Mathematical Analysis II, 2018/19 First semester

Yoh Tanimoto
Dipartimento di Matematica, Università di Roma Tor Vergata
Via della Ricerca Scientifica 1, I-00133 Roma, Italy
email: `hoyt@mat.uniroma2.it`

We basically follow the textbook "Calculus" Vol. I,II by Tom M. Apostol, Wiley.
Lecture notes:
`http://www.mat.uniroma2.it/~tanimoto/teaching/2019MA2/2019MathematicalAnalysisII.pdf`

Summary of the course:

- Sequences and series of functions, Taylor series
- Differential calculus of scalar and vector fields
- Applications of differential calculus, extremal points
- Basic differential equations
- Line integrals
- Multiple integrals
- Surface integrals, Gauss and Stokes theorems

Sep 23. Pointwise and uniform convergence

Mathematical Analysis I and II

In Mathematical Analysis I we learned:

- sequence of numbers $a_1, a_2, \cdots$
- functions $f(x)$ on $\mathbb{R}$: limit $\lim_{x\to a} f(x)$, derivative $f'(x) = \frac{df}{dx}(x)$, integral $\int_a^b f(x)dx$.

In Mathematical Analysis II we will learn:

- sequence of numbers $f_1(x), f_2(x), \cdots$
- functions $f(x,y)$ on $\mathbb{R}^2$, and functions on $\mathbb{R}^n$, vector fields $\boldsymbol{F}(x_1, x_2, \cdots, x_n)$: partial derivatives, multiple integral, line and surface integrals.
- applications to mechanics (Newton's equation, potential and kinematical energy), electrodynamics (Maxwell's equations), statistical analysis (the method of least squares).

In the coming weeks, we learn **sequence of functions**. a goal is **Taylor expansion**: some *nice* functions can be written as $f(x) = \sum_{n=0}^\infty \frac{f^{(n)}(a)}{n!}(x-a)^n$. For example, $e^x = \sum_{n=0}^\infty \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \cdots, \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$.

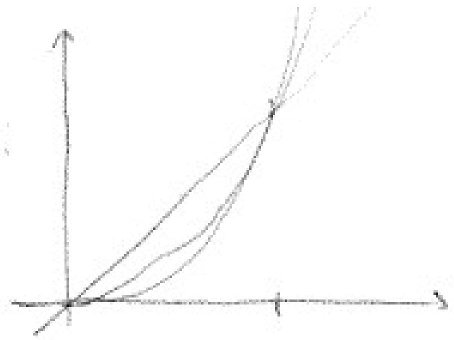
Sequence of functions and convergence

In Mathematical Analysis I we learned sequence of numbers $a_1, a_2, \cdots$, or $\{a_n\}_{n\in\mathbb{N}}$. For example,

- $a_1 = 1, a_2 = 2, a_3 = 3, \cdots$, or $a_n = n$.
- $a_1 = 1, a_2 = 4, a_3 = 9, \cdots$, or $a_n = n^2$.
- $a_1 = 0, a_2 = 1, a_3 = 0, \cdots$, or $a_n = \frac{1}{2}(1 + (-1)^n)$.

Here we consider **sequence of functions** $f_1(x), f_2(x), \cdots$ or $\{f_n(x)\}_{n\in\mathbb{N}}$ for $x \in S \subset \mathbb{R}$. For example,

- $f_1(x) = x, f_2(x) = x^2, f_3(x) = x^3, \cdots$, or $f_n(x) = x^n$.
- $f_1(x) = e^x, f_2(x) = e^{2x}, f_3(x) = e^{3x}, \cdots$, or $f_n(x) = e^{nx}$.
- $f_1(x) = \sin x, f_2(x) = \sin(\sin x), f_3(x) = \sin(\sin(\sin(x))), \cdots$.



Recall that a sequence of numbers $\{a_n\}$ is said to convergent to $a \in \mathbb{R}$ and we write $a_n \to a$ if for each $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that for any $n \geq N$ it holds that $|a_n - a| < \varepsilon$.

Example 1.

- $a_1 = 1, a_2 = \frac{1}{2}, a_3 = \frac{1}{3}, \cdots$, is convergent to 0.
- $a_1 = 0, a_2 = 1, a_3 = 0, \cdots$, or $a_n = \frac{1}{2}(1 + (-1)^n)$, is not convergent.
- $a_1 = \frac{1}{2}, a_2 = \frac{2}{2^2} = \frac{1}{2}, a_3 = \frac{3}{2^3}, \cdots$, or $a_n = \frac{n}{2^n}$, is convergent to 0.

For a sequence of function, there are various concept of convergence. Let us take an example: $f_n(x) = x^n, x \in [0, \infty)$.

- For each $x \in [0, 1)$, $f_n(x) \to 0$.
- For $x = 1$, $f_n(x) = 1$, hence is convergent to 1.
- For each $x \in (1, \infty)$, $f_n(x) \to \infty$, hence is divergent.

Definition 2. Let $S \subset \mathbb{R}$ and $f_n(x)$ be a sequence of functions on S, $f(x)$ a function on S. If $f_n(x) \to f(x)$ for each $x \in S$, then we say that $\{f_n\}$ is **pointwise convergent** to f.

We say that $\{f_n\}$ is **uniformly convergent to** f if $\{f_n\}$ is pointwise convergent to f on S and for each $\varepsilon > 0$ there is N such that for each $n \geq N$ it holds that $|f_n(x) - f(x)| < \varepsilon$ for all $x \in S$.

In the example above, $\{f_n(x)\}$ is uniformly convergent on $[0, s]$ for any $0 < s < 1$, but not uniformly convergent on $[0, 1]$ (exercise).

Consider also $f_n(x) = e^{-nx^2}, x \in \mathbb{R}$. Where is it uniformly convergent and what is the limit?

Sequence of continuous functions

Let $f(x)$ be a function on $S \subset \mathbb{R}$. Recall that f is continuous at $p \in S$ if for each $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(p)| < \varepsilon$ for $x \in S, |x - p| < \delta$. f is said to be continuous on S if it is continuous at each $p \in S$.

Theorem 3. *Assume that $f_n \to f$ uniformly on S and f_n are continuous on S. Then f is continuous on S.*

Proof. Let $p \in S$. For each $\varepsilon > 0$, by uniform convergence, there is N such that for $n \geq N$ it holds that $|f_n(x) - f(x)| < \frac{\varepsilon}{3}$ for $x \in S$. By continuity of $f_N(x)$ at $x = p$, there is $\delta > 0$ such that $|f_N(x) - f_N(p)| < \frac{\varepsilon}{3}$. Therefore, for $|x - p| < \delta$, we have

$$\begin{aligned} |f(x) - f(p)| &= |f(x) - f_N(x) + f_N(x) - f_N(p) + f_n(p) - f(p)| \\ &< |f(x) - f_N(x)| + |f_N(x) - f_N(p)| + |f_n(p) - f(p)| < 3 \cdot \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

This is continuity of f at p. As $p \in S$ is arbtrary, this shows continuity of f on S. $\square$

Recall that, if f is continuous on a closed interval $[a, b]$, then we learned in Analysis I that it is **uniformly continuous**: for each ϵ there is δ that $|f(x) - f(y)| < \epsilon$ whenever $x, y \in [a, b], |x - y| < \delta$. Furthermore, a continuous function on $[a, b]$ has the **absolute miminum and maximum**.

A step function s is a function such that $s(x) = a_k$ for $x \in [x_k, x_{k+1})$, where $a = x_1 < x_2 < \cdots < x_n = n$. For a step function s, its integral is defined by $\int_a^b s(x)dx = \sum_{k=1}^{n-1} a_k(x_{k+1} - x_k)$. A function f on $[a, b]$ is said to be **integrable** if

$$\sup_s \int_a^b s(x)dx = \inf_S \int_a^b S(x)dx,$$

where the sup and inf are taken among step functions $s(x) \leq f(x) \leq S(x)$ on S. In this case, the integral $\int_a^b f(x)dx$ is defined to be the value of this equation above.

Recall that

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx \leq (b - a) \max_{x \in [a,b]} \{|f(x)|\}.$$

Theorem 4. *Let $\{f_n\}$ be a sequence of continuous functions, uniformly convergent to f. Then it holds that*

$$\lim_{n \to \infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Proof. For $\varepsilon > 0$, by uniform convergence there is N such that for $n \geq N$ it holds that $|f_n(x) - f(x)| < \frac{\varepsilon}{b-a}$. Then we obtain

$$\left| \int_a^b f_n(x)dx - \int_a^b f(x)dx \right| \leq \int_a^b |f_n(x) - f(x)|dx \leq (b - a)\frac{\varepsilon}{b - a} = \varepsilon.$$

This shows that $\lim_{n \to \infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx$. $\square$

Oct 25. Power series, Taylor series.

Series of functions, the Weierstrass M-test

Recall that, for a sequence $\{a_n\}$ of numbers, the series $\sum a_n$ is the sequence $\{\sum_{k=1}^n a_k\}$ of numbers, consisting of partial sums. We say that a series $\sum a_n$ is convergent if $\{\sum_{k=1}^n a_k\}$ is convergent.

In the same way, for a sequence of functions $\{f_n\}$, we consider series of function $\sum_n f_n$. This series is said to be pointwise convergent if $\{\sum_{k=1}^n f_k(x)\}$ is pointwise convergent, uniformly convergent if $\{\sum_{k=1}^n f_k(x)\}$ is uniformly convergent.

Just by replacing a sequence by a series, we obtain the following.

Theorem 5. *Assume that series $\sum f_n$ is convergent uniformly to g on S and f_n are continuous on S. Then g is continuous on S.*

Let $\{f_n\}$ be a sequence of continuous functions and $\sum_n f_n$ uniformly convergent to g. Then it holds that

$$\lim_{n\to\infty} \int_a^b \sum_{k=1}^n f_k(x)dx = \int_a^b g(x)dx.$$

Proof. The same proofs apply, by noting that if f_n's are continuous, then $\sum_{k=1}^n f_k$ is continuous. □

Recall some test for convergence of series of numbers.

- (Ratio test) Let $a_n > 0$ and $\frac{a_{n+1}}{a_n} \to L$. If $L < 1$, $\sum_{n=0}^\infty a_n$ converges. If $L > 1$, $\sum_{n=0}^\infty a_n$ diverges.
- (Root test) Let $a_n > 0$ and $(a_n)^{\frac{1}{n}} \to R$. If $R < 1$, $\sum_{n=0}^\infty a_n$ converges. If $R > 1$, $\sum_{n=0}^\infty a_n$ diverges.
- (Comparison test) Let $a_n, b_n > 0, c > 0$ such that $a_n < cb_n$. If $\sum_{n=0}^\infty b_n$ converges, so does $\sum_{n=0}^\infty a_n$.

There is a useful criterion for uniform convergence.

Theorem 6 (The Weierstrass's M-test)**.** *Let f_n be a sequence of functions on $S \subset \mathbb{R}$. If there is a convergent series $\{M_n\}$ of positive numbers such that $|f_n(x)| \le M_n$, then $\sum f_n$ is uniformly convergent.*

Proof. By comparison test, $\sum |f_n(x)|$ is convergent for all $x \in S$, or in other words, $\sum f_n(x)$ is pointwise absolutely convergent. Let $f(x)$ be the limit.

To see uniform convergence, we compute

$$\left| f(x) - \sum_{k=1}^n f_k(x) \right| = \left| \sum_{k=n+1}^\infty f_k(x) \right| \le \sum_{k+1}^\infty |f_k(x)| \le \sum_{k+1}^\infty M_k.$$

As $\sum_n M_n$ is convergent, this last expression tends to 0 as $k \to \infty$, independently of x. This shows uniform convergence. □

Power series

Let $a_n \in \mathbb{C}$ be a sequence of complex numbers. We can consider the series (called a **power series**)

$$\sum_n a_n z^n.$$

This may converge for some z, and diverge for other z.

Example 7. Simplest examples of power series.

- With $a_n = \frac{1}{3^n}$, $\sum_n \frac{z^n}{3^n}$ is convergent for $|z| < 3$, and divervent for $|z| > 3$ (see the theorem below). Indeed, by root test, $\left(\frac{|z|^n}{3^n}\right)^{\frac{1}{n}} = \frac{|z|}{3}$, and hence the series is convergent if $\frac{|z|}{3} < 1$ and divergent, say for positive z, if $\frac{|z|}{3} > 1$.
- With $a_n = \frac{1}{n!}$, $\sum_n \frac{z^n}{n!}$ is convergent for all z. Indeed, by ration test, $\left(\frac{z^{n+1}}{(n+1)!}\right) / \left(\frac{z^n}{n!}\right) = \frac{z}{n+1} \to 0$ for all z, therefore, the series is absolutely convergent for all z.

