Differentiation in Banach Spaces.

PWP Interdysciplinary Doctoral Studies in Mathematical Modeling
UDA-POKL.04.01.01-00-026/13-00
Projekt jest współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

1 Introduction and notations

In this notes we will assume that the Reader is familiar with the basic notions related to normed spaces. We will usually denote the normed space with E or E_i, $i \in \{0, 1, 2, ...\}$ whenever we are referring to more then one space at a time. The norm of the space E will be denoted as $\| \cdot \|$, while the norm in the space E_i will be denoted as $\| \cdot \|_i$. In case of the inner product spaces we will denote the inner product by $\langle \cdot, \cdot \rangle$ (adding the appropriate subscript if necessary). In this case we will always refer to the inner product space as to the normed space with the norm naturally defined by:

$$\|x\| = \sqrt{\langle x, x \rangle},$$

for $x \in E$.

The space of linear and continuous maps between two normed spaces E_i and E_j will be denoted by $L(E_i, E_j)$. We will always assume that such space is naturally equipped with the norm

$$\|A\| = \sup_{\|x\|_i = 1} \|Ax\|_j,$$

where $A \in L(E_i, E_j)$.

The dual of the space E, i.e. $L(E, R)$, will be denoted as E^*. Dual pair $\langle \cdot, \cdot \rangle$ between spaces E^* and E is denoted as

$$\langle f, x \rangle = f(x)$$

for $f \in E^*, x \in E$.

We will refer to some specific examples of spaces, especially including:

- Euclidean spaces, $\mathbb{R}^k$, with the scalar product

  $$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + ... + x_k y_k$$

  for $x = (x_1, ..., x_k)$, $y = (y_1, ..., y_k)$;

- $C(\Omega)$ being the space of continuous functions $f : A \to \mathbb{R}$ defined on $\Omega \subset \mathbb{R}^k$, being the compact subset of the Euclidean space. This space will, by default, be equipped with the norm

  $$\|f\| = \sup_{x \in \Omega} |f(x)|,$$

  for $f \in C(\Omega)$. The most common example will be $C[a, b]$, for $\Omega = [a, b] \subset \mathbb{R}^1$.

- $L^2(\Omega)$ being the space containing the equivalence classes of measurable functions $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^k$ a Lebesque measurable set, and such that $\int_\Omega |f(x)|^2 dx < +\infty$, under the relation of being equal almost everywhere. This space is equipped with the scalar product

  $$\langle f, g \rangle = \int_\Omega f(x) g(x) dx,$$

  for $f, g \in L^2(\Omega)$.

2 Derivative of the map

Before we introduce the basic definitions let us remind the notion of the derivative of the map $F:\mathbb{R}^k\to\mathbb{R}^m$ between the two Euclidean spaces.

Definition 1. Let $U\subset\mathbb{R}^k$ be the open set and $x_0\in U$ be fixed. The linear map $A:\mathbb{R}^k\to\mathbb{R}^m$ is called a derivative of the map $F:U\to\mathbb{R}^m$ at x_0 if

$$\forall_{\varepsilon>0}\exists_{\delta>0}\forall_{x\in U}|x-x_0|\le\delta\Rightarrow|F(x)-F(x_0)-A(x-x_0)|\le\varepsilon|x-x_0|.$$

We know that the linear map $A=DF(x_0)$ is represented by the Jacobi matrix, i.e.

$$DF(x_0)=\left[\frac{\partial F_i(x_0)}{\partial x_j}\right]_{1\le i\le m,1\le j\le k},$$

where $F=(F_1,...,F_m)$.

The very natural generalization of this concept is given in the following definition:

Definition 2. Let $U\subset E_1$ be the open subset and $x_0\in U$. The map $F:E_1\to E_2$ is said to be Frechét differentiable at x_0 if there exists $A\in L(E_1,E_2)$ such that

$$\forall_{\varepsilon>0}\exists_{\delta>0}\forall_{x\in U}\|x-x_0\|_1\le\delta\Rightarrow\|F(x)-F(x_0)-A(x-x_0)\|_2\le\varepsilon\|x-x_0\|_1.$$

We will denote such linear map A as $DF(x_0)$ and call Frechét derivative of the map F at point x_0.

Alternatively we may write

$$\lim_{x\to x_0}\frac{\|F(x)-F(x_0)-A(x-x_0)\|_2}{\|x-x_0\|_1}=0 \tag{1}$$

or

$$\lim_{h\to 0}\frac{\|F(x_0+h)-F(x_0)-A(h)\|_2}{\|h\|_1}=0 \tag{2}$$

This concept is the natural generalization of the definition of the derivative in Euclidean spaces, but as we know there may be different perspectives used. For example we may try to generalize the concept of the partial or directional derivative. As we know, in case of differentiable map, all directional derivatives exist, but the opposite is not necessarily true. So there may appear functions having directional derivatives but not necessarily differentiable. In this case there is a chance to operate in the broader class of maps.

