Deutsche Forschungsgemeinschaft
Priority Programme 1962

A Simple Proof of the Baillon–Haddad Theorem on Open Subsets of Hilbert Spaces

Daniel Wachsmuth, Gerd Wachsmuth

SPP 1962
Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-189

received on April 1, 2022

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

A simple proof of the Baillon–Haddad theorem on open subsets of Hilbert spaces*

Daniel Wachsmuth† Gerd Wachsmuth‡

April 1, 2022

We give a simple proof of the Baillon–Haddad theorem for convex functions defined on open and convex subsets of Hilbert spaces. We also state some generalizations and limitations. In particular, we discuss equivalent characterizations of the Lipschitz continuity of the derivative of convex functions on open and convex subsets of Banach spaces.

Keywords: Baillon–Haddad theorem, cocoercivity, strong smoothness
MSC: 26B25, 47H05, 47N10, 49J50

1 Introduction

A very important result in convex analysis is the Baillon–Haddad theorem which states that the derivative f' is $\frac{1}{L}$-cocoercive whenever $f\colon X \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous derivative, see [Baillon and Haddad, 1977, Corollaire 10]. Here, X is a (real) Banach space. In [Pérez-Aros and Vilches, 2019, Theorem 3.1] it was shown that this remains true if f is defined on an open and convex subset of a (real) Hilbert space. The corresponding proof is quite involved, since it uses generalized second-order derivatives and a reduction to the finite-dimensional situation. We also refer to [Bauschke and Combettes, 2010] for further comments and references. We give a short and direct proof, see Section 2.

It is well known that the L-Lipschitz continuity of the derivative of f is equivalent to a number of important properties of f and its convex conjugate. In Section 3, we investigate which of these equivalences remain valid if f is defined on an open and convex subset of a (real) Banach or Hilbert space.

*This research was supported by the German Research Foundation (DFG) under grant numbers WA 3626/3-2 and WA 3636/4-2 within the priority program "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization" (SPP 1962).

†Institut für Mathematik, Universität Würzburg, 97074 Würzburg, Germany, daniel.wachsmuth@mathematik.uni-wuerzburg.de

‡Brandenburgische Technische Universität Cottbus-Senftenberg, Institute of Mathematics, 03046 Cottbus, Germany, https://www.b-tu.de/fg-optimale-steuerung, gerd.wachsmuth@b-tu.de,

2 The Baillon–Haddad theorem on open subsets of Hilbert spaces

We start by a characterization of differentiable functions with Lipschitz derivatives.

Lemma 2.1. *Let $O \subset H$ be an open and convex subset of the Hilbert space H. Suppose that $f\colon O \to \mathbb{R}$ is (Gâteaux) differentiable. For $L \geq 0$, the following are equivalent.*

(i) *The derivative $f'\colon O \to H^\star$ is L-Lipschitz, i.e.,*

$$\|f'(y) - f'(x)\|_{H^\star} \leq L\|y - x\|_H \qquad \forall x, y \in O. \tag{2.1}$$

(ii) *Both f' and $-f'$ satisfy a one-sided Lipschitz estimate with constant L, i.e.,*

$$|\langle f'(y) - f'(x), y - x\rangle_H| \leq L\|y - x\|_H^2 \qquad \forall x, y \in O. \tag{2.2}$$

(iii) *The function f has a first order Taylor expansion with remainder $\frac{L}{2}\|\cdot\|_H^2$, i.e.,*

$$\left|f(y) - f(x) - \langle f'(x), y - x\rangle_H\right| \leq \frac{L}{2}\|y - x\|_H^2 \qquad \forall x, y \in O. \tag{2.3}$$

Proof. The implication "(i)⇒(ii)" is a simple application of the Cauchy–Schwarz inequality.