Theorem 8. *Assume that $\sum a_n z^n$ converges for some $z = z_0 \neq 0$. Then for $R < |z_0|$, the series converges uniformly for $z, |z| \leq R$ and absolutely convergent.*

Proof. If $\sum a_n z_0^n$ is convergent, then in particular $|a_n z_0^n|$ is bounded, namely, less than M for some $M > 0$. Then, if $|z| < R < |z_0|$, then $|a_n z^n| = |a_n z_0^n| \cdot \left|\frac{z}{z_0}\right|^n < M\frac{R^n}{|z_0|^n}$, where $\frac{R^n}{|z_0|^n} < 1$. As $\sum M\frac{R^n}{|z_0|^n}$ is convergent (it is a geometric series), by the M-test, the series is uniformly and absolutely convergent. □

Theorem 9. *Assume that $\sum a_n z^n$ converges for some $z = z_1 \neq 0$ and not convergent for $z = z_2$. Then there is $r > 0$ such that $\sum a_n z^n$ is convergent for $|z| < r$ and divergent for $|z| > r$.*

Proof. As there is z_1, by Theorem 8, the series $\sum a_n z^n$ is convergent for $|z| < |z_1|$. Let A be the set of positive numbers R for which $\sum a_n z^n$ is convergent if $|z| < R$. As there is z_2, A is a bounded set. Let r be the least upper bound. By definition, if $|z| < r$, then $\sum a_n z^n$ is convergent. On the other hand, if $|z_3| > r$ and $\sum a_n z_3^n$ is convergent, then by Theorem 8, the series must converge for z with $r < |z| < |z_3|$. Namely, $|z| \in A$. This contradicts with the definition of A, therefore, $\sum a_n z^n$ is divergent for $|z| > r$. □

This r is called the radius of convergence for the series $\sum a_n z^n$. If the power series converges for all $z \in \mathbb{C}$, the radius of convergence is ∞ by convention. If it does not converge except $z = 0$, the radius of convergence is 0.

Derivative and integration of power series

Now let $a_n \in \mathbb{R}, x \in \mathbb{R}$. If $\sum a_n x^n$ converges, we can define a function by $f(x) = \sum_{n=0}^{\infty} a_n x^n$.

We learned that it is not always possible to exchange limits and derivative or integration. For power series, the situation is better.

Theorem 10. *Assume that, for all $x \in (-r, r)$, the series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is convergent. Then $f(x)$ is continous and $\int_0^x f(t)dt = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$.*

Proof. Let us take R such that $|x| < R < r$. By Theorem 8, the series is uniformly convergent for $t \in [-R, R]$. Then by Theorem 5, $f(x)$ is continuous in $[-R, R]$ and as $x \in [-R, R]$, we can exchange the limit and integral, namely,

$$\int_0^x f(t)dt = \sum_{n=0}^{\infty} \int_0^x a_n t^n dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}.$$

□

Theorem 11. *Assume that, for all $x \in (-r, r)$, the series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is convergent. Then $f(x)$ is differentiable and and $f'(x) = \sum_{n=1}^{\infty} n a_n x^n$.*

Proof. In this case, r is smaller or equal to the radius of convergence. As $|x| < r$, we can take r_0 such that $|x| < r_0 < r$ and then $\sum_{n=1}^{\infty} na_n x^{n-1} = \sum_{n=1}^{\infty} na_n r_0^n \cdot \frac{x^{n-1}}{r_0^{n-1}}$. The series $\sum_{n=1}^{\infty} a_n r_0^n$ is absolutely convergent and $\frac{n|x|^{n-1}}{r_0^{n-1}}$ is bounded, hence by comparison test, $\sum_{n=1}^{\infty} na_n x^{n-1}$ is (absolutely) convergent.

This function $g(x) = \sum_{n=1}^{\infty} na_n x^{n-1}$ is a power series with the coefficients na_n. By Theorem 10, $\int_0^x g(x) = \sum_{n=1}^{\infty} a_n x^n = f(x) - a_0$. This shows that $f(x)$ is differentiable by the fundamental theorem of calculus and $f'(x) = g(x) = \sum_{n=1}^{\infty} na_n x^{n-1}$. □

Example 12.

- As this is a geometric series, we know $\frac{1}{x+1} = \sum_{n=0}^{\infty}(-1)^n x^n$ for $|x| < 1$. On the other hand, $(\log(x+1))' = \frac{1}{x+1}$. Hence $\log(x+1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1}$.
- We know $\frac{1}{x^2+1} = \sum_{n=0}^{\infty}(-1)^n x^{2n}$ for $|x| < 1$. On the other hand, $(\arctan x)' = \frac{1}{x^2+1}$. Hence $\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}$.

Set 30. Power series, Taylor series.

Shifted power series

Let $\{a_n\} \subset \mathbb{C}$. Instead of $\sum a_n z^n$, we can consider, for $a \in \mathbb{C}$, a shifted power series $\sum a_n (z-a)^n$. The theorem about the radius of convergence holds in a parallel way. If $a_n, a \in \mathbb{R}$, then $f(x) = \sum_{n=0}^{n} a_n (x-a)^n$ defines a function on $(a-r, a+r)$, and the integral and differentiation can be done term by term.

In particular,

Theorem 13. *Let $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$ with $x \in (a-r, a+r)$, where r is the radius of convergence. Then $f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)a_n(x-a)^{n-k}$.*

Corollary 14. *If $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n = \sum_{n=0}^{\infty} b_n (x-a)^n$, then $a_k = b_k = k! f^{(k)}(a)$ for all n.*

Proof. The n-th derivatives $f^{(n)}(a)$ are determined by the function $f(x)$. □

Taylor's series

If $f(x)$ is defined by $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$, then we saw $a_n = \frac{f^{(n)}(a)}{n!}$.
Question: If $f(x)$ is infinitely many times differentiable, we can develop a power series (**Taylor's series for** f) $\sum_{n=0}^{n} \frac{f^{(n)}}{n!}(x-a)^n$. Does it converge to $f(x)$?

Answer: not always. Consider $f(x) = \begin{cases} e^{-\frac{1}{x}} = 0 & (x > 0) \\ 0 & (x \le 0) \end{cases}$ (exercise).

Let $E_n(x) := f(x) - \sum_{k=0}^{n} \frac{f^{(k)}}{k!}(x-a)^k$ be the error term of the n-th approximation. We learned that $E_n(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)dt$. One can prove this by integration by parts: for example, with $n = 2$,

$$\begin{aligned}
\frac{1}{2}\int_a^x (x-t)^2 f^{(3)}(t)dt &= \frac{1}{2}\left[(x-t)^2 f^{(2)}(t)\right]_a^x + \int_a^x (x-t) f^{(2)}(t)dt \\
&= -\frac{1}{2}(x-a)^2 f^{(2)}(a) + \left[(x-t)f'(t)\right]_a^x + \int_a^x f'(t)dt \\
&= -\frac{1}{2}(x-a)^2 f^{(2)}(a) - (x-a)f'(t) + f(x) - f(a) = E_2(x).
\end{aligned}$$

There is a useful criterion for the convergence of Taylor's series.

Theorem 15. *If there is $A, r \geq 0$ such that $|f^{(n)}(t)| \leq A^n$ for $t \in (a-r, a+r)$, then $E_n(x) \to 0$ as $n \to \infty$ for $x \in (a-r, a+r)$.*

Proof. For $x \geq a$, $E_n(x)$ can be estimated as

$$\begin{aligned}
|E_n(x)| &\leq \frac{1}{n!}\int_a^x |x-t|^n |f^{(n+1)}(t)| dt \\
&\leq \frac{1}{n!}\int_a^x (x-t)^n A^n dt \\
&= \frac{1}{(n+1)!}\left[-(x-t)^{n+1}\right]_a^x = \frac{1}{(n+1)!}A^n (x-a)^{n+1}
\end{aligned}$$

This tends to 0 as $n \to \infty$. A similar estimate can be made for $x \leq a$. □

Example 16.

- $f(x) = \sin x$. $f^{(1)}(x) = \cos x, f^{(2)}(x) = -\sin x, f^{(3)}(x) = -\cos x, \cdots$, $|f^{(n)}(x)| \leq 1$. Theorem 15 applies with $a = 0$ and $\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots$.
- Similarly, $\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots$.
- $f(x) = e^x$ on $x \in [-T, T]$. $f^{(n)}(x) = e^x$, hence $|f^{(n)}(x)| \leq e^T$ and Theorem 15 applies with $x = 0$. $e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$.

Applications to ordinary differential equations

An ordinary differential equation is an equation about a function $y(x)$ instead of a number x. For example, $-2y(x) = (1-x^2)y''(x)$. Such equations can be sometimes solved using power series.
Problem: Find a function $y(x)$ such that $-2y(x) = (1-x^2)y''(x)$ with $y(0) = 1, y'(0) = 1$.
Solution:

Step 1. Assume that $y(x) = \sum_{n=0}^{\infty} a_n x^n$. Then

$$y'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}, \quad y''(x) = \sum_{n=2}^{\infty} n(n-1)a_n x^{n-2}.$$

Step 2. $y(x)$ must satisfy

$$\begin{aligned}
-2y(x) = -2\sum_{n=0}^{\infty} a_n x^n &= (1-x^2)y''(x) \\
&= (1-x^2)\sum_{n=2}^{\infty} n(n-1)a_n x^{n-2} \\
&= \sum_{n=2}^{\infty} n(n-1)a_n x^{n-2} - \sum_{n=2}^{\infty} n(n-1)a_n x^n \\
&= \sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n - \sum_{n=0}^{\infty} n(n-1)a_n x^n
\end{aligned}$$

Step 3. By Corollary 14, $-2a_n = (n+2)(n+1)a_{n+2} - n(n-1)a_n$. Equivalently, $(n+2)(n+1)a_{n+1} = [n(n-1)-2]a_n = (n+1)(n-2)a_n$, or $a_{n+2} = \frac{n-2}{n+2}a_n$.

Step 4. $-a_0 = a_2$, $a_4 = 0 = a_6 \cdots$. $a_3 = -\frac{1}{3}a_1, a_5 = \frac{1}{5}a_3 = -\frac{1}{5\cdot 3}a_1, a_7 = \frac{3}{7}a_5 = -\frac{1}{7\cdot 5}a_1$, in general, $a_{2n+1} = -\frac{a_1}{(2n+1)(2n-1)}$.

Step 5. By $y(0) = 1, a_0 = 1$ and $y'(0) = 1, a_1 = 1$. Hence $y(x) = 1 - x^2 + \sum_{n=0}^{\infty} \frac{1}{(2n+1)(2x-1)}x^{2n+1}$.

Step 6. This is convergent for $|x| < 1$.

Binominal series

For $\alpha \in \mathbb{R}, n \in \mathbb{N}$, we define

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}.$$

Theorem 17. *$(1+x)^\alpha = \sum_{n=0}^{n} \binom{\alpha}{n} x^n$ for $|x| < 1$.*

Proof. By ratio test, $\left|\binom{\alpha}{n+1}|x|^{n+1}\right| / \left|\binom{\alpha}{n}|x|^n\right| = |\alpha - n + 1||x|/n \to |x|$, the right-hand side converges for $|x| < 1$.

Put $f(x) = (1+x)^\alpha$, then $f'(x) = \alpha(1+x)^{\alpha-1} = \alpha\frac{f(x)}{1+x}$ and $f(0) = 1$. Put $g(x) = \sum_{n=0}^\infty \binom{\alpha}{n} x^n$. $(x+1)g'(x) = \sum_{n=1}^\infty n\binom{\alpha}{n}x^{n-1}(x+1) = \sum_{n=1}^\infty [n\binom{\alpha}{n} + (n+1)\binom{\alpha}{n+1}]x^n = \alpha g(x)$, and $g(0) = 1$.

Therefore, $f(x)$ and $g(x)$ satisfy the same first-order differential equation and $f(0) = g(0) = 1$, hence $f(x) = g(x)$. Namely, $(1+x)^\alpha = \sum_{n=0}^\infty \binom{\alpha}{n} x^n$ for $|x| < 1$. $\square$

Oct 02. Scalar and vector fields.

Higher dimensional space

Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$. We define the inner product $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{k=1}^n x_k y_k \in \mathbb{R}$ and the norm $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \sqrt{\sum_{k=1}^n x_k^2}$. In linear algebra we leaned

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| \text{ (Cauchy-Schwarz inequality)}$$
$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\| \text{ (Triangle inequality)}$$

A map $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^m$ is called a "field". The case $m = 1$ is a **scalar field**, and in general it is a **vector field**.

Some examples have practical applications:

- $T : \mathbb{R}^3 \supset S \to \mathbb{R}$, temperature in a room
- $\boldsymbol{V} : \mathbb{R}^3 \supset S \to \mathbb{R}^3$, wind velocity
- $\boldsymbol{E} : \mathbb{R}^3 \supset S \to \mathbb{R}^3$, electric field.

We denote $\boldsymbol{f}(x_1, \cdots, x_n)$ by $\boldsymbol{f}(\boldsymbol{x})$, and they represent the same vector field $\mathbb{R}^n \to \mathbb{R}^m$.

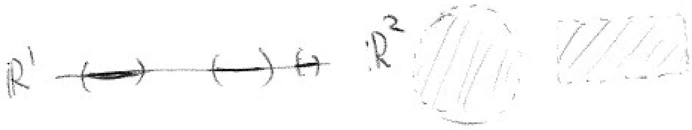
Open balls and open sets

Let $\boldsymbol{a} \in \mathbb{R}^n, r > 0$. The open n-ball with radius r with center $\boldsymbol{a}$ is $B(a; r) := \{\boldsymbol{x} : \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{a}\| < r\}$.

