3 Reminders on Differential Calculus

3.1 Differentiability

Notation

In the following:

- $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ are normed vector spaces;
- $\Omega \subset E$ is an open subset of E;
- $f : \Omega \subset E \to F$ is a function;
- $\mathcal{L}_c(E, F)$ denotes the space of continuous linear maps from E to F, equipped with the operator norm

$$\|L\|_{\mathcal{L}_c(E,F)} := \sup_{\|h\|_E \leq 1} \|Lh\|_F.$$

Definition 3.1 (Directional derivative)**.** *We say that f is* directionally differentiable at $x \in \Omega$ in the direction $h \in E$ *if*

$$\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(x + th) - f(x)}{t}$$

exists in F. The limit is called the directional derivative of f in the direction h and is denoted by

$$Df(x; h) = \partial_h f(x).$$

Definition 3.2 (Fréchet differentiability)**.** *We say that f is* Fréchet differentiable at $x \in \Omega$ *if there exists $L \in \mathcal{L}_c(E, F)$ such that*

$$f(x + h) = f(x) + Lh + r(h), \tag{1}$$

where $r(h)$ denotes a remainder term satisfying

$$\lim_{h \to 0} \frac{\|r(h)\|_F}{\|h\|_E} = 0.$$

If f is differentiable at x, the map L in equation (1) is unique. We call this the derivative (or differential) of f at x and we denote it by $Df(x)$.

Remark 3.1. *In general, Fréchet differentiability depends on the norms on E and F. In finite-dimensional spaces all norms are equivalent, so the notion of Fréchet differentiability is independent of the particular choice of norms.*

Remark 3.2. *Fréchet differentiability at x implies continuity at x and directional differentiability in every direction $h \in E$, with*

$$Df(x; h) = Df(x)h.$$

The converse is false in general.

Remark 3.3. *Directional derivatives in all directions need not even imply continuity. They also need not depend linearly or continuously on the direction.*

Example 3.1. *Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by*

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

Then

$$f(x, x^2) = 1 \qquad (x \neq 0),$$

so f is not continuous at $(0,0)$. However, for every $h = (h_1, h_2) \in \mathbb{R}^2$,

$$f(th_1, th_2) = \begin{cases} t\dfrac{h_1^2}{h_2}, & h_2 \neq 0, \\ th_1, & h_2 = 0, \end{cases}$$

and hence all directional derivatives at $(0,0)$ exist:

$$Df((0,0); h) = \begin{cases} \dfrac{h_1^2}{h_2}, & h_2 \neq 0, \\ h_1, & h_2 = 0. \end{cases}$$

Thus directional differentiability in every direction does not imply continuity or Fréchet differentiability.

Proposition 3.1. *If f is Fréchet differentiable at $x \in \Omega$, then its derivative $Df(x) \in \mathcal{L}_c(E, F)$ is uniquely determined, and*

$$Df(x; h) = Df(x)h \qquad \forall h \in E.$$

Definition 3.3 (Gradient)**.** *Suppose that $F = \mathbb{R}$ and that E is a Hilbert space with inner product $\langle \cdot, \cdot \rangle_E$. If f is Fréchet differentiable at x, the* gradient *of f at x is the unique vector $\nabla f(x) \in E$ such that*

$$\langle \nabla f(x), h \rangle_E = Df(x)h \qquad \forall h \in E.$$

Remark 3.4. *The existence and uniqueness of $\nabla f(x)$ follow from the Riesz representation theorem.*

Remark 3.5. *If $E = \mathbb{R}^m$ with the Euclidean inner product, then*

$$\nabla f(x) = \big(\partial_1 f(x), \ldots, \partial_m f(x)\big).$$

More generally, if $A \in \mathbb{R}^{m \times m}$ is symmetric positive definite, then

$$\langle x, y \rangle_A := x^T A y$$

defines an inner product on $\mathbb{R}^m$. What is the gradient of f with respect to this inner product?

Definition 3.4. *We say that $f \in C^1(\Omega; F)$ if f is Fréchet differentiable at every point of Ω and*

$$Df : \Omega \to \mathcal{L}_c(E, F)$$

is continuous in the operator norm.