To prove "(ii)⇒(iii)", we employ the fundamental theorem of calculus and get

$$\begin{aligned}
\left|f(y) - f(x) - \langle f'(x), y - x\rangle_H\right| &\leq \left|\int_0^1 \langle f'(x + t(y - x)) - f'(x), y - x\rangle_H \,\mathrm{d}t\right| \\
&\leq \int_0^1 \frac{1}{t}\left|\langle f'(x + t(y - x)) - f'(x), t(y - x)\rangle_H\right| \mathrm{d}t \\
&\leq \int_0^1 \frac{1}{t} L\|t(y - x)\|_H^2 \mathrm{d}t = \frac{L}{2}\|y - x\|_H^2.
\end{aligned}$$

In order to check "(iii)⇒(i)", we take an arbitrary $\rho > 0$ and set

$$O_\rho := \{x \in O \mid \forall h \in H, \|h\|_H \leq \rho : x + h \in O\}. \tag{2.4}$$

It is clear that O_ρ is again convex. Let us choose some $x, y \in O_\rho$ with $\|y - x\|_H \leq \rho$ and $d \in H$ with $y + d, x - d \in O$. From (2.3), we get the inequalities

$$\begin{aligned}
f(x - d) - f(y) - \langle f'(y), x - d - y\rangle_H &\leq \frac{L}{2}\|y - x + d\|_H^2, \\
f(y + d) - f(x) - \langle f'(x), y + d - x\rangle_H &\leq \frac{L}{2}\|y - x + d\|_H^2, \\
-(f(y + d) - f(y) - \langle f'(y), \quad d\rangle_H) &\leq \frac{L}{2}\|d\|_H^2, \\
-(f(x - d) - f(x) - \langle f'(x), -d\rangle_H) &\leq \frac{L}{2}\|d\|_H^2.
\end{aligned}$$

Adding these inequalities leads to

$$\langle f'(y) - f'(x), y - x + 2d\rangle_H \le L\|y - x + d\|_H^2 + L\|d\|_H^2.$$

Next, we specialize to $d = (x - y + g)/2$ for some $g \in H$ with $\|g\|_H \le \rho$. Note that $\|d\|_H \le (\|y - x\|_H + \|g\|_H)/2 \le \rho$. Together with $x, y \in O_\rho$ we find $x - d, y + d \in O$. For arbitrary $g \in H$ with $\|g\|_H \le \rho$, this leads to

$$\langle f'(y) - f'(x), g\rangle_H \le \frac{L}{4}\|y - x + g\|_H^2 + \frac{L}{4}\|x - y + g\|_H^2 = \frac{L}{2}\|y - x\|_H^2 + \frac{L}{2}\|g\|_H^2, \quad (2.5)$$

where we used the parallelogram identity. We denote by $G \in H$ the Riesz representative of $f'(y) - f'(x) \in H^\star$. In case $\|G\|_H = \|f'(y) - f'(x)\|_{H^\star} \ge \rho L$, we take $g = \rho G/\|G\|_H$ in (2.5) and obtain

$$\rho\|f'(y) - f'(x)\|_{H^\star} = \langle f'(y) - f'(x), g\rangle_H \le \frac{L}{2}\|y - x\|_H^2 + \frac{L}{2}\|g\|_H^2 \le L\rho^2.$$

This yields $\|G\|_H \le \rho L$. Thus, we can insert $g = G/L$ (in case $L > 0$) in (2.5) and obtain

$$\frac{1}{L}\|f'(y) - f'(x)\|_{H^\star}^2 \le \frac{L}{2}\|y - x\|_H^2 + \frac{1}{2L}\|f'(y) - f'(x)\|_{H^\star}^2.$$

Altogether, this shows

$$\|f'(y) - f'(x)\|_{H^\star} \le L\|y - x\|_H \qquad \forall x, y \in O_\rho, \|y - x\|_H \le \rho.$$

For arbitrary $x, y \in O_\rho$, we can choose $n \in \mathbb{N}$, $n \ge \|y - x\|_H/\rho$ and set $x_i := x + i(y - x)/n$, $i = 0, \ldots, n$. Due to $\|x_i - x_{i-1}\|_H = \|y - x\|_H/n \le \rho$, we can use the above estimate to achieve

$$\begin{aligned}\|f'(y) - f'(x)\|_{H^\star} = \|f'(x_n) - f'(x_0)\|_{H^\star} &\le \sum_{i=1}^{n}\|f'(x_i) - f'(x_{i-1})\|_{H^\star} \\ &\le \sum_{i=1}^{n} L\|x_i - x_{i-1}\|_H = \sum_{i=1}^{n} L\frac{\|y - x\|_H}{n} = L\|y - x\|_H \qquad \forall x, y \in O_\rho.\end{aligned}$$

Finally, $O = \bigcup_{\rho>0} O_\rho$ yields (2.1). □

The implications "(i)⇒(ii)⇒(iii)" remain to hold in the Banach space setting. However, in the proof of "(iii)⇒(i)", we have utilized the parallelogram identity in (2.5). Thus, the proof does not generalize to Banach spaces. Instead, we could employ the triangle inequality, which leads to

$$\frac{L}{4}\|y - x + g\|_H^2 + \frac{L}{4}\|x - y + g\|_H^2 \le L\|y - x\|_H^2 + L\|g\|_H^2.$$

By adapting the remaining part of the proof, we still arrive at (2.1) but with $2L$ instead of L.

