

**ESSENTIAL SMOOTHNESS,
ESSENTIAL STRICT CONVEXITY,
AND LEGENDRE FUNCTIONS IN BANACH SPACES**

HEINZ H. BAUSCHKE*

*Department of Mathematics and Statistics,
University of Guelph, