Dr. Florian Oschmann

Karlova univerzita
Zimní semestr 2025/26

Mathematical analysis II

Lesson from 12./13.11.2025

---

1 Jacobi and Hessian matrix

I'll start with the question "what is the intuition and meaning behind Jacobi and Hessian matrices?"

Starting from 1D: if $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then $f'(x) = \frac{df}{dx}(x)$ is simply a one-dimensional function; you can view it as one single point/place somewhere in the universe. Going to higher dimensions $f : \mathbb{R}^n \to \mathbb{R}^m$ for some $n, m \geq 1$, where now

$$f(x) = f(x_1, ..., x_n) = (f_1, ..., f_m)(x_1, ..., x_n) = (f_1(x), ..., f_m(x)),$$

then we have many places in the universe where different derivatives $\partial_{x_i} f_j = \frac{\partial f_j}{\partial x_i}$ lie. You can compare this with mushrooms in the forest: everywhere is something, on the left you see some yellow $\partial_{x_1} f_2$, and two meters ahead there is a poisonous $\partial_{x_{10}} f_7$. The Jacobi matrix

$$J_f(x) = \begin{pmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_m(x) \end{pmatrix} = \begin{pmatrix} \partial_{x_1} f_1(x) & \dots & \partial_{x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} f_m(x) & \dots & \partial_{x_n} f_m(x) \end{pmatrix}$$

does nothing else than to collect and order these derivatives; it works like a basket with specified slots for each mushroom. The specific case $m = 1$ turns the Jacobi matrix into the usual gradient (say, you just have one type of mushroom and order all of them by color in a row). Applications are, as we know, the implicit function theorem, inverse function theorem (here especially $n = m$ and the determinant $\det J_f$ is important), finding extrema, and, later, change of variables in multi-dimensional integrals. If f is totally differentiable in some point x_0, then this total differential $Df_{x_0}(h)$ is exactly $Df_{x_0}(h) = J_f(x_0) \cdot h$ and especially a linear function of h.

The Hessian matrix $H_f(x)$ does roughly the same, but for *second* derivatives. An important change here is that H_f is defined *just for functions mapping into the reals*, that is, $f : \mathbb{R}^n \to \mathbb{R}$ reps. $m = 1$ in the above (for $m \geq 2$ the Hessian matrix is not defined). It is defined as

$$H_f(x) = \begin{pmatrix} \partial^2_{x_1 x_1} f(x) & \partial^2_{x_1 x_2} f(x) & \dots & \partial^2_{x_1 x_n} f(x) \\ \partial^2_{x_2 x_1} f(x) & \partial^2_{x_2 x_2} f(x) & \dots & \partial^2_{x_2 x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2_{x_n x_1} f(x) & \partial^2_{x_n x_2} f(x) & \dots & \partial^2_{x_n x_n} f(x) \end{pmatrix},$$

and sometimes $\partial^2_{x_i x_j}$ is abbreviated as $\partial_{x_i x_j}$ as you will see in the sequel. Note especially that H_f is always a square matrix, and if second derivatives are continuous, then $H_f = H_f^T$, meaning it is symmetric. Moreover, it helps in classifying extrema:

Again starting from 1D, if a function $f : \mathbb{R} \to \mathbb{R}$ has an extremum in x_0, then $f'(x_0) = 0$ (in higher dimensions this translates into the Jacobi matrix resp. the gradient above). To figure out whether it is a Max or a Min, we can ask what $f''(x_0)$ does: if $f''(x_0) > 0$, then x_0 is a local Min; if $f''(x_0) < 0$, then it is a local Max. In higher dimensions, the second derivative turns into H_f, and the inequalities $<$ and $>$ translate into definiteness of H_f. Indeed, if $H_f(x_0)$ is positive definite, then x_0 is a local Min of f, and if it's negative definite, then we have a local Max. Furthermore, if H_f is indefinite, then x_0 is a saddle point. Note that this needs *strict* definiteness, meaning, if H_f is just (positive or negative) *semi-definite*, then nothing can be said. Definiteness plays also a role in convexity: if H_f is positive [negative] definite everywhere in the domain of definition of f, then f is (strictly) convex [concave] there. There are several criteria to figure out whether a matrix is positive or negative definite, probably the most known is the determinant or Hurwitz (also named Silvester's or principal minors) criterion, which I will not speak about here; see linear algebra/discrete mathematics for this.