So now we will be looking at the differentiation from the different perspective. Let $F:E_1\to E_2$ be given. Let us fix $x,v\in E_1$ and assume that there exists the limit

$$A_x(v)=\lim_{t\to 0}\frac{F(x+tv)-F(x)}{t}.$$

Property 1. *Assume the limit $A_x(v)$ exists. Then for each $\alpha\in\mathbb{R}$ exists the limit $A_x(\alpha v)$ and the equality $A_x(\alpha v)=\alpha A_x(v)$ holds true.*

Definition 3. If the limit

$$A_x(v)=\lim_{t\to 0}\frac{F(x+tv)-F(x)}{t}$$

exists for each $v\in E_1$ and the map $v\mapsto A_x(v)$ is a continuous linear map, then we say the map $F:E_1\to E_2$ is Gâteaux differentiable at x. The map A_x is called Gâteaux derivative of F at x. We will denote it as $DF(x)$, as the previously defined Frechét derivative. As we will see the definitions are essentially different, but this should not lead to any misunderstanding.

Remark 1. If the map $A_x(v)$ is defined for each $v \in E_1$, then the map A_x is homogeneous but it may happen not to be additive.

Example 1. How is it possible that the map A is not additive? Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^2 + x_2^2} & \\ 0 & \text{for } (x_1, x_2) = (0, 0). \end{cases}$$

Let us now take any direction $v = (v_1, v_2) \in \mathbb{R}^2$ and look at the function $\varphi_v : \mathbb{R} \to \mathbb{R}$ given by $\varphi_v(t) = f(tv_1, tv_2)$. Then for $t \neq 0$ and $v \neq (0, 0)$ we can see that

$$\varphi_v(t) = t \frac{v_1^2 v_2}{v_1^2 + v_2^2},$$

what means that

$$A_{(0,0)}(v) = \frac{\partial f}{\partial v}(0, 0) = \frac{v_1^2 v_2}{v_1^2 + v_2^2}.$$

And this map is obviously not additive.

Remark 2. In the literature we may find different (and not necessarily equivalent!) versions of the definition of the Gâteaux derivative. Actually one may not require that the map A is continuous, or even linear. Here we will not follow these more general definitions but will focus on the definition that is easier to handle formally. So we can always assume that $Df(x) \in L(E_1, E_2)$ (no matter if $Df(x)$ denotes Gâteaux or Frechét derivative).

Theorem 1. *If the map $F : E_1 \to E_2$ is Frechét differentiable at x, then it is Gâteaux differentiable as well, and the two derivatives coincide.*

Proof. Assume that F is Frechét differentiable at x_0, so

$$\lim_{h \to 0} \frac{\|F(x_0 + h) - F(x_0) - Df(x_0)h\|}{\|h\|} = 0.$$

Let us now fix $v \in E_1$, $v \neq 0$ so we have

$$\lim_{t \to 0} \frac{\|F(x_0 + tv) - F(x_0) - Df(x_0)(tv)\|}{\|tv\|} = 0,$$

$$\lim_{t \to 0} \frac{\|F(x_0 + tv) - F(x_0) - Df(x_0)(tv)\|}{|t|} = 0,$$

$$\lim_{t \to 0} \frac{F(x_0 + tv) - F(x_0) - Df(x_0)(tv)}{t} = 0,$$

$$\lim_{t \to 0} \frac{F(x_0 + tv) - F(x_0)}{t} = Df(x_0)(v).$$

This shows that $Df(x_0)(v)$ is also the Gâteaux derivative at x_0. □

As we are going to observe in a moment the two definitions are not equivalent. This observation will justify calling the Frechét derivative a *strong derivative*, while Gâteaux derivative *a weak* one. But before we present the very simple example we should observe that Frechét derivative has some properties that are well-known from the calculus in Euclidean spaces.

Let us start with the well-known necessary condition for differentiability.

Theorem 2. *Assume $F : E_1 \to E_2$ is Frechét differentiable at x_0, then F is continuous at x_0.*

Proof. Let us observe that there exists such $\eta > 0$, that

$$\|F(x) - F(x_0) - DF(x_0)(x - x_0)\|_2 \leq \|x - x_0\|_1,$$

for $\|x - x_0\|_1 \leq \eta$. Then we may say

$$\|F(x) - F(x_0)\|_2 - \|DF(x_0)(x - x_0)\|_2 \leq \|F(x) - F(x_0) - DF(x_0)(x - x_0)\|_2 \leq \|x - x_0\|_1,$$

$$\|F(x) - F(x_0)\|_2 \leq (1 + \|DF(x_0)\|) \cdot \|x - x_0\|_1.$$

This implies that for any $\varepsilon > 0$ we may choose $\delta = \min\{\eta, \varepsilon/(1 + \|DF(x_0)\|)\} > 0$ and for $\|x - x_0\|_1 \leq \delta$ we have

$$\|F(x) - F(x_0)\|_2 \leq \varepsilon.$$

Hence F is continuous at x_0. □

Let us now look at the example showing that the definitions of Frechét and Gâteaux derivative are not equivalent.