Theorem 3.1 (From directional to Fréchet differentiability)**.** *Suppose that all directional derivatives $Df(x; h)$ exist for $x \in \Omega$ and $h \in E$. Assume moreover that, for every $x \in \Omega$, the map*

$$h \longmapsto Df(x; h)$$

is linear and continuous, and that the map

$$x \in \Omega \mapsto \big(h \mapsto Df(x; h)\big) \in \mathcal{L}_c(E, F)$$

is continuous. Then $f \in C^1(\Omega; F)$ and

$$Df(x)(h) = Df(x; h) \qquad \forall x \in \Omega.$$

Theorem 3.2 (Chain rule)**.** *Let E, F, G be normed vector spaces, let $\Omega \subset E$ and $U \subset F$ be open, and let*

$$f : \Omega \to F, \qquad g : U \to G.$$

If $x \in \Omega$, $f(x) \in U$, f is Fréchet differentiable at x, and g is Fréchet differentiable at $f(x)$, then $g \circ f$ is Fréchet differentiable at x and

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

$$x \in \Omega \subset E \xrightarrow{\ f\ } f(x) \in U \subset F \xrightarrow{\ g\ } g(f(x)) \in G$$

Example 3.2. *Let $L \in C^2(\mathbb{R})$ and define*

$$J : C([a, b]) \to \mathbb{R}, \qquad J(u) = \int_a^b L(u(t))\, dt,$$

where $C([a, b])$ is equipped with the supremum norm

$$\|u\|_\infty := \max_{t \in [a,b]} |u(t)|.$$

Then J is Fréchet differentiable and

$$DJ(u)[h] = \int_a^b L'(u(t))h(t)\, dt.$$

Indeed, for fixed u,

$$|DJ(u)[h]| \le (b - a) \sup_{|s| \le \|u\|_\infty} |L'(s)|\, \|h\|_\infty,$$

so $DJ(u)$ is a bounded linear functional. Moreover, if $\|h\|_\infty \le 1$, Taylor's formula gives, for every $t \in [a,b]$,

$$L(u(t)+h(t)) = L(u(t)) + L'(u(t))h(t) + \frac{1}{2}L''(u(t)+\theta(t)h(t))h(t)^2$$

for some $\theta(t) \in [0,1]$. Hence

$$\big|J(u+h) - J(u) - DJ(u)[h]\big| \le \frac{b-a}{2} \sup_{|s|\le\|u\|_\infty+1} |L''(s)|\,\|h\|_\infty^2.$$

Therefore

$$J(u+h) = J(u) + DJ(u)[h] + r(h),$$

with r as in the definition above, so J is Fréchet differentiable.

Exercise 3.1. *Let $L \in C^1([a,b]\times\mathbb{R})$, and consider the map*

$$J : u \in C([a,b]) \to \int_a^b L(t,u(t))\,dt.$$

Is it differentiable?

3.2 Higher-Order Derivatives and Taylor Formulas

Definition 3.5 (Twice differentiable). *Let $\bar{x} \in \Omega$. We say that f is* twice Fréchet differentiable at $\bar{x}$ *if f is Fréchet differentiable on some open neighbourhood U of $\bar{x}$ and the map*

$$Df : U \to \mathcal{L}_c(E,F)$$

is Fréchet differentiable at $\bar{x}$. In this case, we define

$$D^2 f(\bar{x}) := D(Df)(\bar{x}) \in \mathcal{L}_c\big(E, \mathcal{L}_c(E,F)\big).$$

Remark 3.6. *There is a canonical identification*

$$\mathcal{L}_c\big(E, \mathcal{L}_c(E,F)\big) \cong \mathcal{L}_c(E\times E, F),$$

where the latter denotes the space of bounded bilinear maps. We write

$$D^2 f(x)[h_1,h_2] := \big[D^2 f(x) h_1\big] h_2.$$

Theorem 3.3 (Symmetry of the second derivative). *If f is twice differentiable at x, then*

$$D^2 f(x)[h_1,h_2] = D^2 f(x)[h_2,h_1] \qquad \forall x \in \Omega, \quad h_1,h_2 \in E.$$

Example 3.3 (The Hessian). *If $E = \mathbb{R}^m$ and $F = \mathbb{R}$, then $D^2 f(x)$ is identified with the Hessian matrix*

$$D^2 f(x) = \begin{pmatrix} \partial_{11} f(x) & \partial_{12} f(x) & \cdots & \partial_{1m} f(x) \\ \partial_{21} f(x) & \partial_{22} f(x) & \cdots & \partial_{2m} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{m1} f(x) & \partial_{m2} f(x) & \cdots & \partial_{mm} f(x) \end{pmatrix} \in \mathbb{R}^{m\times m}.$$

Under the preceding C^2 hypothesis, this matrix is symmetric.