2 Taylor polynomial again

Let's try to figure out the Taylor polynomial of degree 3 at the point $(x, y) = (0, 0)$ of the function

$$f(x, y) = (x + x^2)\sin y.$$

To do so, we first need all derivatives up to third order; thus

$$\begin{aligned} \partial_x f &= (1 + 2x)\sin y, & \partial_{xx} f &= 2\sin y, & \partial_{xxx} f &= 0,\\ \partial_y f &= (x + x^2)\cos y, & \partial_{yy} f &= -f, & \partial_{yyy} f &= -\partial_y f. \end{aligned}$$

Not to forget the mixed partial derivatives (I will already use that they are continuous so the order of differentiation doesn't matter):

$$\begin{aligned} \partial_{xy} f &= (1 + 2x)\cos y = \partial_{yx} f,\\ \partial_{xxy} f &= 2\cos y = \partial_{xyx} f = \partial_{yxx} f,\\ \partial_{xyy} f &= -(1 + 2x)\sin y = \partial_{yxy} f = \partial_{yyx} f. \end{aligned}$$

Now the Taylor polynomial takes the general form (compare the powers of each term with the order and number of derivatives coming with it, see also explanation below)

$$\begin{aligned} T^3_{f;(x_0,y_0)}(x, y) = f(x_0, y_0) &+ \partial_x f(x_0, y_0)(x - x_0) + \partial_y f(x_0, y_0)(y - y_0)\\ &+ \frac{1}{2!}\Big[\partial_{xx} f(x_0, y_0)(x - x_0)^2 + \partial_{xy} f(x_0, y_0)(x - x_0)(y - y_0)\\ &\quad + \partial_{yx} f(x_0, y_0)(y - y_0)(x - x_0) + \partial_{yy} f(x_0, y_0)(y - y_0)^2\Big]\\ &+ \frac{1}{3!}\Big[\partial_{xxx} f(x_0, y_0)(x - x_0)^3 + \partial_{xxy} f(x_0, y_0)(x - x_0)^2(y - y_0)\\ &\quad + \partial_{xyx} f(x_0, y_0)(x - x_0)^2(y - y_0) + \partial_{yxx} f(x_0, y_0)(x - x_0)^2(y - y_0)\\ &\quad + \partial_{xyy} f(x_0, y_0)(x - x_0)(y - y_0)^2 + \partial_{yxy} f(x_0, y_0)(x - x_0)(y - y_0)^2\\ &\quad + \partial_{yyx} f(x_0, y_0)(x - x_0)(y - y_0)^2 + \partial_{yyy} f(x_0, y_0)(y - y_0)^3\Big]. \end{aligned}$$

I highlighted the terms that will survive for us since for the given function f, all other derivatives in $(x_0, y_0) = (0, 0)$ will be zero. Hence, the Taylor polynomial collapses to

$$T^3_{f;(0,0)}(x, y) = xy + x^2 y = (x + x^2)y.$$

Compare the form of the polynomial with the number and order of derivatives: $\partial_{xxy} f(x_0, y_0)$ has 2x the derivative wrt. first variable x and 1x wrt. second one y, hence, the variables coming with this derivative are 2x the first element of the vector $(x - x_0, y - y_0)$ and 1x the second element. Thus, $\partial_{xxy} f(x_0, y_0)$ comes with $(x - x_0)(x - x_0)(y - y_0) = (x - y_0)^2(y - y_0)$. Similarly for $\partial_{yxy} f$, this comes with 1x the second element, 1x the first, and again 1x the second, thus $\partial_{yxy} f(x_0, y_0)$ comes with $(y - y_0)(x - x_0)(y - y_0) = (x - x_0)(y - y_0)^2$. Higher order polynomials work in the same way; essentially, it's just counting and comparing.

