0}^{N}(I-A^{-1}B)^k - \sum_{0}^{N}(I-A^{-1}B)^{k+1} \\
&\qquad = \sum_{0}^{N}(I-A^{-1}B)^k - \sum_{1}^{N+1}(I-A^{-1}B)^{k} \\
&\qquad = I-(I-A^{-1}B)^{N+1}.
\end{aligned}$$

Now, if we consider the norm of this sum,

$$\begin{aligned}
\left\| \sum_{0}^{N}(I-A^{-1}B)^k A^{-1}B - I \right\| &= \|(I-A^{-1}B)^{N+1}\| \\
&\leq \underbrace{\|I-A^{-1}B\|}_{<1}{}^{N+1} \xrightarrow{N\to\infty} 0.
\end{aligned}$$

$$\implies \underbrace{\sum_{k=0}^{\infty}(I-A^{-1}B)^k A^{-1}}_{B^{-1}} B = I.$$

Hence, B is invertible.

2. If $B \to A$ then $A^{-1}B \to I$ as multiplication by a fixed A^{-1} is continuous. Then

$$\lim_{B\to A}\sum_{k=0}^{\infty}(I-A^{-1}B)^k \underbrace{A^{-1}B}_{\to I} = I, \quad \text{i.e.,} \quad \lim_{B\to A}\sum_{k=0}^{\infty}(I-A^{-1}B)^k = I,$$

so

$$\lim_{B\to A} B^{-1} = \sum_{k=0}^{\infty}(I-A^{-1}B)^k A^{-1} = A^{-1}.$$

□

Proof (Theorem 5.5.1). 1. Let $A = Df(p)$ and set $\lambda = \frac{1}{2\|A^{-1}\|}$. Since f is continuously differentiable at p, there exists an open and convex neighborhood of p such that

$$\|Df(p) - Df(x)\| < \lambda \quad \forall x \in U. \tag{$*$}$$

Since we want to show that the function is bijective on some restriction, we first show injectivity i.e., for some fixed $y \in \mathbb{R}^n$ we want at most one x such that $f(x) = y$.

Define $\varphi_y : W \to \mathbb{R}^n$ where $\varphi_y(x) = x + A^{-1}(y - f(x))$ (this choice of φ_y is motivated by Newton's method where we use the iteration scheme $\tilde{x} = x + \frac{f(\tilde{x})-f(x)}{f'(x)} = \varphi_y(x)$). Hence

$$f(x) = y \iff \phi_y(x) = x.$$

That is, x is a fixed point of φ_y.

Now,

$$D\varphi_y(x) = 1 - A^{-1}Df(x) = A^{-1}(A - Df(x))$$
$$\implies \|D\varphi_y(x)\| \leq \underbrace{\|A^{-1}\|}_{=\frac{1}{2\lambda}} \underbrace{\|A - Df(x)\|}_{<\lambda} < \frac{1}{2}$$

With Lemma 5.5.2 this implies that

$$\|\varphi_y(x_1) - \varphi_y(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|. \qquad (**)$$

Hence, $\varphi_y(x)$ is a contraction $\implies$ $\varphi_y(x)$ has *at most* one fixed point (assuming two fixed points x_1, x_2 we get $\varphi_y(x_1) = x_1, \varphi_y(x_2) = x_2 \implies \|\varphi_y(x_1) - \varphi_y(x_2)\| = \|x_1 - x_2\| \leq \frac{1}{2}\|x_1 - x_2\|$ ↯ which contradicts $(**)$).

Taking this fixed point x, we have $\varphi_y(x) = x \implies f(x) = y$ for at most one x. Hence, $f|_U$ is injective.

We would now like to show that $V = f(U)$ is open.

To get the existence of the fixed point we need to find a complete metric space X such that $\varphi_y : X \to X$ i.e., we must find X closed so that we can apply the Banach fixed point theorem.

Let $f(U) = V$ and pick $q \in V$ then there is a unique p such that $f(p) = q$ (from injectivity). Now, define $B = B_r(p)$ with r so small that $B \subset \bar{B} \subset U$. We show that if we choose $y \in B_{r\lambda}(q)$ then $y \in V$, i.e., V is open.