Example 3.4. *Let $L \in C^2(\mathbb{R})$, and consider the functional J in Example 3.2. Then, one can show that J is twice differentiable on its domain and*

$$D^2J(u)[h,k] = \int_a^b L''(u(t))h(t)k(t)\,dt.$$

Definition 3.6. *We say that $f \in C^2(\Omega; F)$ if $f \in C^1(\Omega; F)$ and*

$$Df : \Omega \to \mathcal{L}_c(E, F)$$

is itself of class C^1. Equivalently,

$$D^2 f : \Omega \to \mathcal{L}_c(E \times E, F)$$

is continuous.

3.2.1 Taylor formulas with integral remainder

From now on we take for simplicity $F = \mathbb{R}^d$. Let $f : \Omega \to \mathbb{R}^d$, $x \in \Omega$, and let $h \in E$ such that the segment

$$\{x + th\ ;\ t \in [0,1]\}$$

belongs to Ω for every $t \in [0,1]$. Define

$$g(t) := f(x + th).$$

We consider two cases:

- If $f \in C^1(\Omega; \mathbb{R}^d)$, then $g \in C^1([0,1]; \mathbb{R}^d)$ and

  $$g'(t) = Df(x + th)h.$$

  The fundamental theorem of calculus gives

  $$g(1) = g(0) + \int_0^1 g'(t)\,dt,$$

  and therefore

  $$\boxed{f(x+h) = f(x) + \int_0^1 Df(x+th)h\,dt.}$$

  We refer to this expression as *Taylor formula of order 1 with integral remainder.*

- If $f \in C^2(\Omega; \mathbb{R}^d)$, then $g \in C^2([0,1]; \mathbb{R}^d)$

  $$g''(t) = D^2 f(x + th)[h,h],$$

  and applying the one-dimensional Taylor formula with integral remainder,

  $$g(1) = g(0) + g'(0) + \int_0^1 (1-t)g''(t)\,dt.$$

  This gives us the *Taylor formula of order 2 with integral remainder*:

  $$\boxed{f(x+h) = f(x) + Df(x)h + \int_0^1 (1-t)D^2 f(x+th)[h,h]\,dt.}$$

3.2.2 Taylor–Young formulas

Little-o notation

For functions f and g defined on a normed space (as $h \to 0$):

$$f(h) = o(g(h)) \qquad \Longleftrightarrow \qquad \lim_{h\to 0} \frac{\|f(h)\|}{\|g(h)\|} = 0.$$

The first-order Taylor–Young formula is just another way of expressing Fréchet differentiability at x:

$$\boxed{f(x+h) = f(x) + Df(x)h + o(h).}$$

Theorem 3.4 (Taylor–Young formula of order 2)**.** *If f is twice-differentiable at $x \in \Omega$, then*

$$\boxed{f(x+h) = f(x) + Df(x)h + \frac{1}{2}D^2 f(x)[h,h] + o(\|h\|_E^2).}$$

Proof sketch. Define

$$R(h) = f(x+h) - f(x) - Df(x)(h) - \frac{1}{2}D^2 f(x)(h,h).$$

Then

$$R(0) = 0,$$

and, since R is differentiable near 0,

$$DR(h) = Df(x+h) - Df(x) - D^2 f(x)(h,\cdot).$$

Because Df is differentiable at x with derivative $D^2 f(x)$,

$$\frac{\|DR(h)\|}{\|h\|} \to 0 \qquad \text{as } h \to 0.$$

Moreover,

$$R(h) = \int_0^1 DR(th)(h)\,dt,$$

so

$$\|R(h)\| \le \int_0^1 \|DR(th)\|\,\|h\|\,dt = o(\|h\|^2).$$

This proves the result. □