3 Implicit function theorem (IFT) again

First a word to "why do we care": say we stand on some mountain at a specific point and we want to walk around the mountain, but without going down or up to reduce the energy we need. Which way shall we take? Even worse, does there even exist such a way?

The mountain profile is a function F that assigns to each point (x, y) in the plane a specific height $F(x, y)$. The point where we are standing is (x_0, y_0), and we assume that $F(x_0, y_0) = 0$ (if we are standing higher, we can simply shift our function by this specific height in order to force it to be zero). If the way would be known (by other tourist and signs), then we can write it as a function (a path) $g : \mathbb{R} \to \mathbb{R}$ with $g(x_0) = y_0$, and since this is the way not going up or down, we have $F(x, g(x)) = 0$ for each x. But how do we know g? To figure this out is exactly the outcome of IFT. It can be seen as the question to "does such a way exist and, if yes, how it looks like" instead of "let's follow the signs".

The assumption $\det \frac{\partial F}{\partial y}(x_0, y_0) \neq 0$ means nothing else than there are no crossings in the profile; in other words, if such a path g exists, then it must be unique, at least in some neighborhood of the point we're standing in. If such a crossing would happen, then the function g gives for one single x two or more values for y, thus it will be multivalued and not even a function (in this sense we might take the wrong way and get lost in the mountains). Mathematically speaking, we search for even more solutions to an equation we already know one solution to. Our known solution is (x_0, y_0), and these more solutions are $(x, g(x))$.

Let's now try to find paths. Making the task a bit harder, we have:

Exercise 1. Show that there is some $\delta > 0$ and a function $g : (-\delta, \delta) \to \mathbb{R}$ with $g(0) = \pi$ such that

$$2 \sin g(x) = 3xe^{g(x)}.$$

How the tangent of g on this point looks like? What can you say about monotonicity of g in this point?

Solution. (I recommend to put $2 \sin y - 3xe^y = 0$ into 3D geogebra and have a look on how the paths look like to see what we actually do here; a picture always helps.) We define

$$F(x, y) = 2 \sin y - 3xe^y.$$

(The negative would do as well.) If such a function g exists, it must fulfil $F(x, g(x)) = 0$ for any $x \in (-\delta, \delta)$. Let's first check that $F(0, \pi) = 0$, so the point on the mountain in which we're standing is a good one. Next,

$$\partial_x F = -3e^y, \qquad \partial_y F = 2 \cos y - 3xe^y$$

such that partial derivatives are continuous and $\partial_y F(0, \pi) = -2 \neq 0$ is invertible. Hence, IFT guarantees the existence of some $\delta > 0$ and a continuously differentiable function g fulfilling what we want (that is, at least a (maybe very short) path exists).

To the tangent: We know that $F(x, g(x)) = 0$ for any $x \in (-\delta, \delta)$. Differentiating both sides of this equality wrt. x, using that the derivative of zero stays zero, and chain rule gives (again, I use ∂_1 and ∂_2 to indicate differentiation wrt. first and second variable, respectively)

$$0 = \frac{d}{dx} F(x, g(x)) = \partial_1 F(x, g(x)) + \partial_2 F(x, g(x)) g'(x).$$

