Fréchet derivatives and Gâteaux derivatives

Jordan Bell
jordan.bell@gmail.com
Department of Mathematics, University of Toronto
April 3, 2014

1 Introduction

In this note all vector spaces are real. If $X$ and $Y$ are normed spaces, we denote by $B(X,Y)$ the set of bounded linear maps $X \rightarrow Y$, and write $B(X) = B(X,X)$. $B(X,Y)$ is a normed space with the operator norm.

2 Remainders

If $X$ and $Y$ are normed spaces, let $o(X,Y)$ be the set of all maps $r : X \rightarrow Y$ for which there is some map $\alpha : X \rightarrow Y$ satisfying:

- $r(x) = \|x\| \alpha(x)$ for all $x \in X$,
- $\alpha(0) = 0$,
- $\alpha$ is continuous at 0.

Following Penot,\footnote{Jean-Paul Penot, Calculus Without Derivatives, p. 133, §2.4.} we call elements of $o(X,Y)$ remainders. It is immediate that $o(X,Y)$ is a vector space.

If $X$ and $Y$ are normed spaces, if $f : X \rightarrow Y$ is a function, and if $x_0 \in X$, we say that $f$ is stable at $x_0$ if there is some $\epsilon > 0$ and some $c > 0$ such that $\|x - x_0\| \leq \epsilon$ implies that $\|f(x - x_0)\| \leq c\|x - x_0\|$. If $T : X \rightarrow Y$ is a bounded linear map, then $\|Tx\| \leq \|T\|\|x\|$ for all $x \in X$, and thus a bounded linear map is stable at 0. The following lemma shows that the composition of a remainder with a function that is stable at 0 is a remainder.\footnote{Jean-Paul Penot, Calculus Without Derivatives, p. 134, Lemma 2.41.}

**Lemma 1.** Let $X, Y$ be normed spaces and let $r \in o(X,Y)$. If $W$ is a normed space and $f : W \rightarrow X$ is stable at 0, then $r \circ f \in o(W,Y)$. If $Z$ is a normed space and $g : Y \rightarrow Z$ is stable at 0, then $g \circ r \in o(X,Z)$. 

1
2
Proof. $r \in o(X,Y)$ means that there is some $\alpha : X \to Y$ satisfying $r(x) = ||x|| \alpha(x)$ for all $x \in X$, that takes the value 0 at 0, and that is continuous at 0. As $f$ is stable at 0, there is some $\epsilon > 0$ and some $c > 0$ for which $||w|| \leq \epsilon$ implies that $||f(w)|| \leq c ||w||$. Define $\beta : W \to Y$ by

$$
\beta(w) = \begin{cases} 
\frac{||f(w)||}{||w||} \alpha(f(w)) & w \neq 0 \\
0 & w = 0,
\end{cases}
$$

for which we have

$$(r \circ f)(w) = ||w|| \beta(w), \quad w \in W.$$  

If $||w|| \leq \epsilon$, then $||\beta(w)|| \leq c ||\alpha(f(w))||$. But $f(w) \to 0$ as $w \to 0$, and because $\alpha$ is continuous at 0 we get that $\alpha(f(w)) \to \alpha(0) = 0$ as $w \to 0$. So the above inequality gives us $\beta(w) \to 0$ as $w \to 0$. As $\beta(0) = 0$, the function $\beta : W \to Y$ is continuous at 0, and therefore $r \circ f$ is remainder.

As $g$ is stable at 0, there is some $\epsilon > 0$ and some $c > 0$ for which $||y|| \leq \epsilon$ implies that $||g(y)|| \leq c ||y||$. Define $\gamma : X \to Z$ by

$$
\gamma(x) = \begin{cases} 
g(\frac{||x||\alpha(x)}{||x||}) & x \neq 0 \\
0 & x = 0.
\end{cases}
$$

For all $x \in X$,

$$(g \circ r)(x) = g(\frac{||x||\alpha(x)}{||x||}) = ||x|| \gamma(x).$$

Since $\alpha(0) = 0$ and $\alpha$ is continuous at 0, there is some $\delta > 0$ such that $||x|| \leq \delta$ implies that $||\alpha(x)|| \leq \epsilon$. Therefore, if $||x|| \leq \delta \land 1$ then

$$
g(\frac{||x||\alpha(x)}{||x||}) \leq c ||x|| \alpha(x) \leq c ||x|| \epsilon,$$

and hence if $||x|| \leq \delta \land 1$ then $||\gamma(x)|| \leq c \epsilon$. This shows that $\gamma(x) \to 0$ as $x \to 0$, and since $\gamma(0) = 0$ the function $\gamma : X \to Z$ is continuous at 0, showing that $g \circ r$ is a remainder.$\blacksquare$

If $Y_1, \ldots, Y_n$ are normed spaces where $Y_k$ has norm $\| \cdot \|_k$, then $\| (y_1, \ldots, y_n) \| = \max_{1 \leq k \leq n} \| y_k \|_k$ is a norm on $\prod_{k=1}^n Y_k$, and one can prove that the topology induced by this norm is the product topology.