By means of an example, we demonstrate that the assertion of Lemma 2.1 indeed fails in Banach spaces.

Example 2.2. *We choose* $X = (\mathbb{R}^2, \|\cdot\|_\infty)$*, thus,* $X^\star = (\mathbb{R}^2, \|\cdot\|_1)$*. We define* $f\colon \mathbb{R}^2 \to \mathbb{R}$ *via*

$$f(x_1, x_2) := \frac{1}{2}(x_1^2 - x_2^2)$$

Thus,

$$|\langle f'(y) - f'(x), y - x\rangle_{\mathbb{R}^2}| = |(y_1 - x_1)^2 - (y_2 - x_2)^2| \le \|y - x\|_\infty^2$$

for all $x, y \in \mathbb{R}^2$*, i.e., (2.2) is satisfied with* $L = 1$*. However,*

$$\|f'(x) - f'(0)\|_1 = |x_1| + |-x_2| \le \hat{L}\|x - 0\|_\infty \quad \forall x \in \mathbb{R}^2$$

only holds for $\hat{L} \ge 2$*. Thus,* $f'\colon X \to X^\star$ *is only Lipschitz continuous with constant* 2*.*

Further, we need a characterization of cocoercive operators on Hilbert spaces.

Lemma 2.3 ([Bauschke and Combettes, 2011, Proposition 4.2])**.** *Let* $O \subset H$ *be a subset of the Hilbert space* H*. Then, for* $T\colon O \to H^\star$ *and* $L > 0$ *the following are equivalent.*

(i) T *is* $1/L$*-cocoercive, i.e.,*

$$\langle T(y) - T(x), y - x\rangle_H \ge \frac{1}{L}\|T(y) - T(x)\|_{H^\star}^2 \qquad \forall x, y \in O.$$

(ii) $2T/L - R$ *is nonexpansive, i.e.,*

$$\|2T(y)/L - R(y) - (2T(x)/L - R(x))\|_{H^\star} \le \|y - x\|_H \qquad \forall x, y \in O.$$

Here, $R\colon H \to H^\star$ *is the Riesz isomorphism of* H*.*

We note that this result follows from some simple and straightforward calculations.

As a last prerequisite, we show that the so-called strong smoothness of f implies Gâteaux differentiability.

Lemma 2.4. *Let* $O \subset X$ *be an open and convex subset of the Banach space* X*. Suppose that* $f\colon O \to \mathbb{R}$ *is convex, lower semicontinuous and strongly smooth, i.e.,*

$$f(\lambda x + (1 - \lambda)y) + \frac{L}{2}\lambda(1 - \lambda)\|y - x\|_X^2 \ge \lambda f(x) + (1 - \lambda)f(y) \quad \forall x, y \in O, \lambda \in (0, 1).$$

holds for some $L > 0$*. Then,* f *is Gâteaux differentiable on* O*.*

Proof. For $x \in O$, $h \in X$ and $t > 0$ small enough, we apply the smoothness inequality to $x \pm th$ and $\lambda = 1/2$. This yields

$$f(x) + \frac{L}{2}\|th\|_X^2 \ge \frac{1}{2}f(x + th) + \frac{1}{2}f(x - th).$$

Sorting terms and dividing by $t/2$ gives

$$0 \ge \lim_{t \searrow 0}\Big(\frac{f(x + th) - f(x)}{t} + \frac{f(x - th) - f(x)}{t} - Lt\|h\|_X^2\Big) = f'(x; h) + f'(x; -h).$$

Recall that the existence of the directional derivatives follows from $x \in O = \operatorname{int}(\operatorname{dom} f)$, see [Zălinescu, 2002, Theorem 2.1.13]. This result also gives the sublinearity $f'(x;h) + f'(x;-h) \geq 0$, thus $f'(x;\cdot)$ is linear. Since f is locally Lipschitz continuous on O by [Zălinescu, 2002, Theorems 2.2.11, 2.2.20], the functional $f'(x;\cdot)$ is Lipschitz continuous as well. Thus, $f'(x;\cdot) \in X^\star$ and this shows the Gâteaux differentiability of f. □

Now we are in position to prove the main result.

