Workshop Lecture

for

Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators

Yang Jiayu

Xi'an Jiaotong University

November 18, 2023

1 Frechet and Gateaus derivatives

Throughout this section, $(\mathbb{X}, \|\cdot\|_1)$, $(\mathbb{Y}, \|\cdot\|_2)$ are two Banach space.

Definition 1 Let

$$f : U(\subset \mathbb{X}) \longrightarrow \mathbb{Y}$$

be a function defined on an open subset U of $\mathbb{X}$. We say that f is Gateaus differentiable at $x \in U$, if $\exists\, L \in \mathscr{L}(\mathbb{X}, \mathbb{Y})$ such that

$$\lim_{t \to 0} \frac{\|f(x+tv) - f(x) - L(tv)\|_2}{t} = 0, \qquad \forall\, v \in \mathbb{X}. \tag{1}$$

Denote L by $f'(x)$, called the Gateaus derivate of f at x.

Proposition 2 The Gateaus derivate is unique if it exists.

Proof. If $f'(x), \tilde{f}'(x)$ satisfy 1,

$$\|f'(x)v - \tilde{f}'(x)v\|_2 \le \frac{\|f(x+tv) - f(v) - tf'(x)v\|_2}{t} + \frac{\|f(x+tv) - f(v) - t\tilde{f}'(x)v\|_2}{t}$$

for small $t > 0$. □

Theorem 3 (Mean Value Theorem) Given $x, y \in \mathbb{X}$, assume that f has a Gateaus derivate at each point int the set

$$[x, y] = \{\lambda x + (1-\lambda)y : 0 \le \lambda \le 1\}.$$

Then, $\forall\, \ell \in \mathbb{Y}^*$, $\exists\, \xi \in (0,1)$ such that

$$\langle \ell, f(y) - f(x) \rangle = \langle \ell, f'(x + \xi(y-x))(y-x) \rangle.$$

Proof. Define

$$g(t) = \langle \ell, f(x + t(y-x)) \rangle$$

Then $g \in \mathcal{C}[0,1] \cap \mathcal{C}^1(0,1)$. By Lagrange Mean Value Theorem, $\exists\, \xi \in (0,1)$ such that

$$g'(\xi)(y-x) = g(y) - g(x).$$

But

$$g'(t) = \langle \ell, f'(t)(y-x) \rangle.$$

□

Definition 4 f is Frechet differentiable at $x \in U$, if $\exists\, f'(x) \in \mathscr{L}(\mathbb{X}, \mathbb{Y})$ such that

$$\lim_{\|v\|_1 \to 0} \frac{\|f(x+v) - f(x) - f'(x)v\|_2}{\|v\|_1} = 0, \qquad \forall\, v \in \mathbb{X}. \tag{2}$$

Proposition 5 If f is Frechet differentiable at x then it is continuous at x.

Proof. $\exists\, \delta > 0$,

$$\|v\|_1 < \delta \qquad \Longrightarrow \qquad \|f(x+v) - f(x) - f'(x)v\|_2 < \|v\|_1$$

It follows that

$$\|f(x+v) - f(x)\|_2 < \big(1 + \|f'(x)\|\big)\, \|v\|_1, \qquad \forall\, x+v \in B(x,\delta).$$

□

Remark 6 This property is not shared by the Gateaus derivate. For example, the function

$$f(x,y) = \begin{cases} \dfrac{x^3 y}{x^6 + y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = 0,0 \end{cases}$$

on $\mathbb{R}^2$ is discontinuous at $(0,0)$ since

$$\lim_{\substack{x \to 0 \\ y = kx^3}} \frac{x^3 y}{x^6 + y^2} = \frac{k}{1+k^2}.$$

But

$$f'(0,0) = \lim_{t \to 0} \frac{f(ut, vt)}{t} = \lim_{t \to 0} \frac{u^3 v t^4}{u^6 t^6 + v^2 t^2} = 0.$$

Lemma 7 $\mathbb{X}$ is a Banach space, $x \in \mathbb{X}$, then

$$\|x\|_{\mathbb{X}} = \sup_{\substack{\ell \in \mathbb{X}^* \\ \|\ell\|_{\mathbb{X}^*} \leq 1}} \langle \ell, x \rangle.$$

Proof. It is clear that

$$\langle \ell, x \rangle \leq \|\ell\|_{\mathbb{X}^*} \|x\|_{\mathbb{X}}, \qquad \forall\, \ell \in \mathbb{X}^*.$$

Thus

$$\|x\|_{\mathbb{X}} \geq \sup_{\substack{\ell \in \mathbb{X}^* \\ \|\ell\|_{\mathbb{X}^*} \leq 1}} \langle \ell, x \rangle.$$

To prove the converse inequality, we use the Hahn Banach theorem. Consider the subspace of $\mathbb{X}$

$$\mathbb{M} = \{kx : k \in \mathbb{K}\}.$$

and the bounded linear functional

$$f_0 : \mathbb{M} \longrightarrow \mathbb{R}$$
$$kx \mapsto k\|x\|$$

then $\|f_0\|_{\mathbb{X}^*} = 1$. By Hahn Banach theorem, $\exists\, f \in \mathbb{X}^*$ such that

$$\|f\|_{\mathbb{X}^*} = \|f_0\|_{\mathbb{X}^*} = 1, \qquad f|_{\mathbb{M}} = f_0.$$

thus

$$\|x\|_{\mathbb{X}} = \langle f, x\rangle \le \sup_{\substack{\ell\in\mathbb{X}^* \\ \|\ell\|_{\mathbb{X}^*}\le 1}} \langle \ell, x\rangle.$$

□

Theorem 8 Suppose f is Gateaus differentiable in an open subset U of $\mathbb{X}$. If f' is continuous at $x \in U$, then $f'(x)$ is the Frechet derivate of f at x.

Proof. Since $x \in U$ and $U \subset \mathbb{X}$ open, $\exists\, B(x,\delta) \subset U$. At the same time, since f' is continuous at x, $\forall\, \varepsilon > 0$, we could assume that

$$\|f'(x+v) - f'(x)\| < \varepsilon, \qquad \forall\, v \in B(0,\delta).$$

It follows that, given $v \in B(0,\delta)$

$$\|f(x+tv) - f'(x)\| < \varepsilon, \qquad \forall\, t \in (0,1).$$

By Mean Value Theorem, $\forall\, \ell \in \mathbb{Y}^*$, $\exists\, \xi \in (0,1)$ such that

$$\langle \ell, f(x+v) - f(x)\rangle = \langle \ell, f'(x+\xi v)v\rangle.$$

It follows that

$$\langle \ell, f(x+v) - f(x) - f'(x)v\rangle = \langle \ell, f'(x+\xi v)v - f'(x)v\rangle.$$

By lemma 7, $\forall\, v \in B(0,\delta)$

$$\begin{aligned}\|f(x+v) - f(x) - f'(x)v\|_2 &\le \|f'(x+\xi v)v - f'(x)v\|_2 \\ &\le \|f'(x+\xi v)v - f'(x)v\| \cdot \|v\|_1 < \varepsilon \cdot \|v\|_1\end{aligned}$$

□

Definition 9 If the first Gateaus derivate of f exists over some open subset of $\mathbb{X}$ that contains x and there is an element $f''(x) \in \mathscr{L}(\mathbb{X}, \mathscr{L}(\mathbb{X},\mathbb{Y}))$ that satisfies

$$\lim_{t\to 0} \frac{\|f'(x+tv) - f'(x) - tf''(x)v\|}{t} = 0, \qquad \forall\, v \in \mathbb{X}.$$

Then f'' is called the second Gateaus derivative of f at x.

If

$$\lim_{t\to 0} \frac{\|f'(x+v) - f(x) - f''(x)v\|_2}{\|v\|_1} = 0$$

$f''(x)$ is the second Frechet derivate of f at x.

Remark 10 Since $f''(x)$ is in $\mathscr{L}(\mathbb{X}, \mathscr{L}(\mathbb{X},\mathbb{Y}))$,

$$v_1 \in \mathbb{X} \qquad \Longrightarrow \qquad f''(x)v_1 \in \mathscr{L}(\mathbb{X},\mathbb{Y}).$$

$$v_1, v_2 \in \mathbb{X} \qquad \Longrightarrow \qquad \left(f''(x)v_1\right) v_2 := f''(x)v_1v_2 \in \mathbb{Y}.$$

That is, the mapping

$$h : \mathbb{X} \times \mathbb{X} \longrightarrow \mathbb{Y}$$
$$(v_1, v_2) \mapsto f''(x)v_1v_2$$

is a bilinear form.