**Lemma 2.** If $X$ and $Y_1, \ldots, Y_n$ are normed spaces, then a function $r : X \to \prod_{k=1}^n Y_k$ is a remainder if and only if each of $r_k : X \to Y_k$ are remainders, $1 \leq k \leq n$, where $r(x) = (r_1(x), \ldots, r_n(x))$ for all $x \in X$.

**Proof.** Suppose that there is some function $\alpha : X \to \prod_{k=1}^n Y_k$ such that $r(x) = ||x|| \alpha(x)$ for all $x \in X$. With $\alpha(x) = (\alpha_1(x), \ldots, \alpha_n(x))$, we have

$$r_k(x) = ||x|| \alpha_k(x), \quad x \in X.$$  

Because $\alpha(x) \to 0$ as $x \to 0$, for each $k$ we have $\alpha_k(x) \to 0$ as $x \to 0$, which shows that $r_k$ is a remainder.
Suppose that each \( r_k \) is a remainder. Thus, for each \( k \) there is a function \( \alpha_k : X \to Y_k \) satisfying \( r_k(x) = \|x\| \alpha_k(x) \) for all \( x \in X \) and \( \alpha_k(x) \to 0 \) as \( x \to 0 \).

Then the function \( \alpha : X \to \prod_{k=1}^{n} Y_k \) defined by \( \alpha(x) = (\alpha_1(x), \ldots, \alpha_n(x)) \) satisfies \( r(x) = \|x\| \alpha(x) \). Because \( \alpha_k(x) \to 0 \) as \( x \to 0 \) for each of the finitely many \( k \), \( 1 \leq k \leq n \), we have \( \alpha(x) \to 0 \) as \( x \to 0 \).

\[ \square \]

3 Definition and uniqueness of Fréchet derivative

Suppose that \( X \) and \( Y \) are normed spaces, that \( U \) is an open subset of \( X \), and that \( x_0 \in U \). A function \( f : U \to Y \) is said to be Fréchet differentiable at \( x_0 \) if there is some \( L \in \mathcal{B}(X,Y) \) and some \( r \in o(X,Y) \) such that

\[
f(x) = f(x_0) + L(x-x_0) + r(x-x_0), \quad x \in U.
\]

Suppose there are bounded linear maps \( L_1, L_2 \) and remainders \( r_1, r_2 \) that satisfy the above. Writing \( r_1(x) = \|x\| \alpha_1(x) \) and \( r_2(x) = \|x\| \alpha_2(x) \) for all \( x \in X \), we have

\[
L_1(x-x_0) + \|x-x_0\| \alpha_1(x-x_0) = L_2(x-x_0) + \|x-x_0\| \alpha_2(x-x_0), \quad x \in U,
\]

i.e.,

\[
L_1(x-x_0) - L_2(x-x_0) = \|x-x_0\| (\alpha_2(x-x_0) - \alpha_1(x-x_0)), \quad x \in U.
\]

For \( x \in X \), there is some \( h > 0 \) such that for all \( |t| \leq h \) we have \( x_0 + tx \in U \), and then

\[
L_1(tx) - L_2(tx) = \|tx\| (\alpha_2(tx) - \alpha_1(tx)),
\]

hence, for \( 0 < |t| \leq h \),

\[
L_1(x) - L_2(x) = \|x\| (\alpha_2(x) - \alpha_1(x)).
\]

But \( \alpha_2(tx) - \alpha_1(tx) \to 0 \) as \( t \to 0 \), which implies that \( L_1(x) - L_2(x) = 0 \). As this is true for all \( x \in X \), we have \( L_1 = L_2 \) and then \( r_1 = r_2 \). If \( f \) is Fréchet differentiable at \( x_0 \), the bounded linear map \( L \) in (1) is called the Fréchet derivative of \( f \) at \( x_0 \), and we define \( Df(x_0) = L \). Thus,

\[
f(x) = f(x_0) + Df(x_0)(x-x_0) + r(x-x_0), \quad x \in U.
\]

If \( U_0 \) is the set of those points in \( U \) at which \( f \) is Fréchet differentiable, then \( Df : U_0 \to \mathcal{B}(X,Y) \).

Suppose that \( X \) and \( Y \) are normed spaces and that \( U \) is an open subset of \( X \). We denote by \( C^1(U,Y) \) the set of functions \( f : U \to Y \) that are Fréchet differentiable at each point in \( U \) and for which the function \( Df : U \to \mathcal{B}(X,Y) \) is continuous. We say that an element of \( C^1(U,Y) \) is continuously differentiable. We denote by \( C^2(U,Y) \) those elements \( f \) of \( C^1(U,Y) \) such that

\[
Df \in C^1(U, \mathcal{B}(X,Y));
\]
that is, $C^2(U,Y)$ are those $f \in C^1(U,Y)$ such that the function $Df : U \rightarrow \mathcal{B}(X,Y)$ is Fréchet differentiable at each point in $U$ and such that the function

$$D(Df) : U \rightarrow \mathcal{B}(X,\mathcal{B}(X,Y))$$

is continuous.$^3$

The following theorem characterizes continuously differentiable functions $\mathbb{R}^n \rightarrow \mathbb{R}^m$. $^4$

**Theorem 3.** Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is Fréchet differentiable at each point in $\mathbb{R}^n$, and write

$$f = (f_1, \ldots, f_m).$$

$f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ if and only if for each $1 \leq i \leq m$ and $1 \leq j \leq n$ the function

$$\frac{\partial f_i}{\partial x_j} : \mathbb{R}^n \rightarrow \mathbb{R}$$

is continuous.