Definition 18. Let $S \subset \mathbb{R}^n, \boldsymbol{a} \in S$. $\boldsymbol{a}$ is called an interior point if there is $r > 0$ such that $B(\boldsymbol{a}; r) \subset S$. We denote $\operatorname{int} S := \{\boldsymbol{x} \in S : \boldsymbol{x} \text{ is an interior point.}\}$. S is said to be open if $\operatorname{int} S = S$.

Example 19.

open intervals (not containing the end points)in $\mathbb{R}$
open disks and open rectangles (not containing the edges) in $\mathbb{R}^2$
open balls and open cuboids in $\mathbb{R}^3$



Definition 20. Let $S \subset \mathbb{R}^n, \boldsymbol{a} \notin S$. $\boldsymbol{a}$ is called an exterior point if there is $r > 0$ such that $B(\boldsymbol{a}; r) \cap S = \emptyset$. We denote $\operatorname{Ext} S := \{\boldsymbol{x} \notin S : \boldsymbol{x} \text{ is an exterior point.}\}$. Note that $\operatorname{Ext} S$ is an open set. $\partial S := \mathbb{R}^n \setminus (\operatorname{int} S \cup \operatorname{ext} S)$ is called the bounary of S.

Let $K \subset \mathbb{R}^n$. K is said to be a closed set if $\partial K \subset K$.

Proposition 21. *$\mathbb{R}^n \subset S$ is open if and only if S^{c} is closed.*

Proof. Note that $\mathbb{R}^n = \operatorname{int} S \cup \operatorname{ext} S \cup \partial S$, and this is a disjoint union.

If $\boldsymbol{x} \in \partial S$, then for any $\epsilon > 0$, $B(\boldsymbol{x}; \varepsilon) \cap S \neq \emptyset$, hence $\boldsymbol{x} \in \partial(S^{\mathrm{c}})$. By the same argument, $\partial S = \partial(S^{\mathrm{c}})$.

If S is open, then $\operatorname{int} S = S$, and $S^{\mathrm{c}} = \operatorname{ext} S \cup \partial S$. Hence $\partial(S^{\mathrm{c}}) = \partial S =\subset S^{\mathrm{c}}$ and S^{c} is closed.

If S is not open, then there is $\boldsymbol{x} \in \partial S \cap S$. This means $\boldsymbol{x} \in \partial S^{\mathrm{c}} \cap S$ but $\boldsymbol{x} \notin S^{\mathrm{c}}$ hence S^{c} is not closed. $\square$

Limits

Let $S \subset \mathbb{R}^n$, $\boldsymbol{f} : S \to \mathbb{R}^m$ a vector field, $\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m$ If $\lim_{\|\boldsymbol{x}-\boldsymbol{a}\| \to 0} \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{b}\| = 0$, then we write $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}$. $\boldsymbol{f}$ is said to be continuous at $\boldsymbol{a}$ if $\boldsymbol{f}(\boldsymbol{a}) = \lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x})$.

Theorem 22. *Let $S \subset \mathbb{R}^m$ and $\boldsymbol{f}, \boldsymbol{g} : S \to \mathbb{R}^m$ two vector fields such that $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}, \lim_{\boldsymbol{x} \to \boldsymbol{a}} g(\boldsymbol{x}) = \boldsymbol{c}$.*

(a) $\lim_{\boldsymbol{x} \to \boldsymbol{a}} (\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})) = \boldsymbol{b} + \boldsymbol{c}$.

(b) For $\lambda \in \mathbb{R}$, $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \lambda \boldsymbol{f}(\boldsymbol{x}) = \lambda \boldsymbol{b}$.

(c) $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{b} \cdot \boldsymbol{c}$.

(d) $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \|\boldsymbol{f}(\boldsymbol{x})\| = \|\boldsymbol{b}\|$.

Proof. We do only (c) and (d).

$$\begin{aligned} |\boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{b} \cdot \boldsymbol{c}| &= |(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{b}) \cdot (\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{c}) + (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{b}) \cdot \boldsymbol{c} + \boldsymbol{b} \cdot (\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{c})| \\ &\leq \|(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{b})\| \cdot \|(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{c})\| + \|(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{b})\| \cdot \|\boldsymbol{c}\| + \|\boldsymbol{b}\| \cdot \|(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{c})\| \to 0. \end{aligned}$$

We have $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \|\boldsymbol{f}(\boldsymbol{x})\|^2 = \boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{f}(\boldsymbol{x}) \to \|\boldsymbol{b}\|^2$ by (c), and (d) is valid because the square root is continuous. $\square$

If we write $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots f_m(\boldsymbol{x}))$, then $\boldsymbol{f}$ is continuous if and only if f_k are continuous. Indeed, if $\boldsymbol{f}$ is continuous, then $f_k(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) \cdot e_k$, where $e_k = (0, \cdots, \underset{k\text{-th}}{1}, \cdots, 0)$. Conversely, if each f_k is continuous, then $\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{a})\|^2 = \sum_{k=1}^m (\boldsymbol{f}_k(\boldsymbol{x}) - \boldsymbol{f}_k(a))^2 \to 0$.

Theorem 23. *Let $\boldsymbol{f}, \boldsymbol{g}$ be vector fields such that $\boldsymbol{g} : \mathbb{R}^\ell \supset S \to \mathbb{R}^m, \boldsymbol{f} : \mathbb{R}^m \supset T \to \mathbb{R}^n$ and $\boldsymbol{g}(S) \subset T$, so that $\boldsymbol{f} \circ \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))$ makes sense. If $\boldsymbol{g}$ is continuous at $\boldsymbol{a} \in S$ and $\boldsymbol{f}$ is continuous at $\boldsymbol{g}(\boldsymbol{a})$, then $\boldsymbol{f} \circ \boldsymbol{g}$ is continuous at $\boldsymbol{a}$.*

Proof. We just have to check $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \|\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) - \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{a}))\| = \lim_{\boldsymbol{y} \to \boldsymbol{g}(\boldsymbol{a})} \|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{a}))\| = 0$. $\square$

Example 24.
- $P(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^3$
- $\boldsymbol{f}(x_1, x_2) = (\sin(x_1x_2^2), e^{x_1 \cos x_2})$.

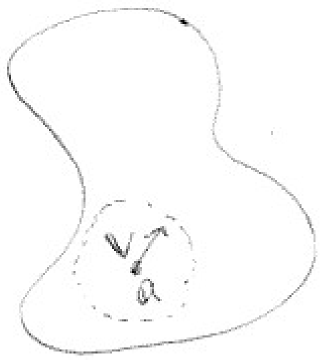
Oct 07. Derivatives of scalar fields.

Directional derivatives

Let $S \subset \mathbb{R}^n$ be an open set, $f : S \to \mathbb{R}$ a scalar field and $\boldsymbol{a} \in B(\boldsymbol{a}; r) \subset S$. In S, there are many directions in which one can approach to the point $\boldsymbol{a}$, hence we need to specify one of them when we take the derivative of f.

Let $\boldsymbol{y} \in \mathbb{R}^n$. We define the **directional derivative** in $\boldsymbol{y}$ of f to be

$$f'(\boldsymbol{a}; \boldsymbol{y}) := \frac{f(\boldsymbol{a} + h\boldsymbol{y}) - f(\boldsymbol{a})}{h}.$$



Note that $\boldsymbol{a} + h\boldsymbol{y} \in S$ for small enough h, hence this limit makes sense. To study it, let us fix $\boldsymbol{y} \in \mathbb{R}^n$ and define $g(t) = f(\boldsymbol{a} + t\boldsymbol{y})$.

Proposition 25. *$g'(0)$ exist if and only if $f'(\boldsymbol{a}; \boldsymbol{y})$ exists and $g'(0) = f'(\boldsymbol{a}; \boldsymbol{y})$.*

Proof. By definition, $\frac{g(t+h)-g(h)}{h} = \frac{f(\boldsymbol{a}+h\boldsymbol{y})-f(\boldsymbol{a})}{h}$. □

Example 26.
- Let $f(\boldsymbol{x}) = f(x_1, x_2) = \sin(x_1 + 2x_2)$ and fix $\boldsymbol{a} = (0,0), \boldsymbol{y}_1 = (1,1)$. Then $g(t) = f(\boldsymbol{a}+h\boldsymbol{y}_1) = f(t\boldsymbol{y}_1) = f(t,t) = \sin 3t$ and $g'(t) = 3\cos 3t$, hence $f'(\boldsymbol{a}; \boldsymbol{y}) = g'(0) = 3$.
- Let $f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ and fix $\boldsymbol{a}, \boldsymbol{y} \in \mathbb{R}^n$. In this example,

$$g(t) = f(\boldsymbol{a} + t\boldsymbol{y}) = \|\boldsymbol{a} + t\boldsymbol{y}\|^2 = \|\boldsymbol{a}\|^2 + 2t\boldsymbol{a} \cdot \boldsymbol{y} + t^2\|\boldsymbol{y}\|^2$$

  and hence $g'(t) = 2\boldsymbol{a} \cdot \boldsymbol{y} + 2t\|\boldsymbol{y}\|^2$, $f(\boldsymbol{a}; \boldsymbol{y}) = g'(0) = 2\boldsymbol{a} \cdot \boldsymbol{y}$.

Proposition 27. *Assume that $f'(\boldsymbol{a} + t\boldsymbol{y}; \boldsymbol{y})$ exists for $0 \le t \le 1$. Then there is $0 \le \theta \le 1$ such that $f(\boldsymbol{a} + \boldsymbol{y}) - f(\boldsymbol{a}) = f'(\boldsymbol{a} + \theta\boldsymbol{y}; \boldsymbol{y})$.*

Proof. Apply the mean value theorem fo $g(t) = f(\boldsymbol{a} + t\boldsymbol{y})$ and obtain that there is $0 \le \theta \le 1$ such that $g(1) - g(0) = g'(\theta)$, namely, $f(\boldsymbol{a} + \boldsymbol{y}) - f(\boldsymbol{a}) = f'(\boldsymbol{a} + \theta\boldsymbol{y}; \boldsymbol{y})$. □

Partial derivatives

For $k = 1, \cdots, n$, let $\boldsymbol{e}_k = (0, \cdots 0, \underset{k\text{-th}}{1}, 0, \cdots 0)$. We define the partial derivative in x_k of $f(x_1, \cdots, x_n)$ at $\boldsymbol{a}$ by

$$\frac{\partial f}{\partial x_k}(\boldsymbol{a}) = f'(\boldsymbol{a}; \boldsymbol{e}_k).$$

There are various notations of **partial derivatives**:

- $\frac{\partial f}{\partial x_k}(\boldsymbol{a}) = D_k f(\boldsymbol{a})$.

- If we consider $\mathbb{R}^2$, then the scalar field is often written as $f(x,y)$ and one denotes $\frac{\partial f}{\partial x}(\boldsymbol{a}) = D_1 f(\boldsymbol{a})$, $\frac{\partial f}{\partial y}(\boldsymbol{a}) = D_2 f(\boldsymbol{a})$. Similarly, if we are in $\mathbb{R}^3$, then for $f(x,y,z)$ we also denote $\frac{\partial f}{\partial z}(\boldsymbol{a}) = D_3 f(\boldsymbol{a})$.

If $D_k f$ exists, one can also consider $D_\ell(D_k f) = \frac{\partial^2 f}{\partial x_\ell \partial x_k}$, and even higher partial derivatives. In general $D_\ell D_k f \neq D_k D_\ell f$.

In practice, in order to compute the partial derivative $\frac{\partial f}{\partial x_k}$, one should consider all other $x_\ell, \ell \neq k$ as constants and take the derivative with respect to x_k.

Example 28.
- (Good function) Let us take $f(x,y) = x^2 + 3xy + y^4$. Then $\frac{\partial f}{\partial x}(x,y) = D_1 f(x,y) = 2x + 3y$, $\frac{\partial f}{\partial y}(x,y) = D_2 f(x,y) = 3x + 4y^3$. Further, $\frac{\partial^2 f}{\partial x^2}(x,y) = 2$, $\frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{\partial^2 f}{\partial x \partial y}(x,y) = 3$, $\frac{\partial^2 f}{\partial y^2}(x,y) = 12y^2$.
- (Bad function) Consider $f(x,y) = \begin{cases} \frac{xy^2}{x^2+y^4} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$. Let $\boldsymbol{b} = (b,c)$ where $b \neq 0$, then $f'((0,0);\boldsymbol{b}) = \lim_{h\to 0} \frac{bc^2h^3}{(b^2h^2+c^4h^4)h} = bc^2$. Similarly, if $\boldsymbol{b} = (0,c)$, then $f'((0,0);\boldsymbol{b}) = \lim_{h\to 0} \frac{0}{c^4h^4} = 0$. Therefore, all the directional derivatives exist. However, if we take $f(t^2,t) = \frac{t^4}{2t^4} = \frac{1}{2}$, hence $f(x,y)$ is not continuous at $(0,0)$.

Total derivatives

Recall that, in $\mathbb{R}^1$, if $f(x)$ is differentiable, then we have

$$f(a+h) = f(a) + hf'(a) + hE(a,h)$$

and $E(a,h) \to 0$ as $h \to 0$. In other words, $f(x)$ can be approximated by $f(a) + hf'(a)$ to the first order in h around a.

