Elements of Calculus | Session 20 (Week 10B)

Prof. Sören Petrat, Constructor University

Lecture notes from Spring 2025

6. Multivariable Calculus

6.1 Total and Partial Derivatives

Topic for Week 10 B: Chain Rule, Gradient, Higher-order Derivatives, Taylor Expansion

A few more remarks about derivatives.

Chain rule

For the total derivative of $f: \mathbb{R}^n \to \mathbb{R}^m$ a generalization of the chain rule holds:

Theorem: If $f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x_0 \in \mathbb{R}^n$ and $g: \mathbb{R}^m \to \mathbb{R}^j$ is differentiable at $f(x_0) \in \mathbb{R}^m$. Then $g \circ f$ (def. as $(g \circ f)(x) = g(f(x))$) is differentiable at x_0 with derivative $\underbrace{Dg\circ f\big|_{x_0}}_{j \times n \text{ matrix}} = \underbrace{Dg\big|_{f(x_0)}}_{j \times m \text{ matrix}} \underbrace{Df\big|_{x_0}}_{m \times n \text{ matrix}}$.

Example: $f: \mathbb{R} \to \mathbb{R}^2, t \mapsto f(t) = \begin{pmatrix} t^2 \\ t^3 \end{pmatrix}$

$g: \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto g(x_1, x_2) = x_1 e^{-x_2}$

$$\Rightarrow Dg\circ f\big|_t = Dg\big|_{f(t)} Df\big|_t = \left(\frac{\partial g}{\partial x_1} \;\; \frac{\partial g}{\partial x_2}\right)\Big|_{f(t)} \frac{\partial f}{\partial t} = \left(e^{-x_2}, -x_1 e^{-x_2}\right)\Big|_{f(t)} \begin{pmatrix} 2t \\ 3t^2 \end{pmatrix}$$

$$= \left(e^{-t^3}, -t^2 e^{-t^3}\right) \begin{pmatrix} 2t \\ 3t^2 \end{pmatrix} = 2t e^{-t^3} - 3t^4 e^{-t^3}.$$

In this simple example we have $g \circ f : \mathbb{R} \to \mathbb{R}$, $t \mapsto g(f(t)) = t^2 e^{-t^3}$, so we can verify with one-variable calculus:

$$\frac{d\, g(f(t))}{dt} = \frac{d}{dt}\left(t^2 e^{-t^3}\right) \underset{\text{product rule}}{=} 2te^{-t^3} + t^2(-3t^2)e^{-t^3} = 2te^{-t^3} - 3t^4 e^{-t^3} \quad \checkmark$$

Gradient

Let us consider the special case of $f : \mathbb{R}^n \to \mathbb{R}$ (real-valued functions).

Here, we call the total derivative the gradient of f or "nabla f", and we write

$$\boxed{Df|_x = \underset{\text{nabla/gradient}}{\nabla^T} f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}^T}$$

, i.e., ∇f is the vector of partial derivatives.

We need to use the transpose since for $f: \mathbb{R}^n \to \mathbb{R}$ the total derivative is a $1 \times n$ matrix, i.e., a row vector.

Note: Often we write $\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}$, which is a linear differential operator.

Let us note two interesting properties:

- Recall that $D_u f = Df\, u = \nabla f\, u = \underbrace{\langle \nabla f, u \rangle}_{\text{scalar/dot product}} = \|\nabla f\| \underbrace{\|u\|}_{=1} \cos\varphi$, with φ the angle between ∇f and u. Hence, if $\nabla f \neq 0$, then f has greatest directional derivative in direction $\frac{\nabla f(x)}{\|\nabla f(x)\|}$.

  In other words, ∇f points in the direction where f changes the most.

- If $\nabla f \neq 0$, then f increases in at least one direction and decreases in the opposite direction.

Hence, $\boxed{\text{if } f \text{ has a local extremum at } x \text{, then } \nabla f(x) = 0.}$

So similarly to one-variable Calculus, $\nabla f(x) = 0$ is a necessary condition for f to have a local maximum or minimum.

Higher-order Derivatives

Consider the example $f(x_1, x_2) = x_1^2 e^{-x_2}$

We find $\frac{\partial f}{\partial x_1} = 2x_1 e^{-x_2}$, $\frac{\partial f}{\partial x_2} = -x_1^2 e^{-x_2}$

We can also take more derivatives: $\frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}\left(2x_1 e^{-x_2}\right) = 2e^{-x_2}$

$$\frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_2}\left(2x_1 e^{-x_2}\right) = -2x_1 e^{-x_2}$$

mixed partial derivatives

$$\frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_1}\left(-x_1^2 e^{-x_2}\right) = -2x_1 e^{-x_2}$$

$$\frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2}\left(-x_1^2 e^{-x_2}\right) = x_1 e^{-x_2}$$

Here, we see that $\frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}$. This is true in general if all partial derivatives are continuous:

<u>Theorem</u> (Clairaut's thm., or Schwarz's thm.):
If $f: \mathbb{R}^n \to \mathbb{R}^m$ has continuous 2nd partial derivatives, then $\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}$ $\forall\, i, j$.

Note: • We usually write $\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$

• There are examples where $\frac{\partial^2 f}{\partial x_i \partial x_j}$ is not continuous, and $\frac{\partial^2 f}{\partial x_i \partial x_j} \neq \frac{\partial^2 f}{\partial x_j \partial x_i}$, e.g.,

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

when we consider the mixed partial derivatives at $(0,0)$.

Note: For $f: \mathbb{R}^n \to \mathbb{R}$, the matrix H with $(H_f(x))_{ij} := \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ is called $\boxed{\text{Hessian matrix of } f.}$

If Schwarz' theorem applies, H_f is symmetric i.e., $(H_f)_{ij} = (H_f)_{ji}$.

E.g., in the example above we found $H_f(x) = \begin{pmatrix} 2e^{-x_2} & -2x_1 e^{-x_2} \\ -2x_1 e^{-x_2} & x_1 e^{-x_2} \end{pmatrix}$.

Taylor Expansion

Similar to functions in $\mathbb{R}$, we can do a Taylor expansion. Let us write it down here up to second order (and for $f: \mathbb{R}^n \to \mathbb{R}$ only).

Theorem (Taylor, 2nd order): Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Let $x \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$ be such that $x + th \in U$ $\forall t \in [0,1]$. Then

$$f(x+h) = f(x) + Df|_x h + \frac{1}{2}\underbrace{\langle h, H_f(x) h\rangle}_{= h^T H_f(x) h} + r_x(h), \text{ with } \frac{\|r_x(h)\|}{\|h\|^2} \xrightarrow{h \to 0} 0.$$

Proof: Follows from applying 1-d Taylor to $g(t) := f(x+th)$. □

We will apply this next time to find an analog of the second derivative test for finding maxima or minima.