Let $y \in B_{r\lambda}(q)$ then,

$$\begin{aligned}
\|\varphi_y(p) - p\| &= \|p + A^{-1}(y - f(p)) - p\| \\
&= \|A^{-1}(y - f(p))\| \\
&\leq \underbrace{\|A^{-1}\|}_{=\frac{1}{2\lambda}} \underbrace{\|y - f(p)\|}_{<r\lambda} \\
&< \frac{r}{2}.
\end{aligned}$$

Set $V = f(U)$ and pick $q \in V$ such that there is a unique p with $f(p) = q$ (we get this from injectivity at q). Now, define $B = B_r(p)$ with r so small such that $\bar{B} \subset U$. We show that if $y \in B_{r\lambda}(q) \implies y \in V$, so that V is open. For any $x \in \bar{B}$

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

As $\bar{B}$ is a complete metric space and $\varphi_y(x) \in \bar{B} \implies \varphi_y : \bar{B} \to \bar{B}$ is a contraction. Then from the Banach fixes point theorem there is a unique fixed point with $\varphi_y(x) = x \implies f(x) = y$ and

$$y \in f(\bar{B}) \subset f(U) = V.$$

2. We show that $g = f^{-1} : V \to U$ is $\mathcal{C}^1$. First choose $y, y+k \in V$ then there are $x, x+h \in U$ such that $f(x) = y$ and $f(x+h) = y+k$. From Lemma 5.5.3, since

$$\|Df(x) - Df(p)\| < \frac{1}{2\|Df(p)\|}$$

then $Df(x)$ has an inverse $T = (Df(x))^{-1}$. Then

$$\begin{aligned}
g(y+k) - g(y) - Tk &= x + h - x - Tk \\
&= h - Tk \\
&= TT^{-1}h - T(f(x+h) - y) \\
&= -T(f(x+h) - f(x) - Df(x) \cdot h) \\
\implies \frac{\|g(y+k) - g(y) - Tk\|}{\|k\|} &\leq \|T\| \frac{\|f(x+h) - f(x) - Df(x) \cdot h\|}{\|k\|}.
\end{aligned}$$

Now, if we somehow relate h and k we're done. So, consider

$$\begin{aligned}
\varphi_y(x+h) - \varphi_y(x) &= x + h + A^{-1}(y - f(x+h)) \\
&\quad - (x + A^{-1} \underbrace{(y - f(x))}_{=0}) \\
&= h + A^{-1}(y - f(x+h)) \\
&= h - A^{-1}k; \\
\implies \|h - A^{-1}k\| &= \|\varphi_y(x+h) - \varphi_y(x)\| \\
&\leq \frac{1}{2}\|h\|; \\
\implies \|h\| &\leq \|h - A^{-1}k + A^{-1}k\| \\
&\leq \underbrace{\|h - A^{-1}k\|}_{\leq \frac{1}{2}\|h\|} + \|A^{-1}k\| \\
\implies \|h\| &\leq 2\|A^{-1}k\| \leq 2\frac{1}{2\lambda}\|k\| = \frac{\|k\|}{\lambda}; \\
\implies \frac{\|g(y+k) - g(y) - Tk\|}{\|k\|} &\leq \frac{\|T\|}{\lambda} \cdot \frac{\|f(x+h) - f(x) 0 Df(x) \cdot h\|}{\|h\|}.
\end{aligned}$$

Clearly, as $k \to 0$, $k \to 0$ as well. Hence, the right hand side approaches zero since $f \in \mathcal{C}^1$. Then, g is differentiable at y and $Df(y) = (Df(x))^{-1}$. Finally

- $x \mapsto Df(x)$ continuous as $f \in \mathcal{C}^1$.
- $A \mapsto A^{-1}$ continuous from Lemma 5.5.3.
- $y \mapsto x$ continuous as g is differentiable (and hence continuous).

Therefore $y \mapsto Dg(y)$ is continuous and $g \in \mathcal{C}^1$.

□

5.6 Implicit function theorem

Motivation We are given a set of n equations in n unknowns

$$\begin{aligned} f_1(x_1, \dots, x_n, y_1, \dots, y_m) &= 0, \\ f_2(x_1, \dots, x_n, y_1, \dots, y_m) &= 0, \\ &\vdots \\ f_n(x_1, \dots, x_n, y_1, \dots, y_m) &= 0, \end{aligned}$$

and we would like to solve for all x_k. In particular, we are interested in solutions of the form $x_1(y_1, \dots, y_m), \dots, x_n(y_1, \dots, y_m)$. This might not be possible globally, so we want to see if local solutions to the equations exist.

Example 5.6.1. Consider the function $f(x, y) = x^2 + y^2 - 1$ for $x, y \in \mathbb{R}$. Solving for $f(x, y) = 0$ is the same as solving for $x^2 + y^2 = 1$. Since this equation describes a circle, we won't have a global solution. However, it should be possible to locally solve for $x(y)$ in open neighborhoods (except possibly at $x = 0$). We know that

$$\frac{\partial f}{\partial x} = 2x, \qquad \frac{\partial f}{\partial y} = 2y.$$

Note that when $x = 0$,

$$\left.\frac{\partial f}{\partial x}\right|_{(x=0,y)} = 0 \implies \frac{\partial f}{\partial x} \text{ is not invertible.}$$

Note. We denote the tuple of x_k-s with $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Similarly, $y = (y_1, \dots, y_m) \in \mathbb{R}^m$ and $(x, y) \in \mathbb{R}^{m+n}$.

Theorem 5.6.1. *Let $E \subset \mathbb{R}^{m+n}$ be open. Let $f : E \to \mathbb{R}^n$ be $\mathcal{C}^1$ such that $f(p, q) = 0$ for some $(p, q) \in E$.*

Assuming $\frac{\partial f}{\partial x}(p, q)$ is invertible there are open sets $U \subset \mathbb{R}^{m+n}, W \subset \mathbb{R}^n$ with $(p, q) \in U, q \in W$ such that

- *For all $y \in W$ there is a unique x such that $(x, y) \in U$ and $f(x, y) = 0$.*
- *Defining this $x := g(y)$, then $g : W \to \mathbb{R}^m$ is $\mathcal{C}^1, g(q) = p$, and $f(g(y), y) = 0$ for all $y \in W$.*
- *The derivative of g at q is given by*

$$Dg(q) = -\left(\frac{\partial f}{\partial x}(p, q)\right)^{-1} \frac{\partial f}{\partial y}(p, q). \qquad (\star)$$

Note. At (p, q),

$$(\star) \implies \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} \cdot \frac{\partial g_j}{\partial y_k} = -\frac{\partial f_i}{\partial y_k}.$$

Example 5.6.2. Continuing with Example 5.6.1, $f(x, y) = x^2 + y^2 - 1, x = g(y) = \sqrt{1 - y^2}$. then implicitly we get

$$\frac{\partial g}{\partial y} = -\frac{1}{2x}\, 2y = -\frac{y}{x} = -\frac{y}{g(y)}.$$

Computing the derivative directly we get

$$\frac{\partial g}{\partial y} = -\frac{2y}{2\sqrt{1-y^2}} = -\frac{y}{\sqrt{1-y^2}} = -\frac{y}{g(y)}.$$

Proof (Theorem 5.6.1). We define $F : E \to \mathbb{R}^{n+m}$ by $F(x, y) = (f(x, y), y)$. Then, $F(p, q) = (0, q)$ and

$$\begin{aligned} DF(p,q) &= \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ 0 & \mathbb{1} \end{pmatrix} \\ \det(DF)(p,q) &= \det\left(\frac{\partial f}{\partial x}\right) \cdot \det \mathbb{1} \neq 0. \qquad \text{(by assumption)} \\ \implies & DF(p,q) \text{ is invertible.} \end{aligned}$$

By the intermediate value theorem, there are open neighborhoods U of (p, q) and V of $F(p, q) = (0, q)$ such that $F : U \to V$ is a diffeomorphism with a $\mathcal{C}^1$ inverse $G : V \to U$. Setting

$$W := \{y \in \mathbb{R}^m : (0, y) \in V\}$$

we get $q \in W \implies W$ is open. If $y \in W$ then $(0, y) \in V$, so $(0, y) = F(x, y)$ for some $(x, y) \in U$ and $f(x, y) = 0$.

From the injectivity of F we know that this x is unique, hence, this procedure defines a unique $g : W \to \mathbb{R}^n$ such that $g(y)$ is the unique $x \in \mathbb{R}^n$ defined above.

With $(x, y) \in U$ and $f(g(y), y) = 0$,since $(p, q) \in U$ then $f(p, q) = 0$ and consequently $g(q) = p$.

Now, we just need to show that $g \in \mathcal{C}^1$. Use $G(0, y) = (g(y), y) \implies G \in \mathcal{C}^1$ and use $(\star)$ with the chain rule for $(g(y), y)$. □

Appendix A

Schedule

Feb. 05, 2018 Riemann–Stieltjes Integral (definition)

Feb. 06, 2018 Riemann–Stieltjes Integral (refinement of partitions, existence criterion)

Feb. 12, 2018 Riemann–Stieltjes Integral (continuous functions, finitely many discontinuities, basic properties)

Feb. 13, 2018 Riemann–Stieltjes Integral (compositions, change of variables)

Feb. 19, 2018 Riemann–Stieltjes Integral (change of variables, fundamental theorem)

Feb. 20, 2018 Riemann–Stieltjes Integral (integration by parts, improper integrals, uniform convergence and integration and differentiation)

Feb. 26, 2018 Convergence, Series, Sequences (review, convergence tests)

Feb. 27, 2018 Convergence, Series, Sequences (more convergence tests, rearrangements)

Mar. 05, 2018 Power Series (Cauchy product, radius of convergence; derivative, integral and Cauchy product of power series)

Mar. 06, 2018 Curves and Differential Equations (definition of curves in $\mathbb{R}^n$, differentiability, speed limit)

Mar. 12, 2018 Curves and Differential Equations (integrability, rectifiability and curve length)

Mar. 13, 2018 Curves and Differential Equations (rectifiability, reparametrization; intro to differential equations, vector fields, initial value problem)

Mar. 19, 2018 Midterm Exam

Mar. 20, 2018 Curves and Differential Equations (Lipschitz condition, Picard–Lindelöf, separation of variables)

Mar. 26, 2018 *no class (Spring Break)*

Mar. 27, 2018 *no class (Spring Break)*

Apr. 02, 2018 *no class (Spring Break)*

Apr. 03, 2018 Basic Topology (definition, continuity)

Apr. 09, 2018 Basic Topology (compactness)

Apr. 10, 2018 Basic Topology (Heine–Borel, connectedness)

Apr. 16, 2018 Basic Topology (connectedness, path connectedness)

Apr. 17, 2018 Differentiation in $\mathbb{R}^n$ (definitions of total derivative, directional derivative, partial derivative)

Apr. 23, 2018 Differentiation in $\mathbb{R}^n$ (connections between total derivative, directional derivative, partial derivative)

Apr. 24, 2018 Differentiation in $\mathbb{R}^n$ (equivalence of total continuous differentiability and continuous partial differentiability, the gradient and extrema)

Apr. 30, 2018 Differentiation in $\mathbb{R}^n$ (higher order derivatives and second order Taylor expansion)

May 01, 2018 *no class (Labor Day)*

May 07, 2018 Differentiation in $\mathbb{R}^n$ (Inverse Function Theorem)

May 08, 2018 Differentiation in $\mathbb{R}^n$ (Implicit Function Theorem)

May 14, 2018 Riemann Integral in $\mathbb{R}^n$ (definition and integrability criteria)

May 15, 2018 Riemann Integral in $\mathbb{R}^n$ (Fubini and change of variables)

May 29, 2018 **Final Exam**

Appendix B

Homeworks

B.1 2018-02-19

B.1.1 [8 points] The Stieltjes integral for discontinuous α

Let $\alpha(x) = 0$ if $x \leq 0$ and $\alpha(x) = 1$ if $x > 0$. Give a precise proof that $\int_{-1}^{1} f \,\mathrm{d}\alpha = f(0)$ for every function $f : \mathbb{R} \mapsto \mathbb{R}$ that is continuous at $x = 0$.

B.1.2 [4 points] Explicit Stieltjes Integral

Let $a < b \in (0, 4)$. Find a monotonically increasing bounded function $\alpha : \mathbb{R} \mapsto \mathbb{R}$ such that

$$\int_0^4 f \,\mathrm{d}\alpha = f(1) + 2f(2) + 3f(3) + \frac{1}{2}\int_a^b f(x)\,\mathrm{d}x$$

for all f for which the integral exists.

B.1.3 [10 points] Integrable and non-integrable functions

Define two functions $f, g : [0, 1] \mapsto \mathbb{R}$ via:

$$f(x) := \begin{cases} 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q}, \\ 1/q & \text{if } x = p/q, \text{ with } p, q \text{ coprime}, \end{cases}$$

$$g(x) := \begin{cases} 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q}, \\ 1 & \text{if } x \in \mathbb{Q}. \end{cases}$$

1. Show that $f \in \mathcal{R}[0, 1]$ and $\int_0^1 f(x)\,\mathrm{d}x = 0$
2. Show that g is not Riemann-integrable.

B.1.4 [10 points] Riemann integrable or not?

1. Show that $f(x) = \mathrm{e}^x$ is Riemann integrable on $[a, b] \subset \mathbb{R}$. What is the value of $\int_a^b \mathrm{e}^x\,\mathrm{d}x$?
2. Show that $f(x) = c$ ($c \in \mathbb{R}$) and $g(x) = x$ are Riemann integrable on $[a, b] \subset \mathbb{R}$. Find

$$\int_a^b c\,\mathrm{d}x \quad \text{and} \quad \int_a^b x\,\mathrm{d}x.$$

3. Show that $f(x) = x^n$ is Riemann integrable on $[0, a] \subset \mathbb{R}$ for every $n \in \mathbb{N}$. What is the value of $\int_0^a x^n \, dx$? *(Hint: You may use the fact that $\sum_{k=1}^N k^n$ is a polynomial in N of degree $n+1$ and with leading coefficient $1/(n+1)$, or — if you know it — use the Stolz-Ceasáro theorem.)*

B.1.5 [8 points] Uniform Continuity

Let $I \subset \mathbb{R}$ be an interval. A function $f : I \mapsto \mathbb{R}$ is called *uniformly continuous* if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, x' \in I$ with $|x - x'| < \delta$ we have that $|f(x) - f(x')| < \varepsilon$ (in other words, the δ does not depend on x).

1. Show that if $I = [a, b]$ is closed and bounded, then every continuous function $f : [a, b] \mapsto \mathbb{R}$ is uniformly continuous. *(Hint: In general, δ may depend on x and thus defines a function $\delta(x)$; show that $\delta(x)$ is continuous.)*

2. Does the answer change if I is no longer closed, and/or no longer bounded?

B.1.6 (Bonus) [3 points] Uniform Continuity Continued

We continue Problem B.1.5.

1. Suppose $f, g : X \mapsto \mathbb{R}$ are uniformly continuous on $X \subset \mathbb{R}$. Is it true that $f + g$ is uniformly continuous on X? How about $f \cdot g$?

2. Does the answer to (1) change if f, g are bounded?

3. Does the answer to (1) change if X is a closed interval?

B.1.7 (Bonus) [5 points] Devil's staircase and Stieltjes integrals

Define a function $\alpha : [0, 1] \mapsto \mathbb{R}$ as Follow: Given $x \in [0, 1]$, write $x = \sum_{i \geq 1} b_i 3^{-i}$ with $b_i \in \{0, 1, 2\}$ (representation of x in base 3). Let n be minimal with $b_n = 1$ (or $n = \infty$ if all $b_i \neq 1$). Then $\alpha(x) := \sum_{i=1}^n a_i 2^{-i}$, where $a_i = 1$ if $b_i \in \{1, 2\}$ and $a_i = 0$ if $b_i = 0$. (For additional credit, you may show check that the value of α is well defined even at points x that have two representations in base 3).

1. Sketch the graph of α.

2. Show that α is continuous and monotonically increasing.

3. Show that for every $\varepsilon > 0$, there are finitely many intervals $I_{\varepsilon,1}, I_{\varepsilon,2}, \ldots, I_{\varepsilon,n}$ with total length ε so that α is constant on $[0, 1] \setminus \cup_{i=1}^n I_{\varepsilon,i}$ (this means that α is constant except on a set of volumen zero, but α is continuous and non-constant).

4. Show that the integrals $\int_0^1 1 \, d\alpha$ and $\int_0^1 x \, d\alpha$ exist, and determine their values. *(Hint: Show that $\int_0^1 (x - 1/2) \, d\alpha = 0$.)*

B.2 2018-02-26

B.2.1 [15 points] Properties of the Riemann–Stieltjes integral

B.2.2 [8 points] When the integral is zero

B.2.3 [8 points] Simultaneous discontinuity of f and α

B.2.4 [9 points] Monotone functions are integrable

B.2.5 (Bonus) [3 points] Sum of α's

B.2.6 (Bonus) [3 points] Integration of composition

B.3 2018-03-05

B.3.1 [10 points] Partial fractions

B.3.2 [6 points] Integration by substitution

B.3.3 [18 points] Lots of integrals ...