## 4 Properties of the Fréchet derivative

If $f : X \rightarrow Y$ is Fréchet differentiable at $x_0$, then because a bounded linear map is continuous and in particular continuous at 0, and because a remainder is continuous at 0, we get that $f$ is continuous at $x_0$.

We now prove that Fréchet differentiation at a point is linear.

**Lemma 4** (Linearity). Let $X$ and $Y$ be normed spaces, let $U$ be an open subset of $X$ and let $x_0 \in U$. If $f_1, f_2 : U \rightarrow Y$ are both Fréchet differentiable at $x_0$ and if $\alpha \in \mathbb{R}$, then $\alpha f_1 + f_2$ is Fréchet differentiable at $x_0$ and

$$D(\alpha f_1 + f_2)(x_0) = \alpha Df_1(x_0) + Df_2(x_0).$$

**Proof.** There are remainders $r_1, r_2 \in o(X,Y)$ such that

$$f_1(x) = f_1(x_0) + Df_1(x_0)(x-x_0) + r_1(x-x_0), \quad x \in U,$n

and

$$f_2(x) = f_2(x_0) + Df_2(x_0)(x-x_0) + r_2(x-x_0), \quad x \in U.$$n

Then for all $x \in U$,

$$(\alpha f_1 + f_2)(x) - (\alpha f_1 + f_2)(x_0) = \alpha f_1(x) - \alpha f_1(x_0) + f_2(x) - f_2(x_0)$$

$$= \alpha Df_1(x_0)(x-x_0) + \alpha r_1(x-x_0) + Df_2(x_0)(x-x_0) + r_2(x-x_0)$$

$$= (\alpha Df_1(x_0) + Df_2(x_0))(x-x_0) + (\alpha r_1 + r_2)(x-x_0),$$

and $\alpha r_1 + r_2 \in o(X,Y)$.

$\square$


The following lemma gives an alternate characterization of a function being Fréchet differentiable at a point.\textsuperscript{5}

\textbf{Lemma 5.} Suppose that $X$ and $Y$ are normed space, that $U$ is an open subset of $X$, and that $x_0 \in U$. A function $f : U \to Y$ is Fréchet differentiable at $x_0$ if and only if there is some function $F : U \to \mathcal{B}(X,Y)$ that is continuous at $x_0$ and for which

$$f(x) - f(x_0) = F(x)(x - x_0), \quad x \in U.$$ 

\textit{Proof.} Suppose that there is a function $F : U \to \mathcal{B}(X,Y)$ that is continuous at $x_0$ and that satisfies $f(x) - f(x_0) = F(x)(x - x_0)$ for all $x \in U$. Then, for $x \in U$,

$$f(x) - f(x_0) = F(x)(x - x_0) - F(x_0)(x - x_0) + F(x_0)(x - x_0)$$

$$= F(x_0)(x - x_0) + r(x - x_0),$$

where $r : X \to Y$ is defined by

$$r(x) = \begin{cases} (F(x + x_0) - F(x_0))(x) & x + x_0 \in U \\ 0 & x + x_0 \notin U. \end{cases}$$

We further define

$$\alpha(x) = \begin{cases} \frac{(F(x + x_0) - F(x_0))(x)}{\|x\|} & x + x_0 \in U, x \neq 0 \\ 0 & x + x_0 \notin U \\ 0 & x = 0, \end{cases}$$

with which $r(x) = \|x\| \alpha(x)$ for all $x \in X$. To prove that $r$ is a remainder it suffices to prove that $\alpha(x) \to 0$ as $x \to 0$. Let $\epsilon > 0$. That $F : U \to \mathcal{B}(X,Y)$ is continuous at $x_0$ tells us that there is some $\delta > 0$ for which $\|x\| < \delta$ implies that $\|F(x + x_0) - F(x_0)\| < \epsilon$ and hence

$$\|(F(x + x_0) - F(x_0))(x)\| \leq \|F(x + x_0) - F(x_0)\| \|x\| < \epsilon \|x\|.$$ 

Therefore, if $\|x\| < \delta$ then $\|\alpha(x)\| < \epsilon$, which establishes that $r$ is a remainder and therefore that $f$ is Fréchet differentiable at $x_0$, with Fréchet derivative $Df(x_0) = F(x_0)$.