Theorem 11 Let f be twice Gateaus differentiable at $x \in \mathbb{X}$. Set

$$g(x) = f(x + tv), \qquad h(x) = f'(x + tv_1)v_2.$$

Then

$$f'(x)v = g'(0), \qquad f''(x)v_1v_2 = h'(0).$$

Proof.
$$\frac{g(t) - g(0)}{t} = \frac{f(x + tv) - f(x)}{t} \qquad \frac{h(t) - h(0)}{t} = \frac{f'(x + tv_1)v_2 - f'(x)v_2}{t}$$

□

Example 12 Let $\mathbb{X}$ be a Hilbert space with inner product $(\cdot, \cdot)$, $\mathbb{Y} = \mathbb{R}$. Set

$$f(x) = (x, x), \qquad \forall\, x \in \mathbb{X}.$$

Then

$$g(t) = f(x + tv) = (x + tv, x + tv) = (x, x) + 2t(x, v) + t^2(v, v).$$

Thus

$$g'(0) = 2(x, v) = f'(x)v, \qquad f'(x) = 2(x, \cdot) \in \mathscr{L}(\mathbb{X}, \mathbb{R}).$$

Now

$$h(t) = f'(x + tv_1)v_2 = 2(x + tv_1, v_2)$$

It follows that

$$h'(0) = 2(v_1, v_2) = f''(x)v_1v_2, \qquad f''(x)(\cdot, \cdot) = 2(\cdot, \cdot).$$

Theorem 13 Suppose that f is Gateaus differentiable over $\mathbb{X}$. If f has a loacl maximum or minimum at $x \in \mathbb{X}$,

$$f'(x)v = 0, \qquad \forall\, v \in \mathbb{X}.$$

Proof. $\forall\, v \in \mathbb{X}$, let

$$g(t) = f(x + tv),$$

then g attains a local maximum or minimum. Therefore $g'(0) = 0$. □

2 Generalized Gram-Schmidt decompositions

Suppose $\{m_j\}_{j=1}^n$ is a collection of linearly independent vectors in a Hilbert space $\mathcal{H}$. Define

$$\mathbb{M}_j = \text{span}\{m_j\}, \qquad 1 \le j \le n.$$

$$\mathbb{S}_k = \sum_{j=1}^{k} \mathbb{M}_j = \text{span}\{m_j\}_{j=1}^k$$

As the m_j are linearly independent, $\mathbb{M}_j \cap \mathbb{M}_k = \{0\}$ and thus

$$\mathbb{M}_i \cap \left(\sum_{j \neq i} \mathbb{M}_j \right) = \{0\}.$$

Now, The Gram-Schmidt algorithm uses the m_j to create a new set of orthonormal vectors $\{e_j\}_{j=1}^n$ with $e_1 = m_1 / \|m_1\|$

$$e_k = \left(m_k - \sum_{j=1}^{k-1} (m_k, e_j) e_j \right) / \left\| m_k - \sum_{j=1}^{k-1} (m_k, e_j) e_j \right\|, \qquad 2 \leq k \leq n.$$

The method of construction ensures that

$$\operatorname{span}\{e_j\}_{j=1}^k = \operatorname{span}\{m_j\}_{j=1}^k.$$

Let

$$\mathbb{N}_j = \operatorname{span}\{e_j\}$$

Then

$$\mathbb{S}_k = \bigoplus_{j=1}^k \mathbb{N}_j.$$

In particular, we see from this that the $\mathbb{N}_k$ are characterized by

$$\begin{aligned} \mathbb{S}_k \cap \mathbb{S}_{k-1}^\perp = \left(\sum_{j=1}^k \mathbb{M}_j \right) \cap \left(\sum_{j=1}^{k-1} \mathbb{M}_j \right)^\perp &= \left(\operatorname{span}\{e_j\}_{j=1}^k \right) \cap \left(\operatorname{span}\{e_j\}_{j=1}^{k-1} \right)^\perp \\ = \operatorname{span}\{e_k\} = \mathbb{N}_k. \end{aligned}$$