Theorem 2.5 (Baillon–Haddad theorem)**.** *Let $O \subset H$ be an open and convex subset of the Hilbert space H. Suppose that $f\colon O \to \mathbb{R}$ is convex. Then, for $L > 0$, the following are equivalent.*

(i) *f is (Gâteaux) differentiable and f' is L-Lipschitz.*

(ii) *$\frac{L}{2}\|\cdot\|_H^2 - f$ is convex and f is lower semicontinuous.*

(iii) *f is (Gâteaux) differentiable and f' is $1/L$-cocoercive.*

Proof. Since H is assumed to be a Hilbert space, it is straightforward to check that (ii) implies the strong smoothness of f, thus f is also Gâteaux differentiable in case (ii), see Lemma 2.4.

We set $h(x) := \frac{L}{2}\|x\|_H^2 - f(x)$ for $x \in O$. For the auxiliary function h we have $h'(x) = LR(x) - f'(x)$. This directly yields

$$\begin{aligned} \text{(i)} \quad &\Leftrightarrow \quad \langle f'(y) - f'(x), y - x\rangle_H \leq L\|y - x\|_H^2 \quad && \forall x, y \in O \\ &\Leftrightarrow \quad \langle h'(y) - h'(x), y - x\rangle_H \geq 0 && \forall x, y \in O \\ &\Leftrightarrow \quad \text{(ii)} \end{aligned}$$

For the other equivalence, we employ Lemma 2.3. To this end, we apply Lemma 2.1 to the function $g\colon O \to \mathbb{R}$ defined via $g(x) := 2f(x)/L - \frac{1}{2}\|x\|_H^2$. Note that $g'(x) = 2f'(x)/L - R$. This yields

$$\begin{aligned} \text{(i)} \quad &\Leftrightarrow \quad 0 \leq \langle f'(y) - f'(x), y - x\rangle_H \leq L\|y - x\|_H^2 \quad && \forall x, y \in O \\ &\Leftrightarrow \quad -\|y - x\|_H^2 \leq \langle g'(y) - g'(x), y - x\rangle_H \leq \|y - x\|_H^2 && \forall x, y \in O \\ &\Leftrightarrow \quad g' \text{ is nonexpansive} \quad \Leftrightarrow \quad \text{(iii)}. \end{aligned}$$

□

The above prove is an adaption of the proof of [Bauschke and Combettes, 2010, Theorem 3.3] to the situation without second-order differentiability and, thus, gives a simple answer to [Bauschke and Combettes, 2010, Remark 3.5]. It is currently not clear whether the equivalence of (i) and (iii) remains to hold if H is only assumed to be a Banach space. Note that both main ingredients of the proof (Lemmas 2.1 and 2.3) cannot be transferred directly to Banach spaces, see also the discussion in the next section. The next example shows that the implication "(i)⇒(ii)" fails in Banach spaces.

Example 2.6. *We choose $X = (\mathbb{R}^2, \|\cdot\|_\infty)$, thus, $X^\star = (\mathbb{R}^2, \|\cdot\|_1)$. We define $f\colon \mathbb{R}^2 \to \mathbb{R}$ via*

$$f(x_1, x_2) := \frac{1}{2}(x_1^2 + x_2^2)$$

Thus,

$$\|f'(y) - f'(x)\|_1 = |y_1 - x_1| + |y_2 - x_2| \le 2\|y - x\|_\infty$$

for all $x, y \in \mathbb{R}^2$. *Moreover, this inequality is satisfied with equality if* $x, y \in \operatorname{span}\{(1,1)\}$. *Hence, (i) is satisfied if and only if* $L \ge 2$.

Define $h(x) := \frac{L}{2}\|x\|_X^2 - f(x) = \max(x_1^2, x_2^2) - \frac{1}{2}(x_1^2 + x_2^2)$. *Then* $h((1,0)) > 0$ *but* $h((1,\pm 1)) = 0$. *Consequently,* h *is not convex, and (ii) is violated. In fact,* h *is not convex for all* $L > 0$.

However, in Hilbert spaces, assertion (ii) is equivalent to the strong smoothness of f. Then, the equivalence between Lipschitzness of f' and strong smoothness of f continues to hold in Banach spaces, see Theorem 3.1 below.