Let me do this in a bit more detailed and formal way: define the function $\Phi : (-\delta, \delta) \to \mathbb{R}^2$ by $\Phi(x) = (\Phi_1(x), \Phi_2(x)) := (x, g(x))$. Then Φ is a path (what a surprise in view of how we called g on the mountain) and it is differentiable since g is with derivative $\Phi'(x) = (1, g'(x))$. Moreover, we have $0 = F(\Phi_1(x), \Phi_2(x)) = F(\Phi(x))$ such that chain rule gives

$$\begin{aligned} 0 &= \nabla_{1,2} F(\Phi_1(x), \Phi_2(x)) \cdot \Phi'(x) = \begin{pmatrix} \partial_1 F(\Phi_1(x), \Phi_2(x)) \\ \partial_2 F(\Phi_1(x), \Phi_2(x)) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ g'(x) \end{pmatrix} \\ &= \partial_1 F(x, g(x)) + \partial_2 F(x, g(x)) g'(x), \end{aligned}$$

which is the formula above. From here, we easily derive

$$0 = \partial_1 F(x, g(x)) + \partial_2 F(x, g(x)) g'(x) \Leftrightarrow g'(x) = -\frac{\partial_1 F(x, g(x))}{\partial_2 F(x, g(x))}.$$

Note what this means: without even knowing the exact form of g, we can calculate its derivative in the given point! (Theoretically, in super super lucky situations you can solve the differential equation to get the exact form of g, but in practice, this never happens.) By this, we find

$$g'(x) = -\frac{-3e^y}{2\cos y - 3xe^y}\bigg|_{(x,g(x))} = -\frac{-3e^{g(x)}}{2\cos g(x) - 3xe^{g(x)}}$$

and thus, by $g(0) = \pi$, we have

$$g'(0) = -\frac{-3e^\pi}{2\cos \pi - 3 \cdot 0 \cdot e^\pi} = -\frac{3}{2} e^\pi < 0,$$

and in turn g is (strictly) decreasing in $x = 0$, which answers the question about monotonicity. Moreover, the tangent is given by

$$t_{g;x_0=0}(x) = g(x_0) + g'(x_0)(x - x_0) = \pi - \frac{3}{2} e^\pi x.$$

Note that the fact that g is decreasing *does not mean* that the "mountain" described by F goes down! It rather tells that g comes from the left, passes the point $(0, \pi)$ and goes to the right in the form of a "U", but all at the same height of the mountain. In differentiating again, we could get expressions for $g''(0), g'''(0)$, and so on such that we might write down the Taylor polynomial of g to get better and better approximations; see the formula for g'' from the example I gave in class, or differentiate the expression for g' given above once more wrt. x to get the explicit outcome (chain rule + quotient rule!).

3.1 Where do functions live?

(This section is just for interested readers and not part of any HW or exam.)

One question for IFT was "do we know how large such δ (or in general, the neighborhood $\hat{U}$ of some point $x_0 \in \mathbb{R}^n$ guaranteed by IFT) can be?" This question is in general (a bit informally) answered by the following:

Let $F : U \times V \to \mathbb{R}^n$, where $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, and let $(x_0, y_0) \in U \times V$ such that $\det \frac{\partial F}{\partial y}(x_0, y_0) \neq 0$. Let $\hat{U} \subset U, \hat{V} \subset V$, and $g : \hat{U} \to \hat{V}$ with $g(x_0) = y_0$ be the sets and function whose existence is guaranteed by IFT. Then the largest neighborhood $\hat{U}_{\max} \subset U$ where g lives (i.e., where it is defined) are all points x that can be connected to x_0 with a (possibly curved) line such that $\det \frac{\partial F}{\partial y}(x, g(x)) \neq 0$ on those lines. More formally, since g is dependent on x_0, let me call it g_{x_0}. Then

$$\hat{U}_{\max} = \{x \in U : \det \frac{\partial F}{\partial y}(x, g_{x_0}(x)) \neq 0\}.$$

One can show that $\hat{U}_{\max}$ is connected, i.e., it is not the union of two disjoint sets. Moreover, in other words, the first point where g fails to exist is where $\frac{\partial F}{\partial y}$ is not invertible anymore. This is somewhat not surprising since the fact that this matrix *is* invertible is the core assumption in IFT. In practice, the sets $\hat{U}_{\max}$ cannot be written down explicitly; in some very special cases, however, they can.