Example 2. For the functional $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = \begin{cases} \frac{x_1^3 x_2}{x_1^6 + x_2^2} & \text{for } (x_1, x_2) \neq (0,0) \\ 0 & \text{for } (x_1, x_2) = (0,0), \end{cases}$$

we can see that

$$A(v) = \frac{\partial f}{\partial v}(0,0) = 0,$$

for each direction $v \in \mathbb{R}^2 \backslash \{(0,0)\}$, so it is Gâteaux differentiable at $(0,0)$. But the functional f is not continuous at $(0,0)$. This makes the example of the map that is Gâteaux differentiable, but not Frechét differentiable.

The above example not only shows that that Gâteaux differentiable map may not be Frechét differentiable, but it may also happen that it is not continuous!

Of course with some additional assumptions the Gâteaux (weak) differentiability is enough to have Frechét derivative. The appropriate condition closely resembles the sufficient condition for differentiability in Euclidean spaces.

Theorem 3. *Assume the map $F : E_1 \to E_2$ is Gâteaux differentiable in some open neighbourhood of $x_0 \in U \subset E_1$, and the map $x \mapsto Df(x) \in L(E_1, E_2)$ is continuous at x_0, then the map F is Frechét differentiable at x_0 as well.*

We will leave this theorem without the proof.

3 Basic applications: extrema

Now we will look at the special case of $E_2 = \mathbb{R}$, i.e. when E is the Banach space and $I : E \to \mathbb{R}$. Then if the Gâteaux derivative $DF(x)$ exists it is an element of the dual space E^*. In this very special case we will denote the Gâteaux derivative $DI(x) = \nabla I(x)$. Hence, for the Gâteaux differentiable functional I, we have $\nabla I : E \to E^*$. In the special case of E being the Hilbert space and due to the Riesz representation theorem we may associate E^* with E and think of $\nabla I : E \to E$.

For the map $I : E \to \mathbb{R}$ we may naturally define the local maximum and minimum of the map I

Definition 4. The point $x_0 \in E$ is the local minimum of the functional I if there exists such open neighbourhood $x_0 \in U \subset E$, that $I(x_0) \leq I(x)$ for $x \in U$.

Definition 5. The point $x_0 \in E$ is the local maximum of the functional I if there exists such open neighbourhood $x_0 \in U \subset E$, that $I(x_0) \geq I(x)$ for $x \in U$.

We say $x_0 \in E$ is an extremum of the map I if it is local minimum or local maximum.

The basic observation should be made that the necessary conditions of the extremum existence are similar to those known in the Euclidean spaces case.

Theorem 4. *Assume $x_0 \in E$ is the local extremum of the Gâteaux differentiable functional $I : E \to \mathbb{R}$, then $\nabla I(x_0) = 0$.*

Proof. Let us fix $v \in E$ and consider the map $\varphi_v : \mathbb{R} \to \mathbb{R}$ given by $\varphi_v(t) = I(x_0 + tv)$. Because I is Gâteaux differentiable, we can see that

$$\varphi_v'(0) = \nabla I(x_0).$$

If x_0 is an extremum of I, then 0 is the extremum of φ_v, so $\varphi'(0) = 0$, what completes the proof. □

We will be interested in minimizing the values of the functional $I : E \to \mathbb{R}$. Of course if the functional is bounded from below, i.e. there exists such $c \in \mathbb{R}$ that

$$\forall_{x \in E} I(x) \geq c$$

we can ask if there exists the global minimum of the functional i.e. $x^* \in E$ satisfying

$$f(x^*) = \inf_{x \in E} I(x).$$

Definition 6. We call the sequence $\{x_n\} \subset E$ the *minimizing sequence* of the functional I if

$$\lim_{n \to +\infty} I(x_n) = \inf_{x \in E} I(x)$$

One could expect that the existence of bounded or convergent minimizing sequence implies the existence of the global minimum of the functional I. Let us have a look at two examples.

Example 3. Let $I : C[0,1] \to \mathbb{R}$ be given by

$$I(u) = \int_0^1 (t + u(t))^2 dt.$$

The functional is bounded from below, we can also see that it is continuous and Gâteaux differentiable. Moreover the sequence $u_n(t) = t^n - t$ is the minimizing sequence, i.e.

$$I(u_n) = \int_0^1 t^{2n} dt = \frac{1}{2n+1} \to 0.$$

But if we look for such function $u^* \in C[0,1]$, that $f(u^*) = 0$ we can see that it must be $u^*(t) = -t$. But the sequence u_n does not converge to u^*.