Thus $\forall\ x \in \mathbb{M}_k$

$$x = \sum_{j=1}^k (x, e_j) e_j = \sum_{j=1}^k \mathscr{P}_{\mathbb{N}_j} x.$$

Let $\mathscr{P}_{\mathbb{N}_j|\mathbb{M}_k}$ be the projetcion operator $\mathscr{P}_{\mathbb{N}_j}$ restricted to $\mathbb{M}_k$. Then

$$\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k} m_k = (m_k, e_k) e_k.$$

and similarly

$$\mathscr{P}_{\mathbb{M}_k|\mathbb{N}_k} e_k = (m_k, e_k) m_k.$$

Therefore,

$$\left(\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k} \right)^{-1} e_k = \frac{m_k}{(m_k, e_k)}$$

and every $x \in \mathbb{M}_k$ can be expressed as

$$x = \sum_{j=1}^k \mathscr{P}_{\mathbb{N}_j} \left(\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k} \right)^{-1} z = \sum_{j=1}^k \mathscr{P}_{\mathbb{N}_j|\mathbb{M}_k} \left(\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k} \right)^{-1} z$$

for some $z \in \mathbb{N}_k$.

Now, consider a Hilbert space $\mathcal{H}$ that can be written as the algebraic direct sum of $n$closed subspace $\{\mathbb{M}_j\}_{j=1}^n$. That is

$$\mathcal{H} = \sum_{j=1}^n \mathbb{M}_j, \qquad \text{where } \mathbb{M}_i \cap \left(\sum_{j \neq i} \mathbb{M}_j \right) = \{0\}.$$

For $1 \leq k \leq n$, define the partial sums

$$\mathbb{S}_k = \sum_{i=1}^{k} \mathbb{M}_i$$

and set

$$\mathbb{N}_k = \mathbb{S}_k \cap \mathbb{S}_{k-1}^{\perp}.$$

where $\mathbb{S}_0 = \{0\}$. Then $\mathbb{N}_k \perp \mathbb{S}_{k-1}$ for all k and $\mathbb{N}_i \perp \mathbb{N}_j$ for $i \neq j$. One can show by introduction that

$$\sum_{i=1}^{k} \mathbb{M}_i = \bigoplus_{i=1}^{k} \mathbb{N}_i, \qquad \forall\, k.$$

In particular

$$\mathcal{H} = \bigoplus_{j=1}^{n} \mathbb{N}_j.$$

Let $\mathscr{P}_{\mathbb{N}_k}$ be the orthogonal projetcion operators onto $\mathbb{N}_k$ for $1 \leq k \leq n$ and for $1 \leq j \leq k \leq n$ define the restriction of $\mathscr{P}_{\mathbb{N}_j}$ to $\mathbb{M}_k$ by

$$\mathscr{P}_{\mathbb{N}_j|\mathbb{M}_k} = \mathscr{P}_{\mathbb{N}_j} x, \qquad \forall\, x \in \mathbb{M}_k.$$

Theorem 14 $\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k}$ is bijective.

Proof. $\forall\, x \in \mathbb{N}_k,\ \exists s_{k-1} \in \mathbb{S}_{k-1},\ x_k \in \mathbb{M}_k$ such that

$$x = s_{k-1} + x_k.$$

As $\mathbb{N}_k \perp \mathbb{S}_{k-1}$,

$$x = \mathscr{P}_{\mathbb{N}_k} x = \mathscr{P}_{\mathbb{N}_k} x_k.$$

Thus $\mathscr{P}_{\mathbb{N}_k}$ maps $\mathbb{M}_k$ onto $\mathbb{N}_k$. If $x \in \mathbb{M}_k$ satisfies $\mathscr{P}_{\mathbb{N}_k} x = 0$, it must be that $x \in \mathbb{M}_k \cap \mathbb{N}_k^{\perp}$. And

$$\mathbb{M}_k \cap \mathbb{N}_k^{\perp} = \mathbb{M}_k \cap \bigoplus_{i \leq k-1} \mathbb{N}_i = \mathbb{M}_k \cap \mathbb{S}_{k-1} = \{0\}.$$

$\square$

The inverse of $\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k}$ can be written as

$$\left(\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k}\right)^{-1} = \sum_{j=1}^{k} \mathscr{P}_{\mathbb{N}_j} \left(\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k}\right)^{-1}.$$

since $\mathcal{H} = \oplus \mathbb{N}_j$.

Thus $\forall\, x \in \mathbb{M}_k$, $\exists\, z \in \mathbb{N}_k$ such that

$$x = \sum_{j=1}^{k} \mathscr{P}_{\mathbb{N}_j} \left(\mathscr{P}_{\mathbb{N}_k|\mathbb{M}_k}\right)^{-1} z.$$