Suppose that $f$ is Fréchet differentiable at $x_0$: there is some $r \in o(X,Y)$ such that

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + r(x - x_0), \quad x \in U,$$

where $Df(x_0) \in \mathcal{B}(X,Y)$. As $r$ is a remainder, there is some $\alpha : X \to Y$ satisfying $r(x) = \|x\| \alpha(x)$ for all $x \in X$, and such that $\alpha(0) = 0$ and $\alpha(x) \to 0$ as $x \to 0$. For each $x \in X$, by the Hahn-Banach extension theorem\textsuperscript{6} there is some $\lambda_x \in X^*$ such that $\lambda_x x = \|x\|$ and $|\lambda_x v| \leq \|v\|$ for all $v \in X$. Thus,

$$r(x) = (\lambda_x x)\alpha(x), \quad x \in X.$$ 

\textsuperscript{5}Jean-Paul Penot, Calculus Without Derivatives, p. 136, Lemma 2.46.

\textsuperscript{6}Walter Rudin, Functional Analysis, second ed., p. 59, Corollary to Theorem 3.3.
Define $F : U \to \mathcal{B}(X,Y)$ by
\[ F(x) = Df(x_0) + (\lambda_{x-x_0})\alpha(x-x_0), \]
i.e. for $x \in U$ and $v \in X$,
\[ F(x)(v) = Df(x_0)(v) + (\lambda_{x-x_0})\alpha(x-x_0) \in Y. \]
Then for $x \in U$,
\[ r(x-x_0) = (\lambda_{x-x_0})\alpha(x-x_0) = F(x)(x-x_0) - Df(x_0)(x-x_0), \]
and hence
\[ f(x) = f(x_0) + F(x)(x-x_0), \quad x \in U. \]
To complete the proof it suffices to prove that $F$ is continuous at $x_0$. But both $\lambda_0 = 0$ and $\alpha(0) = 0$ so $F(x_0) = Df(x_0)$, and for $x \in U$ and $v \in X$,
\[ \|(F(x) - F(x_0))(v)\| = \|\lambda_{x-x_0}v\| \alpha(x-x_0)\|
\[ = \|\lambda_{x-x_0}v\| \|\alpha(x-x_0)\|
\[ \leq \|v\| \|\alpha(x-x_0)\|, \]
so $\|F(x) - F(x_0)\| \leq \|\alpha(x-x_0)\|$. From this and the fact that $\alpha(0) = 0$ and $\alpha(x) \to 0$ as $x \to 0$ we get that $F$ is continuous at $x_0$, completing the proof. \(\square\)

We now prove the chain rule for Fréchet derivatives.\(^7\)

**Theorem 6** (Chain rule). Suppose that $X,Y,Z$ are normed spaces and that $U$ and $V$ are open subsets of $X$ and $Y$ respectively. If $f : U \to Y$ satisfies $f(U) \subseteq V$ and is Fréchet differentiable at $x_0$ and if $g : V \to Z$ is Fréchet differentiable at $f(x_0)$, then $g \circ f : U \to Z$ is Fréchet differentiable at $x_0$, and its Fréchet derivative at $x_0$ is
\[ D(g \circ f)(x_0) = Dg(f(x_0)) \circ Df(x_0). \]

**Proof.** Write $y_0 = f(x_0)$, $L_1 = Df(x_0)$, and $L_2 = Dg(y_0)$. Because $f$ is Fréchet differentiable at $x_0$, there is some $r_1 \in o(X,Y)$ such that
\[ f(x) = f(x_0) + L_1(x-x_0) + r_1(x-x_0), \quad x \in U, \]
and because $g$ is Fréchet differentiable at $y_0$ there is some $r_2 \in o(Y,Z)$ such that
\[ g(y) = g(y_0) + L_2(y-y_0) + r_2(y-y_0), \quad y \in V. \]
For all $x \in U$ we have $f(x) \in V$, and using the above formulas,
\[ g(f(x)) = g(y_0) + L_2(f(x)-y_0) + r_2(f(x)-y_0) \]
\[ = g(y_0) + L_2(L_1(x-x_0) + r_1(x-x_0)) + r_2(L_1(x-x_0) + r_1(x-x_0)) \]
\[ = g(y_0) + L_2(L_1(x-x_0)) + L_2(r_1(x-x_0)) + r_2(L_1(x-x_0) + r_1(x-x_0)). \]
\(^7\)Jean-Paul Penot, *Calculus Without Derivatives*, p. 136, Theorem 2.47.
Define $r_3 : X \to Z$ by $r_3(x) = r_2(L_1 x + r_1(x))$, and fix any $c > \|L_1\|$. Writing $r_1(x) = \|x\| \alpha_1(x)$, the fact that $\alpha(0) = 0$ and that $\alpha$ is continuous at $0$ gives us that there is some $\delta > 0$ such that if $\|x\| < \delta$ then $\|\alpha(x)\| < c - \|L_1\|$, and hence if $\|x\| < \delta$ then $\|r_1(x)\| \leq (c - \|L_1\|) \|x\|$. Then, $\|x\| < \delta$ implies that
\[
\|L_1 x + r_1(x)\| \leq \|L_1 x\| + \|r_1(x)\| \leq \|L_1\| \|x\| + (c - \|L_1\|) \|x\| = c \|x\|.
\]
This shows that $x \mapsto L_1 x + r_1(x)$ is stable at $0$ and so by Lemma 1 that $r_3 \in o(X, Z)$. Then, $r : X \to Z$ defined by $r = L_1 \circ r_1 + r_3$ is a sum of two remainders and so is itself a remainder, and we have
\[
g \circ f(x) = g \circ f(x_0) + L_2 \circ L_1(x - x_0) + r(x - x_0), \quad x \in U.
\]
But $L_1 \in \mathcal{B}(X,Y)$ and $L_2 \in \mathcal{B}(Y,Z)$, so $L_2 \circ L_1 \in \mathcal{B}(X,Z)$. This shows that $g \circ f$ is Fréchet differentiable at $x_0$ and that its Fréchet derivative at $x_0$ is
\[
L_2 \circ L_1 = Dg(y_0) \circ Df(x_0) = Dg(f(x_0)) \circ Df(x_0).
\]
\begin{proof}
The following is the product rule for Fréchet derivatives. By $f_1 \cdot f_2$ we mean the function $x \mapsto f_1(x)f_2(x)$.
\end{proof}