Definition 29. Let $S \subset \mathbb{R}^n$ be open, $f : S \to \mathbb{R}$ a scalar field. We say that f is **differentiable** at $\boldsymbol{a} \in S$ if there is $T_{\boldsymbol{a}} \in \mathbb{R}^n$ and $E(\boldsymbol{a},\boldsymbol{v})$ such that

$$f(\boldsymbol{a}+\boldsymbol{v}) = f(\boldsymbol{a}) + T_{\boldsymbol{a}} \cdot \boldsymbol{v} + \|\boldsymbol{v}\| E(\boldsymbol{a},\boldsymbol{v})$$

for $\boldsymbol{v} \in B(\boldsymbol{a};r)$ and $E(\boldsymbol{a},\boldsymbol{v}) \to 0$ as $\boldsymbol{v} \to \boldsymbol{0}$. $T_{\boldsymbol{a}}$ is called the **total derivative** of f at $\boldsymbol{a}$.

Theorem 30. *If f is differentiable at $\boldsymbol{a}$, then $T_{\boldsymbol{a}} = (D_1 f(\boldsymbol{a}), \cdots D_n f(\boldsymbol{a}))$ and $f'(\boldsymbol{a};\boldsymbol{y}) = T_{\boldsymbol{a}} \cdot \boldsymbol{y}$.*

Proof. As f is differentiable at $\boldsymbol{a}$, it holds that $f(\boldsymbol{a}+\boldsymbol{v}) = f(\boldsymbol{a}) + T_{\boldsymbol{a}} \cdot \boldsymbol{v} + \|\boldsymbol{v}\| E(\boldsymbol{a},\boldsymbol{v})$ where $E(\boldsymbol{a},\boldsymbol{v})$ as $\boldsymbol{v} \to \boldsymbol{0}$. Let us take $\boldsymbol{v} = h\boldsymbol{y}$. Then

$$f'(\boldsymbol{a};\boldsymbol{y}) = \frac{f(\boldsymbol{a}+h\boldsymbol{y}) - f(\boldsymbol{a})}{h} = \frac{T_{\boldsymbol{a}} \cdot h\boldsymbol{y} + h\|\boldsymbol{y}\|E(\boldsymbol{a},h\boldsymbol{y})}{h} = T_{\boldsymbol{a}} \cdot \boldsymbol{y} + \|\boldsymbol{y}\|E(\boldsymbol{a},h\boldsymbol{y}) \to T_{\boldsymbol{a}} \cdot \boldsymbol{y}.$$

Especially, if $\boldsymbol{a} = \boldsymbol{e}_k$, then $D_k f(\boldsymbol{a}) = T_{\boldsymbol{a}} \cdot \boldsymbol{e}_k$. Therefore, we have $T_{\boldsymbol{a}} = (D_1 f(\boldsymbol{a}), \cdots, D_n f(\boldsymbol{a}))$. □

$T_{\boldsymbol{a}} = (D_1 f(\boldsymbol{a}), \cdots, D_n f(\boldsymbol{a})) =: \nabla f(\boldsymbol{a})$ is called the **gradient** of f at $\boldsymbol{a}$.

Proposition 31. *If f is differentiable at $\boldsymbol{a}$, then it is continuous at $\boldsymbol{a}$.*

Proof. We just have to estimate

$$|f(\boldsymbol{a}+\boldsymbol{v}) - f(\boldsymbol{a})| = |T_{\boldsymbol{a}} \cdot \boldsymbol{v} + \|\boldsymbol{v}\|E(\boldsymbol{a},\boldsymbol{v})| \leq \|T_{\boldsymbol{a}}\|\|\boldsymbol{v}\| + \|\boldsymbol{v}\||E(\boldsymbol{a},\boldsymbol{v})| \to 0.$$

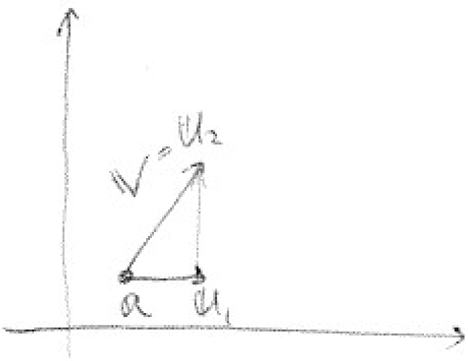
□

Theorem 32. *Assume that $D_1 f, \cdots D_n f$ exist in $B(\boldsymbol{a}; r)$ and are continuous at $\boldsymbol{a}$. Then f is differentiable at $\boldsymbol{a}$.*

Proof. Let us write $\boldsymbol{v} = (v_1, \cdots v_n)$ and introduce $\boldsymbol{u}_k = (v_1, \cdots, v_k, 0, \cdots, 0)$ with $\boldsymbol{u}_0 = (0, \cdots, 0)$. Note that $\boldsymbol{u}_k - \boldsymbol{u}_{k-1} = v_k \boldsymbol{e}_k$. Then we have

$$\begin{aligned} f(\boldsymbol{a}+\boldsymbol{v}) - f(\boldsymbol{a}) &= \sum_{k=1}^{n} (f(\boldsymbol{a}+\boldsymbol{u}_k) - f(\boldsymbol{a}+\boldsymbol{u}_{k-1})) \\ &= \sum_{k=1}^{n} v_k f'(\boldsymbol{a}+\boldsymbol{u}_{k-1}+\theta_k v_k \boldsymbol{e}_k; \boldsymbol{e}_k) \\ &= \sum_{k=1}^{n} v_k f'(\boldsymbol{a}+\boldsymbol{u}_{k-1}; \boldsymbol{e}_k) + \sum_{k=1}^{n} v_k (f'(\boldsymbol{a}+\boldsymbol{u}_{k-1}+\theta_k v_k \boldsymbol{e}_k; \boldsymbol{e}_k) - f'(\boldsymbol{a}+\boldsymbol{u}_{k-1}; \boldsymbol{e}_k)) \\ &= \sum_{k=1}^{n} v_k f'(\boldsymbol{a}+\boldsymbol{u}_{k-1}; \boldsymbol{e}_k) + \|\boldsymbol{v}\| \sum_{k=1}^{n} \frac{v_k}{\|\boldsymbol{v}\|} (D_k f(\boldsymbol{a}+\boldsymbol{u}_{k-1}+\theta_k v_k \boldsymbol{e}_k) - D_k f(\boldsymbol{a})). \end{aligned}$$

As we have $E(\boldsymbol{a}, \boldsymbol{v}) = \sum_{k=1}^{n} \frac{v_k}{\|\boldsymbol{v}\|} (D_k f(\boldsymbol{a}+\boldsymbol{u}_{k-1}+\theta_k v_k \boldsymbol{e}_k) - D_k f(\boldsymbol{a}))$, this tends to 0 as $\boldsymbol{v} \to \boldsymbol{0}$ and $\sum_{k=1}^{n} v_k f'(\boldsymbol{a}+\boldsymbol{u}_{k-1}; \boldsymbol{e}_k) \to \nabla f(\boldsymbol{a}) \cdot \boldsymbol{v}$.



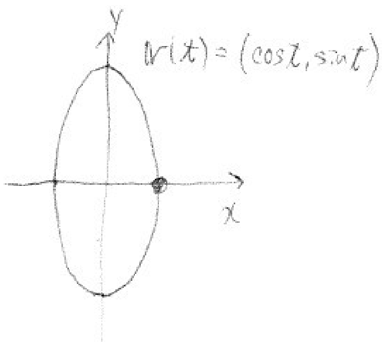
□

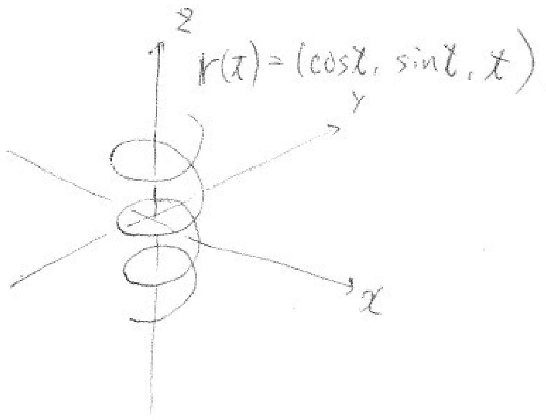
Oct 09. Tangent and chain rule.

Parametrized curves

Let $\boldsymbol{r}(t) = (X_1(t), \cdots, X_n(t))$ be a vector-valued function (defined on an interval $I \subset \mathbb{R}$ and its value is in $\mathbb{R}^n$). Such a vector-valued function $\boldsymbol{r}(t)$ describes a curve C in $\mathbb{R}^n$.

Example 33.
- Let $\boldsymbol{r}(t) = (\cos t, 2\sin t)$ for $t \in [0, 2\pi]$. This describes an ellipse in $\mathbb{R}^2$.
- Let $\boldsymbol{r}(t) = (\cos t, \sin t, t)$ for $t \in \mathbb{R}$. This describes a spiral in $\mathbb{R}^3$.





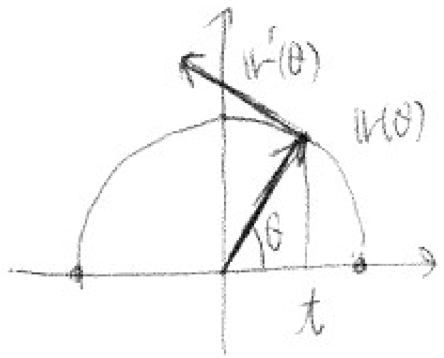
A same curve C can be described in various ways. For example, the following two vector-valued functions

$$\boldsymbol{r}_1(\theta) = (\cos\theta, \sin\theta), \theta \in [0, \pi], \qquad \boldsymbol{r}_2(t) = (-t, \sqrt{1-t^2}), t \in [-1, 1].$$

the upper half-circle $C = \{(x, y) : x^2 + y^2 = 1, y \geq 0\}$. They are both **parametrizations** of C.

When we have a such a parametrization $\boldsymbol{r}(t)$ of C and each component is differentiable, we can take the derivative $\boldsymbol{r}'(t) = (X_1'(t), \cdots, X_n'(t))$. $\boldsymbol{r}'(t)$ is called a tangent vector of C at $\boldsymbol{r}(t)$.

Example 34. For the parametrirzation $\boldsymbol{r}_1(t) = (\cos t, \sin t)$ of the unit circle, we have $\boldsymbol{r}_1'(t) = (-\sin t, \cos t)$ which is indeed tangent to the circle.



If $\boldsymbol{r}(t)$ represents $\boldsymbol{r}(t)$ is the position of the particle where t is the time, hence the motion of a particle in the space $\mathbb{R}^n$, the derivative $\boldsymbol{r}'(t)$ is called the **velocity**.

Chain rule

In $\mathbb{R}$, if $f(t) = g(h(t))$, then we have $f'(t) = h'(t)g'(h(t))$, and this is called the **chain rule**. This can be generalized to the following form: let $g(\boldsymbol{x})$ be a scalar field on $S \subset \mathbb{R}^n$ and $\boldsymbol{r}(t) = (X_1(t), \cdots, X_n(t))$ be a vector-valued function on $I \subset \mathbb{R}$ and $\boldsymbol{r}(t) \subset S$. In this situation,

$$\text{if } f(t) = g(\boldsymbol{r}(t)), \quad \text{then } f'(t) = \nabla g(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t),$$

where $\boldsymbol{r}'(t) = (X_1'(t), \cdots, X_n'(t))$. This has many applications in physics. For example, if $\boldsymbol{r}(t)$ represents the coordinates of a particle and $g(\boldsymbol{x})$ represents the potential energy, then $g(\boldsymbol{r}(t))$ is the potential energy of the particle at time t.

Theorem 35. *Let $S \subset \mathbb{R}^n$ be an open set, $\boldsymbol{r}(t)$ a vector-valued function from an open interval $I \subset \mathbb{R}$ in S and $g : S \to \mathbb{R}$ a scalar field. Define a function $f(t) : J \to \mathbb{R}$ by $f(t) = g(\boldsymbol{r}(t))$. If $\boldsymbol{r}'$ exist at $t \in I$ and g is differentiable at $\boldsymbol{r}(t)$, then f' exists at t and $f'(t) = \nabla g(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t)$.*

Proof. Let t as above. Then, as $\boldsymbol{r}'$ exists at t, for sufficiently small h we have $\boldsymbol{r}(t+h) \in B(\boldsymbol{r}(t); r) \subset S$ and $\frac{\boldsymbol{r}(t+h)-\boldsymbol{r}(t)}{h} \to \boldsymbol{r}'(t)$. In particular, $\frac{\|\boldsymbol{r}(t+h)-\boldsymbol{r}(t)\|}{h} \to \|\boldsymbol{r}'(t)\|$.