Let us determine the maximal existence intervals for the function g from Exercise 1. To this end, we need to figure out when $\partial_y F = 0$, i.e.,

$$2\cos y - 3xe^y = 0.$$

This yields

$$x = x(y) = \frac{2}{3}e^{-y}\cos y,$$

that is, depending on y, we can find the corresponding point x. Since pairs $(x(y), y)$ shall also be part of the mountain's path, we additionally need $F(x(y), y) = 0$ there; inserting the expression found for $x = x(y)$ into F therefore yields

$$0 = F(x(y), y) = 2\sin y - 3\Big(\frac{2}{3}e^{-y}\cos y\Big)e^y = 2(\sin y - \cos y),$$

thus we need to solve $\sin y = \cos y$, which holds for any y of the form

$$y = y_k = \frac{\pi}{4} + k\pi, \quad k \in \mathbb{Z}.$$

In turn, all points (x, y) where $\partial_y F(x, y) = 0$ are given by

$$x = x_k = \frac{2}{3}\cos\Big(\frac{\pi}{4} + k\pi\Big)e^{-(\frac{\pi}{4}+k\pi)} = (-1)^k\frac{\sqrt{2}}{3}e^{-\frac{\pi}{4}-k\pi}, \qquad y = y_k = \frac{\pi}{4} + k\pi, \qquad k \in \mathbb{Z}.$$

These points give rise to infinitely many intervals, which is a consequence of the periodicity of cos. Now we have $g(0) = \pi \in (y_k, y_{k+1})$ if and only if $k = 0$; thus, the interval $I_k = (x_k, x_{k+1})$ where $x_0 = 0 \in I_k$ is also given for $k = 0$, and the implicit function g has maximal interval of existence given by

$$\hat{U}_{\max} = \Big(-\frac{\sqrt{2}}{3}e^{-\frac{\pi}{4}-\pi}, \frac{\sqrt{2}}{3}e^{-\frac{\pi}{4}}\Big) \approx (-0.0093, 0.215).$$

Here we needed to switch x_k and x_{k+1} because one of them is negative, whereas the other is positive, and we don't want I_k to be empty. Note that this interval is somehow super short, but the best we can get. The full range of g, meaning $\hat{V}_{\max} = g[\hat{U}_{\max}]$, is then

$$\hat{V}_{\max} = (y_0, y_1) = \Big(\frac{\pi}{4}, \frac{\pi}{4} + \pi\Big) \approx (0.785, 3.927).$$

We could now search similarly for another function $\tilde{g}$ satisfying $\tilde{g}(0) = 0$ (i.e., we want to have another starting point on the mountain; I suggest to repeat the steps done above, i.e., to prove that $\tilde{g}$ exists and how the tangent looks like). Obviously the points (x_k, y_k) stay the same, and the value for k where $\tilde{g}(0) = 0 \in (y_k, y_{k+1})$ is now given by $k = -1$. The maximal existence interval and range for this function $\tilde{g}$ are then

$$\hat{U}_{\max} = \Big(-\frac{\sqrt{2}}{3}e^{-\frac{\pi}{4}+\pi}, \frac{\sqrt{2}}{3}e^{-\frac{\pi}{4}}\Big) \approx (-4.974, 0.215),$$
$$\hat{V}_{\max} = \Big(\frac{\pi}{4} - \pi, \frac{\pi}{4}\Big) \approx (-2.356, 0.785).$$

Note also that none of the above $\hat{U}_{\max}$ is symmetric, i.e., there is no $\delta > 0$ such that we can write $\hat{U}_{\max} = (x_0 - \delta, x_0 + \delta)$. However, IFT does not make a statement about the maximal size of the neighborhood, it just tells that there exists *some* neighborhood (i.e., *some* $\delta > 0$). Indeed, we can always find some (possibly very small) $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset \hat{U}_{\max}$, so the outcome of IFT is always fine.