**Theorem 7** (Product rule). Suppose that $X$ is a normed space, that $U$ is an open subset of $X$, that $f_1, f_2 : U \to \mathbb{R}$ are functions, and that $x_0 \in U$. If $f_1$ and $f_2$ are both Fréchet differentiable at $x_0$, then $f_1 \cdot f_2$ is Fréchet differentiable at $x_0$, and its Fréchet derivative at $x_0$ is
\[
D(f_1 \cdot f_2)(x_0) = f_2(x_0)Df_1(x_0) + f_1(x_0)Df_2(x_0).
\]

**Proof.** There are $r_1, r_2 \in o(X, \mathbb{R})$ with which
\[
f_1(x) = f_1(x_0) + Df_1(x_0)(x - x_0) + r_1(x - x_0), \quad x \in U
\]
and
\[
f_2(x) = f_2(x_0) + Df_2(x_0)(x - x_0) + r_2(x - x_0), \quad x \in U.
\]
Multiplying the above two formulas,
\[
f_1(x)f_2(x) = f_1(x_0)f_2(x_0) + f_2(x_0)f_1(x_0)(x - x_0) + f_1(x_0)Df_2(x_0)(x - x_0)
+ Df_1(x_0)(x - x_0)Df_2(x_0)(x - x_0) + r_1(x - x_0)r_2(x - x_0)
+ f_1(x_0)r_2(x - x_0) + r_2(x - x_0)Df_1(x_0)(x - x_0)
+ Df_1(x_0)(x - x_0)Df_2(x_0)(x - x_0).
\]
Define $r : X \to \mathbb{R}$ by
\[
r(x) = Df_1(x_0)x Df_2(x_0)x + r_1(x)r_2(x) + f_1(x_0)r_2(x) + r_2(x)Df_1(x_0)x
+ f_2(x_0)r_1(x) + r_1(x)Df_2(x_0)x.
\]
for which we have, for \( x \in U \),
\[
f_1(x)f_2(x) = f_1(x_0)f_2(x_0) + f_2(x_0)Df_1(x_0)(x-x_0)+f_1(x_0)Df_2(x_0)(x-x_0) + r(x-x_0).
\]

Therefore, to prove the claim it suffices to prove that \( r \in o(X, \mathbb{R}) \). Define \( \alpha : X \to \mathbb{R} \) by \( \alpha(0) = 0 \) and \( \alpha(x) = \frac{Df_1(x_0)xDf_2(x_0)x}{\|x\|} \) for \( x \neq 0 \). For \( x \neq 0 \),
\[
|\alpha(x)| = \frac{|Df_1(x_0)x||Df_2(x_0)x|}{\|x\|} \\
\leq \frac{\|Df_1(x_0)\|\|x\|\|Df_2(x_0)\|\|x\|}{\|x\|} \\
= \|Df_1(x_0)\|\|Df_2(x_0)\|\|x\|.
\]

Thus \( \alpha(x) \to 0 \) as \( x \to 0 \), showing that the first term in the expression for \( r \) belongs to \( o(X, \mathbb{R}) \). Likewise, each of the other five terms in the expression for \( r \) belongs to \( o(X, \mathbb{R}) \), and hence \( r \in o(X, \mathbb{R}) \), completing the proof.

5 Dual spaces

If \( X \) is a normed space, we denote by \( X^\ast \) the set of bounded linear maps \( X \to \mathbb{R} \), i.e. \( X^\ast = \mathcal{B}(X, \mathbb{R}) \). \( X^\ast \) is itself a normed space with the operator norm. If \( X \) is a normed space, the dual pairing \( \langle \cdot, \cdot \rangle : X \times X^\ast \to \mathbb{R} \) is
\[
\langle x, \psi \rangle = \psi(x), \quad x \in X, \psi \in X^\ast.
\]

If \( U \) is an open subset of \( X \) and if a function \( f : U \to \mathbb{R} \) is Fréchet differentiable at \( x_0 \in U \), then \( Df(x_0) \) is a bounded linear map \( X \to \mathbb{R} \), and so belongs to \( X^\ast \). If \( U_0 \) are those points in \( U \) at which \( f : U \to \mathbb{R} \) is Fréchet differentiable, then
\[
Df : U_0 \to X^\ast.
\]