We need to estimate

$$\frac{f(t+h) - f(t)}{h} = \frac{g(\boldsymbol{r}(t+h)) - g(\boldsymbol{r}(t))}{h}.$$

As g is differentiable at $\boldsymbol{r}(t)$, we have

$$\frac{g(\boldsymbol{r}(t+h)) - g(\boldsymbol{r}(h))}{h} = \nabla g(\boldsymbol{r}(t)) \cdot \frac{\boldsymbol{r}(t+h) - \boldsymbol{r}(t)}{h} + \frac{\|\boldsymbol{r}(t+h) - \boldsymbol{r}(t)\|}{h} E(\boldsymbol{r}(t); \boldsymbol{r}(t+h) - \boldsymbol{r}(t)).$$

The last term tends to 0 as $h \to 0$ because $\boldsymbol{r}(t+h) - \boldsymbol{r}(t) \to \boldsymbol{0}$, hence $f'(t) = \nabla g(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(h)$. $\square$

Example 36. Let $\boldsymbol{x} = (x, y, z) \in \mathbb{R}$, $g(\boldsymbol{x}) = -\frac{1}{\|\boldsymbol{x}\|}, \boldsymbol{r}(t) = (t, 0, 0)$, $f(t) = g(\boldsymbol{r}(t))$. We have $\nabla g(\boldsymbol{x}) = (\frac{x}{\|\boldsymbol{x}\|^3}, \frac{y}{\|\boldsymbol{x}\|^3}, \frac{z}{\|\boldsymbol{x}\|^3})$ and $\boldsymbol{r}'(t) = (1, 0, 0)$. Then, for $t > 0$, $\nabla g(\boldsymbol{r}(t)) = (\frac{1}{t^2}, 0, 0)$ and $f'(t) = \frac{1}{t^2}$.

The chain rule can be applied to compute the derivative of some function on $\mathbb{R}$.

Example 37.

- $f(t) = t^t$ for $t > 0$. With $g(x, y) = x^y$ and $\boldsymbol{r}(t) = (t, t)$, we have $f(t) = g(\boldsymbol{r}(t))$. As $\nabla g(x, y) = (yx^{y-1}, \log x x^y)$ and $\boldsymbol{r}'(t) = (1, 1)$, we have $f'(t) = \nabla g(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) = t^t + \log t \cdot t^t$.
- $f(t) = \int_{-t^4}^{t^2} e^{s^2} ds$. We do not know the indefinite integral of e^{s^2}. However, we put it $F(x, y) = \int_x^y e^{s^2} ds$. Then, $f(t) = F(-t^4, t^2) = F(\boldsymbol{r}(t))$, where $\boldsymbol{r}(t) = (-t^4, t^2)$. With $\boldsymbol{r}'(t) = (-4t^3, 2t^2)$ and $D_1 F(x, y) = -e^{x^2}, D_2 F(x, y) = e^{y^2}$, we have $f'(t) = 2t^2 e^{t^2} + 4t^3 e^{t^2}$.

Level sets

Let f be a non-constant scalar field on $S \subset \mathbb{R}^2$. Assume that $c \in \mathbb{R}$ and the equation $f(x, y) = c$ defines a curve in $\mathbb{R}$ and it has a tangent at each of its point. Then it holds that

- The gradient vector $\nabla f(\boldsymbol{a})$ is normal to C if $\boldsymbol{a} \in C$. Indeed, assume that C can be written as $\boldsymbol{r}(t)$. Then, as f is constant along C, we have $\frac{d}{dt} f(\boldsymbol{r}(t)) = \nabla f(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) = 0$. As $\boldsymbol{r}'(t)$ is a tangent vector to C, $\nabla f(\boldsymbol{r}(t))$ is normal to C.
- The directional derivative of f is 0 along C, and it has the largest value in the direction of $\nabla f(\boldsymbol{r}(t))$.

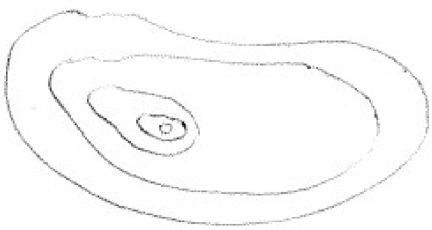
The tangent line at $\boldsymbol{a} = (a, b)$ is represented by the equation

$$D_1 f(\boldsymbol{a})(x - a) + D_2 f(\boldsymbol{a})(y - b) = 0.$$

Indeed, this passes through (a, b) and is orthogonal to $\nabla f(\boldsymbol{a}) = (D_1 f(\boldsymbol{a}), D_2 f(\boldsymbol{a}))$.

More generally, if f is a scalar field on $S \subset \mathbb{R}^n$ and $c \in \mathbb{R}$, $L(c) = \{\boldsymbol{x} \in S : f(\boldsymbol{x}) = c\}$ is called the level set of f. In $\mathbb{R}^2$ it is often a curve, and in $\mathbb{R}^3$ it is often a surface.

In $\mathbb{R}^2$, if $f(x, y)$ represents the height of the point (x, y) in a map of a region $S \subset \mathbb{R}^2$, then the set $L(c)$ is called an isopleth. $\nabla f(x, y)$ represents the direction in which the slop is the largest.



Example 38.

- Let $f(x, y, z) = x^2 + y^2 + z^2$. Then, for $c > 0$, the set $L(c) = \{(x, y, z) : x^2 + y^2 + z^2 = c\}$ is a sphere. If $c < 0$, $L(c)$ is empty and $L(0)$ is one point $(0, 0, 0)$.
- Let $f(x, y, z) = x^2 + y^2 - z^2$. Then, for $c > 0$, $L(c) = \{(x, y, z) : x^2 + y^2 - z^2 = c\}$ is a paraboloid.

Let f be a differentiable scalar field on $S \subset \mathbb{R}^3$ and assume that a level set $L(c)$ is a surface. We argue that $\nabla f(\boldsymbol{a})$ is perpendicular to the tangent of $L(c)$ at $\boldsymbol{a}$.

Let $\boldsymbol{a} \in L(c)$ and take a curve Γ that passes through $\boldsymbol{a}$ and is contained in $L(c)$. Let us take a parametrization $\alpha(t)$ of Γ and assume that $\boldsymbol{a} = \alpha(t_1)$. Then, as $\alpha(t)$ is contained in $L(c)$, $f(\alpha(t)) = c$, and the derivative vanishes: $\frac{d}{dt} f(\alpha(t)) = 0$. We know that $\frac{d}{dt} f(\alpha(t_1)) =$

$\nabla f(\boldsymbol{\alpha}(t_1)) \cdot \boldsymbol{\alpha}'(t_1)$. Therefore, the tangent vector $\boldsymbol{\alpha}'(t_1)$ is orthogonal to $\nabla f(\boldsymbol{\alpha}(t_1))$. This holds for any curve passing through $\boldsymbol{a}$ and contained in $L(c)$, hence $\nabla f(\boldsymbol{\alpha}(t_1))$ is perpendicular to the tangent plane of f.

The tangent plane at $\boldsymbol{a} = (a, b, c)$ is represented by the equation

$$D_1 f(\boldsymbol{a})(x-a) + D_2 f(\boldsymbol{a})(y-b) + D_3 f(\boldsymbol{a})(z-c) = 0.$$

Oct 14. Derivatives of vector fields.

Let $\boldsymbol{f} : \mathbb{R}^n \supset S \to \mathbb{R}^m$ be a vector field. For $\boldsymbol{a} \in S$ and $\boldsymbol{y} \in \mathbb{R}^n$, we define, as before, the directional derivative by

$$\boldsymbol{f}'(\boldsymbol{a}; \boldsymbol{y}) = \lim_{h \to 0} \frac{\boldsymbol{f}(\boldsymbol{a} + h\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{a})}{h}.$$

$\boldsymbol{f}'(\boldsymbol{a}; \boldsymbol{y})$ is a vector in $\mathbb{R}^n$. This means that, if $\boldsymbol{f} = (f_1, \cdots, f_n)$, then the directional derivative is $\boldsymbol{f}'(\boldsymbol{a}; \boldsymbol{y}) = (f_1'(\boldsymbol{a}; \boldsymbol{y}), \cdots, f_n'(\boldsymbol{a}; \boldsymbol{y}))$.

$D_k \boldsymbol{f}(\boldsymbol{a}) = \frac{\partial \boldsymbol{f}}{\partial x_k}(\boldsymbol{a}) = \boldsymbol{f}'(\boldsymbol{a}; \boldsymbol{e}_k)$ as with scalar fields. In $\mathbb{R}^3$, we often write $\boldsymbol{f} = (f_x, f_y, f_z)$.

Example 39. Let $\boldsymbol{E}(x, y, z, t), \boldsymbol{B}(x, y, z, t)$ be electric and magnetic fields. They are vector fields $\mathbb{R}^4 \to \mathbb{R}^3$.

- $\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 4\pi\rho$ (Gauss' law).
- $\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$ (absence of magnetic charge).

Similarly to the case of scalar field, we say that $\boldsymbol{f}$ is **differentiable** at $\boldsymbol{a}$ if there is a linear tranformation $T_{\boldsymbol{a}} : \mathbb{R}^n \in \mathbb{R}^m$ and $\boldsymbol{E}(\boldsymbol{a}, \boldsymbol{v})$ such that

$$\boldsymbol{f}(\boldsymbol{a} + \boldsymbol{v}) = \boldsymbol{f}(\boldsymbol{a}) + \boldsymbol{T}_{\boldsymbol{a}}(\boldsymbol{v}) + \|\boldsymbol{v}\| E(\boldsymbol{a}, \boldsymbol{v})$$

for $\boldsymbol{v} \in B(\boldsymbol{a}; r)$ and $\boldsymbol{E}(\boldsymbol{a}, \boldsymbol{v}) \to \boldsymbol{0}$ as $\boldsymbol{v} \to \boldsymbol{0}$. We also have

Theorem 40. *If $\boldsymbol{f}$ is differentiable at $\boldsymbol{a}$, then $\boldsymbol{f}$ is continuous at $\boldsymbol{a}$ and $\boldsymbol{T}_{\boldsymbol{a}}(\boldsymbol{y}) = \boldsymbol{f}'(\boldsymbol{a}; \boldsymbol{y})$.*

We omit the proof, as it is parallel to the case of scalar fields.

Recall that a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ can be written as a matrix if we fix a basis. We write $\boldsymbol{T}_{\boldsymbol{a}}$ as

$$\boldsymbol{T}_{\boldsymbol{a}} = \begin{pmatrix} D_1 f_1(\boldsymbol{a}) & D_2 f_1(\boldsymbol{a}) & \cdots & D_n f_1(\boldsymbol{a}) \\ D_1 f_2(\boldsymbol{a}) & D_2 f_2(\boldsymbol{a}) & \cdots & D_n f_2(\boldsymbol{a}) \\ \vdots & \vdots & & \vdots \\ D_1 f_m(\boldsymbol{a}) & D_2 f_m(\boldsymbol{a}) & \cdots & D_n f_m(\boldsymbol{a}) \end{pmatrix}$$

and called the **Jacobian matrix** of $\boldsymbol{f}$ at $\boldsymbol{a}$. It is sometimes denoted by $T_{\boldsymbol{a}} = \boldsymbol{f}'(\boldsymbol{a})$.

Example 41. Linear transformation. Let $\boldsymbol{f}(x, y) = (ax + by, cx + dy)$, then $\boldsymbol{T}_{\boldsymbol{a}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

Chain rule for vector fields

Note that, for a linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$, $\|T(\boldsymbol{v})\| \le M\|v\|$ for some $M > 0$. Indeed, we can write $T(v) = \sum_{k=1}^m \boldsymbol{u}_k \langle \boldsymbol{e}_k, v \rangle$ and hence $\|T(v)\| \le \sum \|\boldsymbol{u}_k\| \cdot \|v\|$.

Theorem 42. *Let $\boldsymbol{f} : \mathbb{R}^n \supset S \to T \subset \mathbb{R}^m$ and $\boldsymbol{g} : \mathbb{R}^m \supset T \to \mathbb{R}^\ell$. If $\boldsymbol{f}$ is differentiable at $\boldsymbol{a} \in S$ and $\boldsymbol{g}$ is differentiable at $\boldsymbol{f}(\boldsymbol{a})$. Then $\boldsymbol{h} = \boldsymbol{g} \circ \boldsymbol{f} : S \to \mathbb{R}^\ell$ is differentiable at $\boldsymbol{a}$ and $\boldsymbol{h}'(\boldsymbol{a}) = \boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{a})) \circ \boldsymbol{f}'(\boldsymbol{a})$, the composition of linear operators.*

Proof. We consider the difference $\boldsymbol{h}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{h}(\boldsymbol{a}) = \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y}))-\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a}))$. By the differentiablity of $\boldsymbol{g}$ and $\boldsymbol{f}$, there are $\boldsymbol{E_g}, \boldsymbol{E_f}$ such that

$$
\begin{aligned}
&\boldsymbol{h}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{h}(\boldsymbol{a}) \\
&= \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y}))-\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a})) \\
&= \boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{a}))(\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})) + \|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})\|\boldsymbol{E_g}(\boldsymbol{f}(\boldsymbol{a}),\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})) \\
&= \boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{a}))(\boldsymbol{f}'(\boldsymbol{a})(\boldsymbol{y}) + \|\boldsymbol{y}\|\boldsymbol{E_f}(\boldsymbol{a},\boldsymbol{y})) + \|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})\|\boldsymbol{E_g}(\boldsymbol{f}(\boldsymbol{a}),\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})) \\
&= \boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{a}))(\boldsymbol{f}'(\boldsymbol{a})(\boldsymbol{y})) + \boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{a}))(\|\boldsymbol{y}\|\boldsymbol{E_f}(\boldsymbol{a},\boldsymbol{y})) + \|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})\|\boldsymbol{E_g}(\boldsymbol{f}(\boldsymbol{a}),\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a}))
\end{aligned}
$$

and as $\boldsymbol{y}\to\boldsymbol{0}$, $\|\boldsymbol{y}\|\to 0$ and $\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})\to\boldsymbol{0}$, moreover, $\frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{a})\|}{\|y\|}$ is bounded. Hence by definition, $\boldsymbol{h}$ is differentiable and $\boldsymbol{h}'(\boldsymbol{a}) = \boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{a}))\circ\boldsymbol{f}'(\boldsymbol{a})$. □

Polar coordinates

An example of composition of vector field is given by a change of coordinages. Let $g(x,y)$ be a scalar field, and $x = X(r,\theta) = r\cos\theta, y = Y(r,\theta) = r\sin\theta$ be the polar coordinate, the map $\boldsymbol{f}(r,\theta) = (X(r,\theta), Y(r,\theta))$ can be considered as a vector field from $\mathbb{R}_+\times[0,2\pi)\to\mathbb{R}^2$. We would like to compute derivatives of $\varphi(r,\theta) = g(X(r,\theta),Y(r,\theta))$.