3 Convex functions on open, convex subsets of Banach spaces

In this section, we address generalizations and limitations of Theorem 2.5. In particular, we are interested in convex functions defined on an open subset of a Banach space. We investigate, which of the claims of Theorem 2.5 and conditions well-known to be equivalent to L-Lipschitz continuity of the derivative remain true in this general situation.

Theorem 3.1. *For a convex and lower semicontinuous function* $f\colon O \to \mathbb{R}$, *where* O *is an open and convex subset of a Banach space* X, *we consider the following assertions with some fixed* $L > 0$.

(i) *The function* f *is strongly smooth*

$$f(\lambda x + (1-\lambda)y) + \frac{L}{2}\lambda(1-\lambda)\|y-x\|_X^2 \ge \lambda f(x) + (1-\lambda) f(y) \quad \forall x, y \in O, \lambda \in (0,1).$$

(ii) *The descent lemma holds*

$$f(y) \le f(x) + \langle x^\star, y - x\rangle_X + \frac{L}{2}\|y - x\|_X^2 \qquad \forall x, y \in O, x^\star \in \partial f(x).$$

(iii) *We have*

$$\langle y^\star - x^\star, y - x\rangle_X \le L\|y - x\|_X^2 \qquad \forall (x, x^\star), (y, y^\star) \in \operatorname{graph} \partial f.$$

(iv) *The subdifferential is Lipschitz continuous (thus single-valued)*

$$\|y^\star - x^\star\|_{X^\star} \le L\|y - x\|_X \qquad \forall (x, x^\star), (y, y^\star) \in \operatorname{graph} \partial f.$$

(v) *The subdifferential is cocoercive*

$$\langle y^\star - x^\star, y - x\rangle_X \ge \frac{1}{L}\|y^\star - x^\star\|_{X^\star}^2 \qquad \forall (x, x^\star), (y, y^\star) \in \operatorname{graph} \partial f.$$

(vi) *We have*

$$f(y) \geq f(x) + \langle x^\star, y - x\rangle_X + \frac{1}{2L}\|y^\star - x^\star\|_{X^\star}^2 \qquad \forall (x, x^\star), (y, y^\star) \in \operatorname{graph} \partial f.$$

Any of these conditions imply the Gâteaux differentiability of f on O. Moreover, the following relations hold.

(a) *We have "(i)⇔(ii)⇔(iii)⇔(iv)" and "(vi)⇒(v)⇒(i)".*

(b) *In case that X is a Hilbert space, we have additionally "(iv)⇔(v)". This results in "(i)⇔(ii)⇔(iii)⇔(iv)⇔(v)" and these conditions are implied by (vi).*

(c) *In case that $O = X$, all assertions are equivalent.*

Proof. The case (c) follows from [Zălinescu, 2002, Corollary 3.5.7 and Remark 3.5.2] and case (b) is implied by Theorem 2.5. Lemma 2.4 shows that (i) implies the Gâteaux differentiability of f. Thus, it remains to check the implications from (a). For later reuse we recall that $\partial f(x) \neq \emptyset$ for all $x \in O$, since f is continuous on O, see [Zălinescu, 2002, Theorems 2.2.20 and 2.4.9].

"(i)⇒(ii)": We already know that f is Gâteaux differentiable. Moreover, (i) gives

$$\frac{f(y + \lambda(x - y)) - f(y)}{\lambda} + \frac{L}{2}(1 - \lambda)\|y - x\|_X^2 \geq f(x) - f(y).$$

for arbitrary $\lambda \in (0, 1)$ and $\lambda \searrow 0$ results in

$$\langle f'(y), x - y\rangle_X + \frac{L}{2}\|y - x\|_X^2 \geq f(x) - f(y).$$

Since $\partial f(y) = \{f'(y)\}$, this gives (ii) with exchanged roles of x and y.

"(ii)⇒(i)": We set $x_\lambda := \lambda x + (1 - \lambda)y \in O$ and choose an arbitrary $x_\lambda^\star \in \partial f(x_\lambda)$. Thus,

$$\begin{aligned} f(x) &\leq f(x_\lambda) + \langle x_\lambda^\star, x - x_\lambda\rangle_X + \frac{L}{2}\|x - x_\lambda\|_X^2, \\ f(y) &\leq f(x_\lambda) + \langle x_\lambda^\star, y - x_\lambda\rangle_X + \frac{L}{2}\|y - x_\lambda\|_X^2. \end{aligned}$$

We multiply the first inequality by λ and the second one by $(1-\lambda)$. Adding the resulting inequalities and using $x - x_\lambda = (1 - \lambda)(x - y)$, $y - x_\lambda = \lambda(y - x)$, we get (i).