In the case that \( X \) is a Hilbert space with inner product \( \langle \cdot, \cdot \rangle \), the Riesz representation theorem shows that \( R : X \to X^\ast \) defined by \( R(x)(y) = \langle y, x \rangle \) is an isometric isomorphism. If \( f : U \to \mathbb{R} \) is Fréchet differentiable at \( x_0 \in U \), then we define
\[
\nabla f(x_0) = R^{-1}(Df(x_0)),
\]
and call \( \nabla f(x_0) \in X \) the gradient of \( f \) at \( x_0 \). With \( U_0 \) denoting the set of those points in \( U \) at which \( f \) is Fréchet differentiable,
\[
\nabla f : U_0 \to X.
\]
(To define the gradient we merely used that \( R \) is a bijection, but to prove properties of the gradient one uses that \( R \) is an isometric isomorphism.)

Example. Let \( X \) be a Hilbert space, \( A \in \mathcal{B}(X) \), \( v \in X \), and define
\[
f(x) = \langle Ax, x \rangle - \langle x, v \rangle, \quad x \in X.
\]

8
For all \( x_0, x \in X \) we have, because the inner product of a real Hilbert space is symmetric,
\[
f(x) - f(x_0) = \langle Ax, x \rangle - \langle x, v \rangle - \langle Ax_0, x_0 \rangle + \langle x_0, v \rangle \\
= \langle Ax, x \rangle - \langle Ax_0, x \rangle + \langle Ax_0, x_0 \rangle - \langle x, x_0 \rangle \\
= \langle A(x - x_0), x \rangle + \langle Ax_0, x - x_0 \rangle - \langle x, x_0 \rangle \\
= \langle x - x_0, A^* x \rangle + \langle x - x_0, Ax_0 \rangle - \langle x, x_0 \rangle \\
= \langle x - x_0, A^* x + Ax_0 - v \rangle \\
= \langle x - x_0, A^* x - A^* x_0 + A^* x_0 + Ax_0 - v \rangle \\
= \langle x - x_0, (A^* + A)x_0 - v \rangle + \langle x - x_0, A^*(x - x_0) \rangle.
\]
With \( Df(x_0)(x-x_0) = \langle x - x_0, (A^* + A)x_0 - v \rangle \) or \( Df(x_0)(x) = \langle x, (A^* + A)x_0 - v \rangle \), we have that \( f \) is Fréchet differentiable at each \( x_0 \in X \). Furthermore, its gradient at \( x_0 \) is
\[
\nabla f(x_0) = (A^* + A)x_0 - v.
\]
For each \( x_0 \in X \), the function \( f : X \to \mathbb{R} \) is Fréchet differentiable at \( x_0 \), and thus
\[
Df : X \to X^*,
\]
and we can ask at what points \( Df \) has a Fréchet derivative. For \( x_0, x, y \in X \),
\[
(Df(x) - Df(x_0))(y) = \langle y, (A^* + A)x - v \rangle - \langle y, (A^* + A)x_0 - v \rangle \\
= \langle y, (A^* + A)(x - x_0) \rangle.
\]
For \( D(Df)(x_0)(x - x_0)(y) = \langle y, (A^* + A)(x - x_0) \rangle \), in other words with
\[
D^2 f(x_0)(x)(y) = D(Df)(x_0)(x)(y) = \langle y, (A^* + A)x \rangle,
\]
we have that \( Df \) is Fréchet differentiable at each \( x_0 \in X \). Thus
\[
D^2 f : X \to \mathcal{B}(X, X^*).
\]
Because \( D^2 f(x_0) \) does not depend on \( x_0 \), it is Fréchet differentiable at each point in \( X \), with \( D^3 f(x_0) = 0 \) for all \( x_0 \in X \). Here \( D^3 f : X \to \mathcal{B}(X, \mathcal{B}(X, X^*)) \).

### 6 Gâteaux derivatives

Let \( X \) and \( Y \) be normed spaces, let \( U \) be an open subset of \( X \), let \( f : U \to Y \) be a function, and let \( x_0 \in U \). If there is some \( T \in \mathcal{B}(X, Y) \) such that for all \( v \in X \) we have
\[
\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} = Tv,
\]
then we say that \( f \) is Gâteaux differentiable at \( x_0 \) and call \( T \) the Gâteaux derivative of \( f \) at \( x_0 \).\(^8\) It is apparent that there is at most one \( T \in \mathcal{B}(X, Y) \) that

\(^8\)Our definition of the Gâteaux derivative follows Jean-Paul Penot, *Calculus Without Derivatives*, p. 127, Definition 2.23.
satisfies (2) for all \( v \in X \). We write \( f'(x_0) = T \). Thus, \( f' \) is a map from the set of points in \( U \) at which \( f \) is Gâteaux differentiable to \( \mathcal{B}(X,Y) \). If \( V \subseteq U \) and \( f \) is Gâteaux differentiable at each element of \( V \), we say that \( f \) is Gâteaux differentiable on \( V \).