By the chain rule, $g'(\boldsymbol{x}) = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y})$ and

$$\boldsymbol{f}'(r,\theta) = \begin{pmatrix} \frac{\partial X}{\partial r} & \frac{\partial X}{\partial \theta} \\ \frac{\partial Y}{\partial r} & \frac{\partial Y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

we obtain

$$
\begin{aligned}
\frac{\partial\varphi}{\partial r}(r,\theta) &= \frac{\partial g}{\partial x}(r\cos\theta, r\sin\theta)\cos\theta + \frac{\partial g}{\partial y}(r\cos\theta, r\sin\theta)\sin\theta \\
\frac{\partial\varphi}{\partial \theta}(r,\theta) &= -\frac{\partial g}{\partial x}(r\cos\theta, r\sin\theta)r\sin\theta + \frac{\partial g}{\partial y}(r\cos\theta, r\sin\theta)r\cos\theta
\end{aligned}
$$

This can also be written as

$$
\begin{aligned}
\frac{\partial\varphi}{\partial r}(r,\theta)\cos\theta - \frac{1}{r}\frac{\partial\varphi}{\partial\theta}(r,\theta)\sin\theta &= \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) \\
\frac{\partial\varphi}{\partial r}(r,\theta)\sin\theta + \frac{1}{r}\frac{\partial\varphi}{\partial\theta}(r,\theta)\cos\theta &= \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)
\end{aligned}
$$

Sufficient condition for equality of mixed partial derivatives

Let f be a scalar field. In general, $D_1D_2f \neq D_2D_1f$. Let us take

$$f(x,y) = \frac{xy(x^2-y^2)}{x^2+y^2} \text{ for } (x,y)\neq(0,0), \quad 0 \text{ for } (x,y)=(0,0).$$

Then $D_1f(x,y) = \frac{y(x^4+4x^2y^2-y^4)}{(x^2+y^2)^2}$ for $(x,y)\neq(0,0)$ and $D_1f(0,0) = \lim_{h\to 0}\frac{0-0}{h} = 0$. Then, $D_2D_1f(0,0) = \lim_{h\to 0}\frac{h\cdot(-h^4)}{h^4\cdot h} = -1$. Similarly, $D_1D_2f(0,0) = 1$, so $D_1D_2f\neq D_2D_1f$.

Theorem 43. *Let f be a scalar field and assume that $D_1f, D_2, D_1D_2, D_2D_1f$ exist in an open set S. If $(a,b)\in S$ and D_1D_2f, D_2D_1f are both continuous at (a,b), then $D_1D_2f(a,b) = D_2D_1f(a,b)$.*

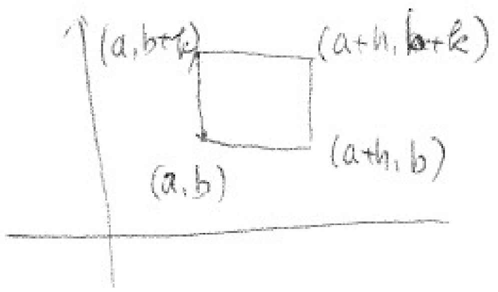
Proof. By the mean value theorem applied to $G(x) = f(x, b+k) - f(x, b), G'(x) = D_1 f(x, b+k) - D_1 f(x, b)$,

$$\begin{aligned}
&(f(a+h, b+k) - f(a+h, b)) - (f(a, b+k) - f(a, b)) \\
&= hG'(a + \theta_1 h) \\
&= h(D_1 f'(a + \theta_1 h, b+k) - D_1 f'(a + \theta_1 h, b)) \\
&= hk D_2 D_1 f(a + \theta_1 h, b + \varphi_1 k),
\end{aligned}$$

where $0 \leq \theta, \varphi \leq 1$, and we applied the mean value theorem to $D_1 f(a+\theta_1 h, y) = H(y)$. Similarly, $(f(a+h, b+k) - f(a+h, b)) - (f(a, b+k) - f(a, b)) = hk D_1 D_2 f(a + \theta_2 h, b + \varphi_2 k)$, hence

$$D_1 D_2 f(a + \theta_2 h, b + \varphi_2 k) = D_2 D_1 f(a + \theta_2 h, b + \varphi_2 k).$$

As $h, k \to 0$, this shows $D_1 D_2 f(a, b) = D_2 D_1 f(a, b)$.



□

Oct 16. Partial differential equations.

A partial differential equation is an equation about a scalar field or a vector field involving its partial derivatives.

Example 44. Some (linear) partial differential equations.

- $\frac{\partial f}{\partial t}(x, t) = k \frac{\partial^2 f}{\partial x^2}(x, y)$, where k is a constant (heat equation)
- $\frac{\partial^2 f}{\partial x^2}(x, y) + \frac{\partial^2 f}{\partial y^2}(x, y) = 0$ (Laplace's equation)
- $\frac{\partial^2 f}{\partial x^2}(x, t) - c^2 \frac{\partial^2 f}{\partial t^2}(x, t) = 0$, where c is a constant (wave equation)

Maxwell's equations, Navier-Stokes equations, Einstein's equations...

In general, PDE's have many solutions, and need to specify a boundary condition (or an initial condition):

Consider $\frac{\partial f}{\partial x}(x, y) = 0$. For any function $g(y)$, $f(x, y) = g(y)$ is a solution, and it holds that $f(0, y) = g(y)$. In general, such a condition is called a boundary condition.

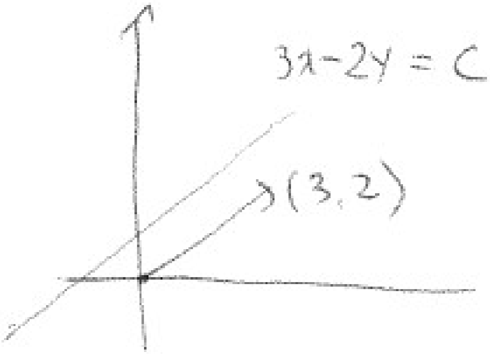
First order linear PDE

Let us consider $3\frac{\partial f}{\partial x}(x, y) + 2\frac{\partial f}{\partial y}(x, y) = 0$ and find all its solutions.

Recall that $\nabla f(x, y) = (\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y))$. The equation can be written as

$$(3\boldsymbol{e}_1 + 2\boldsymbol{e}_2) \cdot \nabla f(x, y) = 0.$$

We know that this is equivalent to $f'((x, y); 3\boldsymbol{e}_1 + 2\boldsymbol{e}_2) = 0$. In other words, f is constant along the vector $3\boldsymbol{e}_1 + 2\boldsymbol{e}_2$, and hence on the lines $2x - 3y = c$. The function $f(x, y)$ depends only on $2x - 3y$.



Actually, if g is any differentiable function, $f(x, y) = g(2x - 3y)$ is a solution. Indeed, by the chain rule, $\frac{\partial f}{\partial x}(x, y) = 2g'(2x-3y)$, $\frac{\partial f}{\partial y}(x, y) = -3g'(2x-3y)$ and hence $3\frac{\partial f}{\partial x}(x, y)+2\frac{\partial f}{\partial y}(x, y) = 0$. Therefore, we have proved that a *general solution* is $g(2x - 3y)$ for some differentiable function g.

Conversely, a general solution is of the form $g(2x - 3y)$. Indeed, let $u = 2y - x, v = 2x - 3y$. This can be solved: $x = 3u + 2v, y = 2u + v$ Define $h(u, v) = f(3u + 2v, 2u + v)$. We have

$$\frac{\partial h}{\partial u}(u, v) = 3\frac{\partial f}{\partial x}(3u + 2v, 2u + v) + 2\frac{\partial f}{\partial y}(3u + 2v, 2u + v) = 0.$$

Namely, h is only a function of v: $h(u, v) = g(v)$. Or $f(x, y) = g(2x - 3y)$.

With the same method, we can prove

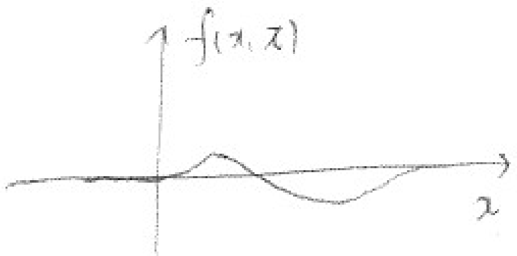
Theorem 45. *Let g be a differentiable function, $a, b \in \mathbb{R}, (a, b) \neq (0, 0)$. Define $f(x, y) = g(bx - ay)$. Then f satisfies the equation*

$$a\frac{\partial f}{\partial x}(x, y) + b\frac{\partial f}{\partial y}(x, y) = 0. \tag{1}$$

Conversely, every solution of (1) is of the form $g(bx - ay)$.

One-dimensional wave equation

Let x be the coordinate on a spanned string and t be the time and $f(x, t)$ be the displacement of the string at (x, t).



When $f(x, y)$ is small, it should satisfy

$$\frac{\partial^2 f}{\partial x^2}(x, t) = c^2\frac{\partial^2 f}{\partial t^2}(x, t), \tag{2}$$

where c is a constant which depends on the string. This can be derived from the equation of motion $m\frac{d^2\boldsymbol{r}}{dt^2} = F$, where $\boldsymbol{r}(t)$ is each small piece of the string and F is the tension of the string.

Theorem 46. *Let F be a twice differentiable function, G a differentiable function. Then*

$$f(x, t) = \frac{F(x + ct) + F(x - ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} G(s)ds \tag{3}$$

satisfies $\frac{\partial^2 f}{\partial^2 x}(x,t) = c^2 \frac{\partial^2 f}{\partial t^2}(x,t)$, $f(x,0) = F(x), \frac{\partial f}{\partial t}(x,0) = G(x)$. *Conversely, any solution of* (2) *is of the form, if* $\frac{\partial^2 f}{\partial x \partial t}(x,t) = \frac{\partial^2 f}{\partial t \partial x}(x,t)$.

Proof. Let $f(x,t)$ as above. Then

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,t) &= \frac{F'(x+ct)+F'(x-ct)}{2} + \frac{1}{2c}(G(x+ct)-G(x-ct)) \\
\frac{\partial^2 f}{\partial x^2}(x,t) &= \frac{F''(x+ct)+F''(x-ct)}{2} + \frac{1}{2c}(G'(x+ct)-G'(x-ct)) \\
\frac{\partial f}{\partial t}(x,t) &= \frac{cF'(x+ct)-cF'(x-ct)}{2} + \frac{1}{2}(G(x+ct)+G(x-ct)) \\
\frac{\partial^2 f}{\partial t^2}(x,t) &= \frac{c^2F''(x+ct)+c^2F''(x-ct)}{2} + \frac{c}{2}(G'(x+ct)-G'(x-ct))
\end{aligned}$$

therefore, $\frac{\partial^2 f}{\partial^2 x}(x,t) = c^2 \frac{\partial^2 f}{\partial^2 t}(x,t)$.

Conversely, assume that f satisfies (2). Introduce $u = x+ct, v = x-ct$. Then $x = \frac{u+v}{2}, t = \frac{u-v}{2c}$ and define $g(u,v) = f(x,t) = f(\frac{u+v}{2}, \frac{u-v}{2c})$. Then by the chain rule,

$$\begin{aligned}
\frac{\partial g}{\partial u}(u,v) &= \frac{1}{2}\frac{\partial f}{\partial x}(\tfrac{u+v}{2}, \tfrac{u-v}{2c}) + \frac{1}{2c}\frac{\partial f}{\partial t}(\tfrac{u+v}{2}, \tfrac{u-v}{2c}) \\
\frac{\partial^2 g}{\partial v \partial u}(u,v) &= \frac{1}{4}\frac{\partial^2 f}{\partial x^2}(\tfrac{u+v}{2}, \tfrac{u-v}{2c}) - \frac{1}{4c}\frac{\partial^2 f}{\partial x \partial t}(\tfrac{u+v}{2}, \tfrac{u-v}{2c}) + \frac{1}{4}\frac{\partial^2 f}{\partial x \partial t}(\tfrac{u+v}{2}, \tfrac{u-v}{2c}) - \frac{1}{4c}\frac{\partial^2 f}{\partial t^2}(\tfrac{u+v}{2}, \tfrac{u-v}{2c}) \\
&= 0
\end{aligned}$$

by the assumption. Therefore, $\frac{\partial g}{\partial u}(u,v) = \underline{\varphi}(u)$ and $g(u,v) = \varphi_1(u) + \varphi_2(v)$. In other words, $f(x,t) = \varphi_1(x+ct) + \varphi_2(x-ct)$. We define $f(x,0) = \varphi_1(x) + \varphi_2(x) =: F(x)$, then we have $F'(x) = \varphi_1'(x) + \varphi_2'(x)$, and furthermore, $\frac{\partial f}{\partial t}(x,t) = c\varphi_1(x+ct) - c\varphi_2(x-ct)$, and we define $\frac{\partial f}{\partial t}(x,0) = c\varphi_1'(x) - c\varphi_2'(x) =: G(x)$.