"(ii)⇒(iii)": This follows from adding the inequality with $(x, x^\star)$ and $(y, y^\star)$ exchanged.

"(iii)⇒(ii)": We choose an arbitrary $n \in \mathbb{N}$ and define

$$x_0 := x, \quad x_n := y, \quad x_i := x + \frac{i}{n}(y - x), \quad x_0^\star := x^\star, \quad x_n^\star := y^\star.$$

Further we choose some arbitrary $x_i^\star \in \partial f(x_i)$ for $i = 1, \ldots, n-1$. Thus,

$$\begin{aligned} f(y) - f(x) - \langle x^\star, y - x\rangle_X &= f(x_n) - f(x_0) - \langle x_0^\star, x_n - x_0\rangle_X \\ &= \sum_{i=1}^n f(x_i) - f(x_{i-1}) - \langle x_0^\star, x_i - x_{i-1}\rangle_X \\ &\le \sum_{i=1}^n \langle x_i^\star - x_0^\star, x_i - x_{i-1}\rangle_X = \sum_{i=1}^n \frac{1}{i}\langle x_i^\star - x_0^\star, x_i - x_0\rangle_X \\ &\le L\sum_{i=1}^n \frac{1}{i}\|x_i - x_0\|_X^2 = L\sum_{i=1}^n \frac{i}{n^2}\|x_1 - x_0\|_X^2 = L\frac{n+1}{2n}\|y - x\|_X^2. \end{aligned}$$

Now, the claim follows from $n \to \infty$.

"(ii)⇒(iv)": Since (ii) implies (i), f is differentiable on O. For an arbitrary $x \in O$, there exists $\rho > 0$ with $B_\rho(x) := \{y \in X \mid \|y - x\|_X \le \rho\} \subset O$. Now, let additionally $y \in B_\rho(x)$ with $x \ne y$ and $\varepsilon > 0$ be given. By the Hahn-Banach theorem we get $z \in X$ with $\|z\|_X = \|y - x\|_X \le \rho$ and $(1-\varepsilon)\|f'(y) - f'(x)\|_{X^\star}\|z\|_X \le \langle f'(y) - f'(x), z\rangle_X$. Thus, $x + z \in B_\rho(x) \subset O$ and we have

$$\begin{aligned} &(1-\varepsilon)\|f'(y) - f'(x)\|_{X^\star}\|z\|_X \\ &\quad\le 0 + \langle f'(y) - f'(x), z\rangle_X \\ &\quad\le [f(x+z) - f(y) - \langle f'(y), x + z - y\rangle_X] + \langle f'(y) - f'(x), z\rangle_X \\ &\quad= [f(x+z) - f(x) - \langle f'(x), z\rangle_X] + [f(x) - f(y) - \langle f'(y), x - y\rangle_X] \\ &\quad\le \frac{L}{2}\|z\|_X^2 + \frac{L}{2}\|y - x\|_X^2 = L\|y - x\|_X^2. \end{aligned}$$

Dividing by $\|z\|_X = \|y - x\|_X$ and passing to the limit $\varepsilon \searrow 0$ yields

$$\|f'(y) - f'(x)\|_{X^\star} \le L\|y - x\|_X$$

for all $x, y \in O$ such that $y \in B_\rho(x) \subset O$ with $\rho > 0$.

Now, if $x, y \in O$ are arbitrary, there exists $\rho > 0$ with $B_\rho(x), B_\rho(y) \subset O$. We pick $n \in \mathbb{N}$ with $\|y - x\|_X / n < \rho$ and define $x_i := x + \frac{i}{n}(y - x)$, $i = 0, \ldots, n$. By convexity we have $B_\rho(x_i) \subset O$ for all $i = 0, \ldots, n$. Thus,

$$\|f'(y) - f'(x)\|_{X^\star} \le \sum_{i=1}^n \|f'(x_i) - f'(x_{i-1})\|_{X^\star} \le \sum_{i=1}^n L\|x_i - x_{i-1}\|_X = L\|y - x\|_X$$

and this shows (iv).