B.3.4 [6 points] Uniform convergence of second derivatives

B.3.5 (Bonus) [4 points] Null sets

B.3.6 (Bonus) [4 points] A criteria for Riemann integrability

B.4 2018-03-12

B.4.1 [15 points] Convergent and divergent series

B.4.2 [8 points] Taylor series of logarithm

B.4.3 [6 points] Dirichlet's test on convergence of series

B.4.4 [14 points] The Riemann zeta function

B.4.5 (Bonus) [4 points] Abel's test on convergence of series

B.5 2018-04-03

B.5.1 [8 points] More about convergence: root and ratio test

- State carefully the root and the ratio test for series: in which cases do they imply convergence (absolute or conditional?) or divergence, in which cases are they inconclusive?
- Give three examples that were not yet treated in class: one each where the root test proves convergence of a series, where it proves divergence, and where it is not conclusive.
- Same for the ratio test (different examples, please).

B.5.2 [8 points] More about convergence

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers. Show that convergence of $\sum_n a_n$ implies convergence of $\sum_n \frac{\sqrt{a_n}}{n}$.

B.5.3 [16 + 3 points] More about convergence: infinite products

B.5.4 [8 + 5 points] More about convergence: Dirichlet series

B.6 2018-04-03

B.6.1 [12 points] Rectifiable graphs and their arclength

B.6.2 [12 + 2 points] The snowflake: a non-rectifiable curve (the "von-Koch"–curve)

B.6.3 [16 points] The hyperbolic metric

B.6.4 (Bonus) [2 points] Devil's staircase

B.6.5 (Bonus) [4 points] Space–Filling curves

B.7 2018-04-10

B.7.1 [20 points] Picard–Lindelöf

B.7.2 [8 + 3 points] The exponential differential equation

B.7.3 [12 points] Separation of variables

B.7.4 (Bonus) [5 points] Exponential separation for Lipschitz vector fields

B.8 2018-04-17

Note: This homework sheet is not so hard once you get used to compactness.

B.8.1 [5 points] Subspace topology

Let (X, τ_X) be a topological space and $Y \subset X$. Prove that

$$\tau_Y = \{U_i \cap Y : U_i \in \tau_X\}$$

is indeed a topology on Y. *Note: This is trivial, but to get familiar with topologies, carefully write down a nice formal proof.*

B.8.2 [10 points] Continuity and compactness

Prove that the continuous image of any compact set is compact.

B.8.3 [25 points] Some compactness lemmas

Prove the following lemmas about compactness:

1. Every closed subset of a compact topological space is compact.
2. In a Hausdorff space, every compact set is closed.
3. Every continuous bijective map from a compact set to a Hausdorff space has a continuous inverse (i.e., is a homeomorphism)

B.8.4 (Bonus) [8 points] Connectedness and path connectedness

B.9 2018-24-2018

B.9.1 [10 points] Connectedness

Prove that every path connected topological space is connected.

Hint: Recall the definitions of both path connectedness and connectedness. One strategy is to assume that the space is path connected but not connected, and derive a contradiction.

B.9.2 [10 points] Uniform continuity and compactness

B.9.3 [10 points] Partial derivatives

B.9.4 [10 points] Derivatives, partial derivatives, and continuity

B.9.5 (Bonus) [8 points] Banach fixed point theorem

B.10 2018-05-01

B.10.1 [12 points] The chain rule

B.10.2 [14 points] Directional derivatives and total derivatives

B.10.3 [14 points] Local maxima and the gradient

B.10.4 [8 points] Length

B.11 2018-05-08

B.11.1 [8 points] Cauchy–Riemann differential equations and harmonic functions

B.11.2 [18 points] Twice differentiable

B.11.3 [6 points] Second order Taylor

B.11.4 [8 points] The wave equation

B.12 2018-05-15

B.12.1 [10 points] Very basic vector calculus

B.12.2 [12 points] Maxima and minima

B.12.3 [10 points] Two-dimensional polar coordinates

B.12.4 [8 points] Three dimensional polar coordinates

B.12.5 (Bonus) [8 points] Newton's method in several variables