We can expresss $\varphi_1'(x), \varphi_2'(x)$ as $\varphi_1'(x) = \frac{1}{2}F'(x) + \frac{1}{2c}G(x), \quad \varphi_2'(x) = \frac{1}{2}F'(x) - \frac{1}{2c}G(x)$, or

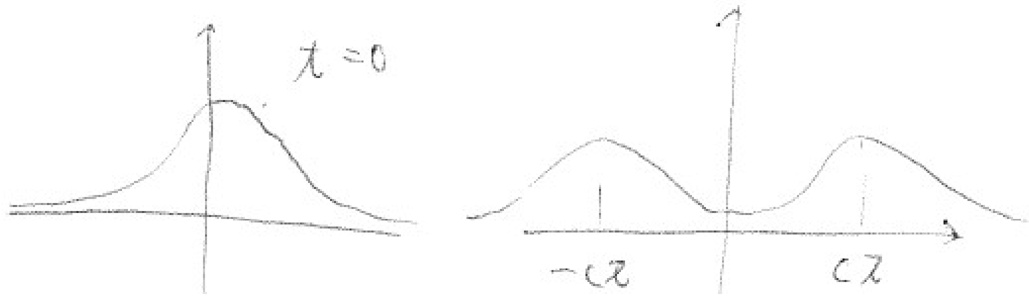
$$\varphi_1(y)-\varphi_1(0) = \frac{1}{2}(F(y)-F(0)) + \frac{1}{2c}\int_0^y G(s)ds, \varphi_2(y)-\varphi_2(0) = \frac{1}{2}(F(y)-F(0)) - \frac{1}{2c}\int_0^y G(s)ds.$$

and hence, by noting that $\varphi_1(0) + \varphi_2(0) = F(0)$,

$$\begin{aligned}
f(x,y) &= \varphi_1(x+ct) + \varphi_2(x-ct) \\
&= \varphi_1(0) + \varphi_2(0) + \frac{F(x+ct)+F(x-ct)-2F(0)}{2} + \int_{x-ct}^{x+ct} G(s)ds \\
&= \frac{F(x+ct)+F(x-ct)}{2} + \int_{x-ct}^{x+ct} G(s)ds.
\end{aligned}$$

□

Example 47. Take $F(x) = e^{-x^2}, G(x) = 0$. Then $f(x,t) = \frac{F(x+ct)+F(x-ct)}{2}$.



Oct 21. Implicit functions and partial derivatives

Recall that a **function** or a **scalar field** $f(\cdot)$ defined on a subset S of $\mathbb{R}^n$ assigns to each point $x \in S$ a real number $f(x)$, and it is represented by a curve or a surface.

Example 48. Explicitly given functions.

- $f(x) = x^2$
- $f(x, y) = \cos xe^y$
- $f(x, y, z) = e^{198x^7} z + (x - 2345)^{32} + (x^2 + 28)(y^3 - 2\pi)...$

Sometime a function is defined implicitly: consider the equation

$$x^2 + y^2 = 1.$$

This defines a circle. By solving this equation, we obtain

$$y = \pm\sqrt{1 - x^2}.$$

Namely, the curve $x^2 + y^2 = 0$ **defines implicitly** the function $y = f(x) = \pm\sqrt{1 - x^2}$. Similarly, the equation $x^2 + y^2 + z^2 = 1$ represents a sphere. It defines the function (scalar field) $z = f(x, y) = \pm\sqrt{1 - x^2 - y^2}$.

In general, if $F(x, y, z)$ is a function, the equation $F(x, y, z) = 0$ **may** define a function (but not always). Furthermore, even if it defines a function, it is **not always possible to solve it explicitly**. Can you solve the following equation in z?

$$F(x, y, z) = y^2 + xz + z^2 - e^z - 4 = 0$$

We **assume** that there is a function $f(x, y)$ such that $F(x, y, f(x, y)) = 0$. Even if **we do not know** the explicit form of $f(x, y)$, we can obtain **some information** about $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$.

Consider $g(x, y) = F(x, y, f(x, y)) = 0$ as a function of two variables x, y. Obviously we have $\frac{\partial g}{\partial x} = \frac{\partial g}{\partial y} = 0$. On the other hand, one can see it as

$$g(x, y) = F(u_1(x, y), u_2(x, y), u_3(x, y)) \quad \text{with } u_1(x, y) = x, \quad u_2(x, y) = y, \quad u_3(x, y) = f(x, y).$$

By the chain rule,

$$0 = \frac{\partial g}{\partial x}(x, y) = \frac{\partial F}{\partial x}(x, y, f(x, y)) \cdot 1 + \frac{\partial F}{\partial y}(x, y, f(x, y)) \cdot 0 + \frac{\partial F}{\partial z}(x, y, f(x, y)) \cdot \frac{\partial f}{\partial x}(x, y),$$

therefore,

$$\frac{\partial f}{\partial x}(x, y) = -\frac{\frac{\partial F}{\partial x}(x, y, f(x, y))}{\frac{\partial F}{\partial z}(x, y, f(x, y))}.$$

Similarly,

$$\frac{\partial f}{\partial y}(x, y) = -\frac{\frac{\partial F}{\partial y}(x, y, f(x, y))}{\frac{\partial F}{\partial z}(x, y, f(x, y))}.$$

Note that $F(x, y, z)$ is **explicitly given**.

Example 49. $F(x, y, z) = y^2 + xz + z^2 - e^z - C$, where $C \in \mathbb{R}$. Assume the existence of $f(x, y)$ such that $F(x, y, f(x, y)) = 0$. Find the value of C such that $f(0, e) = 2$ and compute $\frac{\partial f}{\partial x}(0, e), \frac{\partial f}{\partial y}(0, e)$.

Solution. $F(0, e, f(0, e)) = F(0, e, 2) = e^2 + 0 + 2^2 - e^2 - C = 0 \Longrightarrow C = 4$. Note that

- $\frac{\partial F}{\partial x}(x, y, z) = z$ and hence $\frac{\partial F}{\partial x}(x, y, f(x, y)) = f(x, y)$

- $\frac{\partial F}{\partial y}(x,y,z) = 2y$ and hence $\frac{\partial F}{\partial y}(x,y,f(x,y)) = 2y$
- $\frac{\partial F}{\partial z}(x,y,z) = x + 2z - e^z$ and hence $\frac{\partial F}{\partial z}(x,y,f(x,y)) = x + f(x,y) - e^{f(x,y)}$

Therefore,

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y) &= -\frac{\frac{\partial F}{\partial x}(x,y,f(x,y))}{\frac{\partial F}{\partial z}(x,y,f(x,y))} = -\frac{f(x,y)}{x+2f(x,y)-e^{f(x,y)}} \\
\frac{\partial f}{\partial x}(0,e) &= -\frac{f(0,e)}{0+2f(0,e)-e^{f(0,e)}} = \frac{2}{e^2-4}. \\
\frac{\partial f}{\partial y}(x,y) &= -\frac{\frac{\partial F}{\partial y}(x,y,f(x,y))}{\frac{\partial F}{\partial z}(x,y,f(x,y))} = -\frac{2y}{x+2f(x,y)-e^{f(x,y)}} \\
\frac{\partial f}{\partial y}(0,e) &= -\frac{2e}{0+2f(0,e)-e^{f(0,e)}} = \frac{2e}{e^2-4}.
\end{aligned}$$

More generally, if $F(x_1, \cdots, x_n) = 0$ defines a function $x_n = f(x_1, \cdots, x_{n-1})$, then

$$\frac{\partial f}{\partial x_k}(x_1, \cdots, x_{n-1}) = -\frac{\frac{\partial F}{\partial x_k}(x_1, \cdots, x_{n-1}, f(x_1, \cdots, x_{n-1}))}{\frac{\partial F}{\partial x_n}(x_1, \cdots, x_{n-1}, f(x_1, \cdots, x_{n-1}))}.$$

Next, let us consider two surfaces $F(x,y,z) = 0$ and $G(x,y,z) = 0$ and assume that their intersection is a curve $(X(z), Y(z), z)$. Namely, $F(X(z), Y(z), z) = 0, G(X(z), Y(z), z) = 0$.

Example 50. The unit sphere $F(x,y,z) = x^2 + y^2 + z^2 = 0$ and the xz-plane $G(x,y,z) = y = 0$ has the intersection $x^2 + z^2 = 0 \Longrightarrow x = X(z) = \pm\sqrt{1-z^2}, Y(z) = 0$.

Even if $X(z)$ and $Y(z)$ are only **implicitly given**, we can compute their derivatives. As before, put $f(z) = F(X(z), Y(z), z) = 0, g(z) = G(X(z), Y(z), z) = 0$. By the chain rule,

$$0 = f'(z) = X'(z)\frac{\partial F}{\partial x}(X(z), Y(z), z) + Y'(z)\frac{\partial F}{\partial x}(X(z), Y(z), z) + \frac{\partial F}{\partial z}(X(z), Y(z), z).$$

Similarly,

$$0 = g'(z) = X'(z)\frac{\partial G}{\partial x}(X(z), Y(z), z) + Y'(z)\frac{\partial G}{\partial x}(X(z), Y(z), z) + \frac{\partial G}{\partial z}(X(z), Y(z), z).$$

From these, we obtain

$$\begin{pmatrix} X'(z) \\ Y'(z) \end{pmatrix} = \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{pmatrix}^{-1} (X(z), Y(z), z) \cdot \begin{pmatrix} -\frac{\partial G}{\partial z} \\ -\frac{\partial G}{\partial z} \end{pmatrix} (X(z), Y(z), z).$$

Example 51. Computations of partial derivatives.

- $x = u + v, y = uv^2$ defines $u(x,y), v(x,y)$. Compute $\frac{\partial v}{\partial x}$.

  Solution. By eliminating u, we obtain $xv^2 - v^3 - y = 0$. In other words, $F(x,y,v) = 0$ where $F(x,y,v) = xv^2 - v^3 - y$. By the formula above, with $\frac{\partial F}{\partial x} = v^2, \frac{\partial F}{\partial v} = 2xv - 3v^2$, $\frac{\partial v}{\partial x} = -\frac{v(x,v)^2}{2xv(x,y) - 3v(x,y)^2}$.

- Assume that $g(x,y) = 0$ defines implicitly $Y(x)$. Let $f(x,y)$ be another function. Then $h(x) = f(x, Y(x))$ is a function of x. By the chain rule,

$$\begin{aligned}
h'(x) &= \frac{\partial f}{\partial x}(x, Y(x)) + \frac{\partial f}{\partial y}(x, Y(x))Y'(x) \\
&= \frac{\partial f}{\partial x}(x, Y(x)) - \frac{\frac{\partial g}{\partial x}(x, Y(x))}{\frac{\partial g}{\partial y}(x, Y(x))}\frac{\partial f}{\partial y}(x, Y(x))
\end{aligned}$$

- Let u be defined by $F(u+x, yu) = u$. Let $u = g(x,y)$, then $g(x,y) = F(g(x,y)+x, yg(x,y))$, and

$$\begin{aligned}
&\frac{\partial g}{\partial x}(x,y) \\
&= \frac{\partial F}{\partial X}(g(x,y)+x, yg(x,y))\left(\frac{\partial g}{\partial x}(x,y)+1\right) + \frac{\partial F}{\partial Y}(g(x,y)+x, yg(x,y)) \cdot y\frac{\partial g}{\partial x}(x,y) \\
&\Longrightarrow \frac{\partial g}{\partial x}(x,y) = \frac{-\frac{\partial F}{\partial X}(g(x,y)+x, yg(x,y))}{\frac{\partial F}{\partial Y}(g(x,y)+x, yg(x,y)) + y\frac{\partial F}{\partial X}(g(x,y)+x, yg(x,y)) - 1}
\end{aligned}$$

- $2x = v^2 - u^2, y = uv$ defines implicitly $u(x,y), v(x,y)$ (it is also possible to solve them: $2x + (\frac{y}{v})^2 = v^2, (\frac{y}{u})^2 - u^2 = 2x$). Compute $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}$.
  Solution. By differentiating with respect to x,
  $$2 = 2v\frac{\partial v}{\partial x} - 2u\frac{\partial u}{\partial x}, \qquad 0 = \frac{\partial u}{\partial x}v + u\frac{\partial v}{\partial x}$$
  From which one obtains
  $$\frac{\partial u}{\partial x} = -\frac{u}{u^2+v^2}, \qquad \frac{\partial v}{\partial x} = \frac{v}{u^2+v^2}.$$
  Similarly,
  $$\frac{\partial u}{\partial y} = \frac{v}{u^2+v^2}, \qquad \frac{\partial v}{\partial y} = \frac{u}{u^2+v^2}.$$

Oct 23. Minima, maxima and saddle points

Various extremal points

Let $S \subset \mathbb{R}^n$ be an open set, $f : S \to \mathbb{R}$ be a scalar field and $\boldsymbol{a} \in S$. Recall that $B(\boldsymbol{a}, r) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{a}\| < r\}$ is the r-ball centered at $\boldsymbol{a}$.