Example 4. We can make the example given above even more evident if we consider the same functional on the space $C_0[0,1]$ of continuous functions satisfying boundary conditions $u(0 = u(1) = 0$. As we can see the sequence $u_n(t) = t^n - t$ belongs to $C_0[0,1]$, and it is again the minimizing sequence of the functional I but the functional I does not achieve its minimum in the space $C_0[0,1]$.

One may specify different sufficient conditions for the existence of the extremum, some of them similar to conditions known in Euclidean spaces, including these related to the second derivative. We are not going to talk about higher order derivatives but rather concentrate on the results related to convexity.

Definition 7. Let E be the Banach space and $U \subset E$ its convex subset. The functional $I : U \to \mathbb{R}$ is called convex if for each pair of points $x_1, x_2 \in U$ and any number $t \in [0,1]$ the inequality holds

$$I(tx_1 + (1-t)x_2) \leq tI(x_1) + (1-t)I(x_2).$$

Theorem 5. *Assume $U \subset E$ is the convex and open subset of the Banach space E. Let $I : U \to \mathbb{R}$ be the Gâteaux differentiable functional. Then the two conditions are equivalent*

(i) I is convex;

(ii) for each pair of points $x_1, x_2 \in U$ the inequality holds

$$I(x_2) - I(x_1) - \langle \nabla I(x_1), x_2 - x_1 \rangle \geq 0.$$

Proof. First we are going to prove that (i) implies (ii). Let us take any $x_1, x_2 \in U$ and define the map $\varphi : [0,1] \to \mathbb{R}$ by

$$\varphi(t) = I(x_1 + t(x_2 - x_1)).$$

As it may be observed the function φ is differentiable and we can find its derivative as

$$\varphi'(t) = \langle \nabla I(x_1 + t(x_2 - x_1)), x_2 - x_1 \rangle.$$

That is why by the mean value (Lagrange) theorem there exists such $\tau \in (0,1)$ that

$$I(x_2) - I(x_1) = \varphi(1) - \varphi(0) = \varphi'(\tau) \geq \varphi'(0),$$

as φ is a convex function. Hence

$$I(x_2) - I(x_1) \geq \langle \nabla I(x_1), x_2 - x_1 \rangle,$$

what completes the proof of the first implication.

Now let us assume (ii) and prove that I is convex. Let us have $x_1, x_2 \in U$ fixed and let $x_0 = (1-t)x_1 + tx_2$ for $t \in [0,1]$. Then we have

$$I(x_1) \geq I(x_0) + \langle \nabla I(x_0), x_1 - x_0 \rangle$$

$$I(x_2) \geq I(x_0) + \langle \nabla I(x_0), x_2 - x_0 \rangle$$

and

$$(1-t)I(x_1) + tI(x_2) \geq I(x_0) + \langle \nabla I(x_0), (1-t)x_1 + tx_2 - x_0 \rangle$$

$$(1-t)I(x_1) + tI(x_2) \geq I((1-t)x_1 + tx_2),$$

what proves (i). □

The convexity makes the necessary condition for the existence of the minimum a sufficient one as well – as the next theorem says.

Theorem 6. *Let $U \subset E$ be the open and convex subset of the Banach space E, and the functional $I : U \to \mathbb{R}$ be convex. Then $x_0 \in U$ is the minimum of I if and only if $\nabla I(x_0) = 0$.*

Proof. By Theorem 4 any local minimum $x_0 \in U$ satisfies $\nabla I(x_0) = 0$. Hence it is enough to show that each zero of the gradient map $\nabla I(x_0) = 0$ must be a local minimum.

Let us assume, contrary to our claim, that there exists $x \in U$ such that $I(x) < I(x_0)$. Let us take any $t \in (0,1)$ and have a look at

$$I(x_0 + t(x - x_0)) \leq (1-t)I(x_0) + tI(x) \leq I(x_0) + t(I(x) - I(x_0))$$

and

$$I(x_0 + t(x - x_0)) - I(x_0) \leq t(I(x) - I(x_0))$$

$$\frac{I(x_0 + t(x - x_0)) - I(x_0)}{t} \leq (I(x) - I(x_0)) < 0$$

and passing to the limit $t \to 0$ we have

$$\langle \nabla I(x_0), x - x_0 \rangle < 0,$$

what contradicts the assumption of $\nabla I(x_0) = 0$. □

References

[1] Jean Pierre Aubin, Applied functional analysis, John Wiley & Sons, 2000

[2] Atkinson, Han, Theoretical Numerical Analysis, A Functional Analysis Framework, Springer, 2009.