"(vi)⇒(v)": This follows from adding the inequality with $(x, x^\star)$ and $(y, y^\star)$ exchanged.

"(v)⇒(iv)" and "(iv)⇒(iii)" are straightforward. □

Remark 3.2.

(i) *In the case (b), the missing implication "(i)⇒(vi)" cannot hold due to the counterexample by [Drori, 2020, Section 2], which is for the Euclidean case $X = \mathbb{R}^n$.*

(ii) *In the general case, it is currently not clear whether any of the (equivalent) conditions (i), (ii), (iii), (iv) implies condition (v).*

The next lemma shows that the failure of (vi) in case (b) is due to the missing convexity of the range of f'. Since any of the properties in Theorem 3.1 imply Gâteaux differentiability, we will formulate all the following results for a Gâteaux differentiable function to simplify the presentation.

Lemma 3.3 ((v)⇒(vi))**.** *Let $O \subset X$ be an open and convex subset of the Banach space X. Suppose that the convex, lower semicontinuous and Gâteaux differentiable function $f\colon O \to \mathbb{R}$ satisfies Theorem 3.1 (v), i.e., f' is cocoercive with constant $L > 0$. Further, let $x, y \in O$ be given, such that*

$$[f'(x), f'(y)] := \{(1-\lambda)f'(x) + \lambda f'(y) \mid \lambda \in [0,1]\} \subset f'(O) := \{f'(z) \mid z \in O\}.$$

Then,

$$f(y) \geq f(x) + \langle f'(x), y - x\rangle_X + \frac{1}{2L}\|f'(y) - f'(x)\|_{X^\star}^2.$$

That is, Theorem 3.1 (vi) holds on a subset of O.

Proof. Let $n \in \mathbb{N}$ be arbitrary and define

$$x_i^\star := \Big(1 - \frac{i}{n}\Big)f'(x) + \frac{i}{n}f'(y). \tag{3.1}$$

Due to our assumption, there exists $x_i \in O$, $i = 1, \ldots, n-1$, such that $f'(x_i) = x_i^\star$ for all $i = 1, \ldots, n-1$. We further set $x_0 := x$ and $x_n := y$. Now, we have

$$\begin{aligned}
f(y) - f(x) - \langle f'(x), y - x\rangle_X &= \sum_{i=1}^n f(x_i) - f(x_{i-1}) - \langle f'(x_0), x_i - x_{i-1}\rangle_X \\
&\geq \sum_{i=1}^n \langle f'(x_{i-1}) - f'(x_0), x_i - x_{i-1}\rangle_X = \sum_{i=1}^n (i-1)\langle f'(x_i) - f'(x_{i-1}), x_i - x_{i-1}\rangle_X \\
&\geq \frac{1}{L}\sum_{i=1}^n (i-1)\|f'(x_i) - f'(x_{i-1})\|_{X^\star}^2 = \frac{1}{Ln^2}\sum_{i=1}^n (i-1)\|f'(y) - f'(x)\|_{X^\star}^2 \\
&= \frac{n(n-1)}{2Ln^2}\|f'(y) - f'(x)\|_{X^\star}^2.
\end{aligned}$$

Passing to the limit $n \to \infty$ yields the claim. □

Under slightly stronger assumptions on f than in the previous lemma, condition (ii) implies (v).

Lemma 3.4 ((ii)⇒(v))**.** *Let $O \subset X$ be an open and convex subset of the Banach space X. Suppose that $f\colon O \to \mathbb{R}$ is convex, lower semicontinuous, and Gâteaux differentiable. Let $L > 0$ be given such that Theorem 3.1 (ii), which is*

$$f(y) \leq f(x) + \langle f'(x), y - x\rangle_X + \frac{L}{2}\|y - x\|_X^2 \qquad \forall x, y \in O,$$

is satisfied. Let $\rho > 0$ and $x, y \in O_\rho$ be given with O_ρ defined in (2.4). If

$$[f'(x), f'(y)] \subset f'(O_\rho),$$

then

$$\langle f'(y) - f'(x), y - x \rangle_X \geq \frac{1}{L} \|f'(y) - f'(x)\|_{X^\star}^2 \tag{3.2}$$

holds, which shows that Theorem 3.1 (v) holds on a subset of O.