Definition 52. (Minima and maxima)

- If $f(\boldsymbol{a}) \le f(\boldsymbol{x})$ (respectively $f(\boldsymbol{a}) \ge f(\boldsymbol{x})$) for all $\boldsymbol{x} \in S$, then $f(\boldsymbol{a})$ is said to be the **absolute** minimum (resp. maximum) of f.

- If $f(\boldsymbol{a}) \le f(\boldsymbol{x})$ (respectively $f(\boldsymbol{a}) \ge f(\boldsymbol{x})$) for $\boldsymbol{x} \in B(\boldsymbol{a}, r)$ for some r, then $f(\boldsymbol{a})$ is said to be a **relative** minimum (resp. maximum)

Theorem 53. *If f is differentiable and has a relative minumum (resp. maximum) at $\boldsymbol{a}$, then $\nabla f(\boldsymbol{a}) = 0$.*

Proof. We prove the statement only for a relative minumum, because the other case is analogous. For any unit vector $\boldsymbol{y}$, consider $g(u) = f(\boldsymbol{a} + u\boldsymbol{y})$. As $\boldsymbol{a}$ is a relative minumum, g has a relative minumum at $u = 0$, therefore, $g'(0) = 0$, and $f'(\boldsymbol{a}; \boldsymbol{y}) = 0$ for any $\boldsymbol{y}$. This implies that $\nabla f(\boldsymbol{a}) = \mathbf{0}$. □

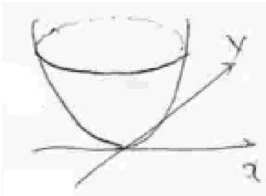
Remark 54. $\nabla f(\boldsymbol{a}) = \mathbf{0}$ does not imply that f takes a relative minumum or maximum at $\boldsymbol{a}$. Even in $\mathbb{R}$, $f(x) = x^3$ has $f'(0) = 0$ but 0 is not a relative minumum either a relative maximum.

Definition 55. (Stationary points)

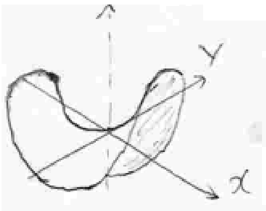
- If $\nabla f(\boldsymbol{a}) = \mathbf{0}$, then $\boldsymbol{a}$ is called a **stationary point**.
- If $\nabla f(\boldsymbol{a}) = \mathbf{0}$ and $\boldsymbol{a}$ is neither a relative minumum nor a relative maximum, then $\boldsymbol{a}$ is called a **suddle point**.

Example 56. (Stationary points)

- $f(x, y) = x^2 + y^2$. $\nabla f(\boldsymbol{x}) = (2x, 2y), \nabla f(0, 0) = (0, 0)$. $f(0, 0)$ is the absolute minumum.



- $f(x, y) = xy$. $\nabla f(\boldsymbol{x}) = (y, x), \nabla f(0, 0) = 0$. $f(0, 0)$ is a saddle:
  - $x > 0, y > 0$, then $f(x, y) > 0$.
  - $x > 0, y < 0$, then $f(x, y) < 0$.



Second-order Taylor formula

Let f be a differentiable function. We learned that $f(\boldsymbol{a} + \boldsymbol{y}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a}) \cdot \boldsymbol{y} + \|\boldsymbol{y}\| E_1(\boldsymbol{a}, \boldsymbol{y})$ and $E_1(\boldsymbol{a}, \boldsymbol{y}) \to 0$ as $\|\boldsymbol{y}\| \to 0$.

Let f have continuous second partial derivatives and let us denote them by $D_{ij} f = \frac{\partial^2 f}{\partial x_i, \partial x_j}$. Define the **Hessian matrix** by

$$H(\boldsymbol{x}) = \begin{pmatrix} D_{11}f(\boldsymbol{x}) & D_{12}f(\boldsymbol{x}) & \cdots & D_{1n}f(\boldsymbol{x}) \\ \vdots & & & \vdots \\ D_{n1}f(\boldsymbol{x}) & D_{n2}f(\boldsymbol{x}) & \cdots & D_{nn}f(\boldsymbol{x}) \end{pmatrix}$$

This is a *real symmetric matrix.* For $\boldsymbol{y} = (y_1, \cdots, y_n)$, $\boldsymbol{y} H(\boldsymbol{x}) \boldsymbol{y}^{\mathrm{t}} \in \mathbb{R}$.

Theorem 57. *Let f be a scalar field with continuous second partial derivatives on $B(\boldsymbol{a}; r)$. Then, for $\boldsymbol{y}$ such that $\boldsymbol{a} + \boldsymbol{y} \in B(\boldsymbol{a}; r)$ there is $0 < c < 1$ such that*

$$f(\boldsymbol{a} + \boldsymbol{y}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a}) \cdot \boldsymbol{y} + \frac{1}{2} \boldsymbol{y} H(\boldsymbol{a} + c\boldsymbol{y}) \boldsymbol{y}^{\mathrm{t}},$$
$$f(\boldsymbol{a} + \boldsymbol{y}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a}) \cdot \boldsymbol{y} + \frac{1}{2} \boldsymbol{y} H(\boldsymbol{a}) \boldsymbol{y}^{\mathrm{t}} + \|\boldsymbol{y}\|^2 \cdot E_2(\boldsymbol{a}, \boldsymbol{y})$$

and $E_2(\boldsymbol{a}, \boldsymbol{y}) \to 0$ as $\|\boldsymbol{y}\| \to 0$.

Proof. Let us define $g(u) = f(\boldsymbol{a} + u\boldsymbol{y})$. We apply the Taylor formula to g to get $g(1) = g(0) + g'(0) + \frac{1}{2} g''(c)$ with $0 < c < 1$. Since we have $g(u) = f(a_1 + uy_1, \cdots, a_n + uy_n)$, by chain rule,

$$g'(u) = \sum_{j=1}^{n} D_j f(a_1 + uy_1, \cdots, a_n + uy_n) y_j = \nabla f(\boldsymbol{a} + u\boldsymbol{y}) \cdot \boldsymbol{y},$$

where $D_j f = \frac{\partial f}{\partial x_j}$. Similarly,

$$g''(u) = \sum_{i,j=1}^{n} D_{ij} f(a_1 + uy_1, \cdots, a_n + uy_n) y_i y_j = \boldsymbol{y} H(\boldsymbol{a} + u\boldsymbol{y}) \boldsymbol{y}^{\mathrm{t}},$$

from which the first equation follows. As for the second equation, we define E_2 by $E_2(\boldsymbol{a}, \boldsymbol{y}) = \frac{1}{2}(\boldsymbol{y}H(\boldsymbol{a} + c\boldsymbol{y}) - H(\boldsymbol{a})^{\mathrm{t}})\boldsymbol{y}^{\mathrm{t}}/\|\boldsymbol{y}\|^2$. Then

$$|E_2(\boldsymbol{a}, \boldsymbol{y})| \leq \frac{1}{2} \sum_{i,j=1}^{n} \frac{|y_i y_j|}{\|\boldsymbol{y}\|^2} |D_{ij} f(\boldsymbol{a} + c\boldsymbol{y}) - D_{ij} f(\boldsymbol{a})| \to 0$$

as $\|y\| \to 0$, by the continuity of $D_{ij} f(\boldsymbol{a})$.

□

Classifiying stationary points

We give a criterion to deterine whether $\boldsymbol{a}$ is a minumum/maximum/saddle when $\nabla f(\boldsymbol{a}) = \boldsymbol{0}$.

Theorem 58. *Let* $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ *be a real symmetric matrix, and* $Q(\boldsymbol{y}) = \boldsymbol{y}A\boldsymbol{y}^{\mathrm{t}}$. *Then,*

- $Q(\boldsymbol{y}) > 0$ *for all* $\boldsymbol{y} \neq \boldsymbol{0}$ *if and only if all eigenvalues of* A *are positive.*
- $Q(\boldsymbol{y}) < 0$ *for all* $\boldsymbol{y} \neq \boldsymbol{0}$ *if and only if all eigenvalues of* A *are negative.*

Proof. A real symmetric matrix A can be diagonalized by an orthogonal matrix C, namely, $L = C^{\mathrm{t}}AC = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$. If all $\lambda_j > 0$, then $Q(\boldsymbol{y}) = \boldsymbol{y}CC^{\mathrm{t}}ACC^{\mathrm{t}}\boldsymbol{y}^{\mathrm{t}} = \boldsymbol{v}L\boldsymbol{v}^{\mathrm{t}} = \sum_j \lambda_j v_j > 0$, where $\boldsymbol{v} = \boldsymbol{y}C$. If $Q(\boldsymbol{y}) > 0$ for all $\boldsymbol{y} \neq \boldsymbol{0}$, then especially for $\boldsymbol{y}_k = \boldsymbol{u}_k C$ where $\boldsymbol{u}_k = (0, \cdots, 0, \underset{k\text{-th}}{1}, 0, \cdots, 0)$, and $Q(\boldsymbol{y}_k) = \lambda_k > 0$. □

Theorem 59. *Let* f *be a scalar field with continuous second derivatives on* $B(\boldsymbol{a}; r)$. *Assume that* $\nabla f(\boldsymbol{a}) = \boldsymbol{0}$. *Then,*

(a) If all the eigenvalues λ_j *of* $H(\boldsymbol{a})$ *are positive, then* f *has a relative minumum at* $\boldsymbol{a}$.

(b) If all the eigenvalues λ_j *of* $H(\boldsymbol{a})$ *are negative, then* f *has a relative maximum at* $\boldsymbol{a}$.

(c) If some $\lambda_k > 0$ *and* $\lambda_\ell <$, *then* $\boldsymbol{a}$ *is a saddle.*

Proof. (a) Let $Q(\boldsymbol{y}) = \boldsymbol{y}H(\boldsymbol{a})\boldsymbol{y}^{\mathrm{t}}$. Let h be te smallest eigenvalue of $H(\boldsymbol{a})$, $h > 0$ and diagonalize $h(\boldsymbol{a})$ by C. We set $\boldsymbol{y}C = \boldsymbol{v}$, then $\|\boldsymbol{y}\| = \|\boldsymbol{v}\|$. Furthermore,

$$\boldsymbol{y}H(\boldsymbol{a})\boldsymbol{y}^{\mathrm{t}} = \boldsymbol{v}CH(\boldsymbol{a})C^{\mathrm{t}}\boldsymbol{v}^{\mathrm{t}} = \sum_j \lambda_j v_j^2 > h \sum_j v_j^2 = h\|\boldsymbol{v}\|^2 = h\|\boldsymbol{y}\|^2.$$

By Theorem 57,

$$f(\boldsymbol{a} + \boldsymbol{y}) = f(\boldsymbol{a}) + \boldsymbol{y}H(\boldsymbol{a})\boldsymbol{y}^{\mathrm{t}} + \|\boldsymbol{y}\|^2 E_2(\boldsymbol{a}, \boldsymbol{y}).$$

As $H_2(\boldsymbol{a}, \boldsymbol{y}) \to 0$ as $\|\boldsymbol{y}\| \to 0$, there is r_1 such that if $\|\boldsymbol{y}\| < r_1$, then $|E_2(\boldsymbol{a}, \boldsymbol{y})| < \frac{h}{2}$. Now

$$f(\boldsymbol{a} + \boldsymbol{y}) = f(\boldsymbol{a}) + \frac{1}{2}\boldsymbol{y}H(\boldsymbol{a})\boldsymbol{y}^{\mathrm{t}} + \|\boldsymbol{y}\|^2 E_2(\boldsymbol{a}, \boldsymbol{y}) > f(\boldsymbol{a}) + \frac{h}{2}\|\boldsymbol{y}\|^2 - \frac{h}{2}\|\boldsymbol{y}\| > f(\boldsymbol{a}),$$

hence f has a relative minumum at $\boldsymbol{a}$.

(b) This case is similar as above.

(c) Let $\boldsymbol{y}_k$ be an eigenvector with eigenvalue λ_k, $\boldsymbol{y}_\ell$ be an eigenvector with eigenvalue λ_ℓ. As in (a), $f(\boldsymbol{a}+c\boldsymbol{y}_k) > f(\boldsymbol{a})$ and as in (b) $f(\boldsymbol{a}+c\boldsymbol{y}_\ell) < f(\boldsymbol{a})$ for small c, hence $\boldsymbol{a}$ is a saddle. $\square$

Example 60. $f(x,y) = xy$. $\lambda\varphi(x,y) = (y,x), H(x,y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ $(0,0)$ is a stationary point and $H(0,0)$ has eigenvalues $1,-1$, hence $(0,0)$ is a saddle.