**Example.** Define \( f : \mathbb{R}^2 \to \mathbb{R} \) by \( f(x_1,x_2) = \frac{x_1^6}{x_1^6+x_2^6} \) for \( (x_1,x_2) \neq (0,0) \) and \( f(0,0) = 0 \). For \( v = (v_1,v_2) \in \mathbb{R}^2 \) and \( t \neq 0 \),

\[
\frac{f(0+tv) - f(0)}{t} = \frac{f(tv_1,tv_2)}{t} = \begin{cases} \frac{1}{t} \cdot \frac{tv_1^6v_2}{tv_1^6+tv_2^6} & v \neq (0,0) \\ 0 & v = (0,0) \end{cases}
\]

Hence, for any \( v \in \mathbb{R}^2 \), we have \( \frac{f(0+tv)-f(0)}{t} \to 0 \) as \( t \to 0 \). Therefore, \( f \)

is Gâteaux differentiable at \( (0,0) \) and \( f'(0,0)v = 0 \in \mathbb{R} \) for all \( v \in \mathbb{R}^2 \), i.e. \( f'(0,0) = 0 \). However, for \( (x_1,x_2) \neq (0,0) \),

\[
f(x_1,x_2^2) = \frac{x_1^6}{x_1^6+x_2^6} = \frac{1}{2},
\]

from which it follows that \( f \) is not continuous at \( (0,0) \). We stated in §4 that if a function is Fréchet differentiable at a point then it is continuous at that point, and so \( f \) is not Fréchet differentiable at \( (0,0) \). Thus, a function that is Gâteaux differentiable at a point need not be Fréchet differentiable at that point.

We prove that being Fréchet differentiable at a point implies being Gâteaux differentiable at the point, and that in this case the Gâteaux derivative is equal to the Fréchet derivative.

**Theorem 8.** Suppose that \( X \) and \( Y \) are normed spaces, that \( U \) is an open subset of \( X \), that \( f \in Y^U \), and that \( x_0 \in U \). If \( f \) is Fréchet differentiable at \( x_0 \), then \( f \) is Gâteaux differentiable at \( x_0 \) and \( f'(x_0) = Df(x_0) \).

**Proof.** Because \( f \) is Fréchet differentiable at \( x_0 \), there is some \( r \in o(X,Y) \) for which

\[
f(x) = f(x_0) + Df(x_0)(x-x_0) + r(x-x_0), \quad x \in U.
\]

For \( v \in X \) and nonzero \( t \) small enough that \( x_0 + tv \in U \),

\[
\frac{f(x_0 + tv) - f(x_0)}{t} = \frac{Df(x_0)(x_0 + tv - x_0) + r(x_0 + tv - x_0)}{t} = \frac{tDf(x_0)v + r(tv)}{t}.
\]

Writing \( r(x) = \|x\| \alpha(x) \),

\[
\frac{f(x_0 + tv) - f(x_0)}{t} = \frac{tDf(x_0) + \|tv\| \alpha(tv)}{t} = Df(x_0)v + \|v\| \alpha(tv).
\]

Hence,

\[
\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} = Df(x_0)v.
\]

This holds for all \( v \in X \), and as \( Df(x_0) \in \mathcal{B}(X,Y) \) we get that \( f \) is Gâteaux differentiable at \( x_0 \) and that \( f'(x_0) = Df(x_0) \). \( \square \)
If $X$ is a vector space and $u, v \in X$, let
\[ [u, v] = \{(1 - t)u + tv : 0 \leq t \leq 1\}, \]
namely, the line segment joining $u$ and $v$. The following is a mean value theorem for Gâteaux derivatives.\(^9\)

**Theorem 9** (Mean value theorem). Let $X$ and $Y$ be normed spaces, let $U$ be an open subset of $X$, and let $f : U \to Y$ be Gâteaux differentiable on $U$. If $u, v \in U$ and $[u, v] \subset U$, then
\[ \|f(u) - f(v)\| \leq \sup_{w \in [u, v]} \|f'(w)\| \cdot \|u - v\|. \]

**Proof.** If $f(u) = f(v)$ then immediately the claim is true. Otherwise, $f(v) - f(u) \neq 0$, and so by the Hahn-Banach extension theorem\(^10\) there is some $\psi \in Y^*$ satisfying $\psi(f(v) - f(u)) = \|f(v) - f(u)\|$ and $\|\psi\| = 1$. Define $h : [0, 1] \to \mathbb{R}$ by
\[ h(t) = (f((1 - t)u + tv), \psi). \]
For $0 < t < 1$ and $\tau \neq 0$ satisfying $t + \tau \in [0, 1]$, we have
\[ \frac{h(t + \tau) - h(t)}{\tau} = \frac{1}{\tau} \langle f((1 - t - \tau)u + (t + \tau)v), \psi \rangle - \frac{1}{\tau} \langle f((1 - t)u + tv), \psi \rangle = \langle \frac{f((1 - t)u + tv + (v - u)\tau) - f((1 - t)u + tv)}{\tau}, \psi \rangle. \]
Because $f$ is Gâteaux differentiable at $(1 - t)u + tv$,
\[ \lim_{\tau \to 0} \frac{f((1 - t)u + tv + (v - u)\tau) - f((1 - t)u + tv)}{\tau} = f'((1 - t)u + tv)(v - u), \]
so because $\psi$ is continuous,
\[ \lim_{\tau \to 0} \frac{h(t + \tau) - h(t)}{\tau} = \langle f'((1 - t)u + tv)(v - u), \psi \rangle, \]
which shows that $h$ is differentiable at $t$ and that
\[ h'(t) = \langle f'((1 - t)u + tv)(v - u), \psi \rangle. \]