Proof. First, we consider the case that $x, y \in O_\rho$ are given such that $\|f'(y) - f'(x)\|_{X^\star} < L\rho$. Take $d \in X$ with $\|d\|_X < \rho$. Then using the convexity and property (ii), we get

$$\begin{aligned}\langle f'(y) - f'(x), d \rangle_X &= \langle f'(y) - f'(x), y - x \rangle_X + \langle f'(y) - f'(x), d - y + x \rangle_X \\ &\leq \langle f'(y) - f'(x), y - x \rangle_X + f(x + d) - f(y) + f(y - d) - f(x) \\ &\leq \langle f'(y) - f'(x), y - x \rangle_X + \langle f'(y) - f'(x), -d \rangle_X + L\|d\|_X^2.\end{aligned}$$

This implies

$$\langle f'(y) - f'(x), d \rangle_X - \frac{L}{2}\|d\|_X^2 \leq \frac{1}{2}\langle f'(y) - f'(x), y - x \rangle_X.$$

For arbitrary $\varepsilon > 0$, we can choose $d \in X$ with $\|d\|_X = \frac{1}{L}\|f'(y) - f'(x)\|_{X^\star}$ such that $\langle f'(y) - f'(x), d \rangle_X \geq \frac{1}{L}\|f'(y) - f'(x)\|_{X^\star}^2 - \varepsilon$. This implies

$$\frac{1}{2L}\|f'(y) - f'(x)\|_{X^\star}^2 - \varepsilon \leq \frac{1}{2}\langle f'(y) - f'(x), y - x \rangle_X.$$

Since $\varepsilon > 0$ was arbitrary, this shows (3.2) in case that $\|f'(y) - f'(x)\|_{X^\star}$ is small.

Now, let $x, y \in O_\rho$ be given such that $[f'(x), f'(x)] \subset f'(O_\rho)$. We use the same construction as in the proof of Lemma 3.3. Let a number $n \in \mathbb{N}$ be given such that $\|f'(y) - f'(x)\|_{X^\star} < L\rho n$. Let $x_i^\star$ be given as in (3.1) and choose $x_i \in O_\rho$ with $f'(x_i) = x_i^\star$ for all $i = 1, \ldots, n-1$. We set $x_0 = x$ and $x_n := y$. By construction, we have $\|f'(x_{i+1}) - f'(x_i)\|_{X^\star} = \frac{1}{n}\|f'(y) - f'(x)\|_{X^\star} < L\rho$. Then, the first part of the proof gives

$$\frac{1}{L}\|f'(x_i) - f'(x_{i-1})\|_{X^\star}^2 \leq \langle f'(x_i) - f'(x_{i-1}), x_i - x_{i-1} \rangle_X \quad \forall i = 1, \ldots, n.$$

By construction, this implies

$$\frac{1}{Ln^2}\|f'(y) - f'(x)\|_{X^\star}^2 \leq \frac{1}{n}\langle f'(y) - f'(x), x_i - x_{i-1} \rangle_X \quad \forall i = 1, \ldots, n.$$

Summation over $i = 1, \ldots, n$ yields the claim. □

References

Baillon, Jean-Bernard and Haddad, Georges (1977). “Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones”. In: *Israel Journal of Mathematics* 26.2, pp. 137–150. DOI: 10.1007/bf03007664.

Bauschke, Heinz H. and Combettes, Patrick L. (2010). "The Baillon–Haddad Theorem Revisited". In: *Journal of Convex Analysis* 17.3&4, pp. 781–787. URL: https://www.heldermann.de/JCA/JCA17/JCA173/jca17051.htm.

Bauschke, Heinz H. and Combettes, Patrick L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces*. Berlin: Springer. DOI: 10.1007/978-1-4419-9467-7.

Drori, Yoel (2020). "On the Properties of Convex Functions over Open Sets". In: *Journal of Convex Analysis* 27.4, pp. 1303–1314. URL: https://www.heldermann.de/JCA/JCA27/JCA274/jca27069.htm.

Pérez-Aros, Pedro and Vilches, Emilio (2019). "An Enhanced Baillon–Haddad Theorem for Convex Functions Defined on Convex Sets". In: *Applied Mathematics & Optimization* 83.3, pp. 2241–2252. DOI: 10.1007/s00245-019-09626-6.

Zălinescu, Constantin (2002). *Convex Analysis in General Vector Spaces*. World Scientific. DOI: 10.1142/5021.