$h : [0, 1] \to \mathbb{R}$ is a composition of continuous functions so it is continuous. Applying the mean value theorem, there is some $\theta$, $0 < \theta < 1$, for which
\[ h'(\theta) = h(1) - h(0). \]


On the one hand,
\[ h'(\theta) = \langle f'((1 - \theta)u + \theta v)(v - u), \psi \rangle. \]

On the other hand,
\[ h(1) - h(0) = \langle f(v), \psi \rangle - \langle f(u), \psi \rangle = \langle f(v) - f(u), \psi \rangle = \|f(v) - f(u)\|. \]

Therefore
\[ \|f(v) - f(u)\| = \left| \langle f'((1 - \theta)u + \theta v)(v - u), \psi \rangle \right| \]
\[ \leq \|\psi\| \|f'((1 - \theta)u + \theta v)(v - u)\| \]
\[ = \|f'((1 - \theta)u + \theta v)(v - u)\| \]
\[ \leq \|f'((1 - \theta)u + \theta v)\| \|v - u\| \]
\[ \leq \sup_{w \in [u,v]} \|f'(w)\| \|v - u\|. \]

\[ \square \]

### 7 Antiderivatives

Suppose that \( X \) is a Banach space and that \( f : [a,b] \to X \) be continuous. Define \( F : [a,b] \to X \) by
\[ F(x) = \int_a^x f. \]

Let \( x_0 \in (a,b) \). For \( x \in (a,b) \), we have
\[ F(x) - F(x_0) = \int_a^x f - \int_a^{x_0} f = \int_{x_0}^x f = f(x_0)(x - x_0) + \int_{x_0}^x (f - f(x_0)), \]
from which it follows that \( F \) is Fréchet differentiable at \( x_0 \), and that
\[ DF(x_0)(x - x_0) = f(x_0)(x - x_0). \]

If we identify \( f(x_0) \in X \) with the map \( x \mapsto f(x_0)x \), namely if we say that \( X = \mathcal{B}(\mathbb{R}, X) \), then \( DF(x_0) = f(x_0) \).

Let \( X \) be a normed space, let \( Y \) be a Banach space, let \( U \) be an open subset of \( X \), and let \( f \in C^1(U,Y) \). Suppose that \( u, v \in U \) satisfy \([u,v] \subset U \). Write \( I = (0,1) \) and define \( \gamma : I \to U \) by \( \gamma(t) = (1 - t)u + tv \). We have
\[ D\gamma(t) = v - u, \quad t \in I, \]
and thus by Theorem 6,
\[ D(f \circ \gamma)(t) = Df(\gamma(t)) \circ D\gamma(t), \quad t \in I, \]
that is,
\[ D(f \circ \gamma)(t) = Df(\gamma(t)) \circ (v - u), \quad t \in I, \]
\[ D(f \circ \gamma)(t) = Df(\gamma(t))(v-u), \quad t \in I. \]

If \( t \in I \) and \( t + h \in I \), then

\[
D(f \circ \gamma)(t + h) - D(f \circ \gamma)(t) = Df(\gamma(t + h))(v-u) - Df(\gamma(t))(v-u) \\
= (Df(\gamma(t + h)) - Df(\gamma(t)))(v-u),
\]

and hence

\[
\|D(f \circ \gamma)(t + h) - D(f \circ \gamma)(t)\| \leq \|Df(\gamma(t + h)) - Df(\gamma(t))\| \|v-u\|.
\]

Because \( Df : U \to \mathcal{B}(X,Y) \) is continuous, it follows that

\[
\|D(f \circ \gamma)(t + h) - D(f \circ \gamma)(t)\| \to 0
\]
as \( h \to 0 \), i.e. that \( D(f \circ \gamma) \) is continuous at \( t \), and thus that

\[ D(f \circ \gamma) : I \to \mathcal{B}(\mathbb{R}, Y) \]
is continuous. If we identify \( \mathcal{B}(\mathbb{R}, Y) \) with \( Y \), then

\[ D(f \circ \gamma) : I \to Y. \]

On the one hand,

\[
\int_0^1 D(f \circ \gamma) = (f \circ \gamma)(1) - (f \circ \gamma)(0) = f(v) - f(u).
\]

On the other hand,

\[
\int_0^1 D(f \circ \gamma) = \int_0^1 Df(\gamma(t))(v-u)dt = \left(\int_0^1 Df((1-t)u + tv)dt\right)(v-u);
\]

here,

\[
\int_0^1 Df((1-t)u + tv)dt \in \mathcal{B}(X,Y).
\]

Therefore

\[
f(v) - f(u) = \left(\int_0^1 Df((1-t)u + tv)dt\right)(v-u).
\]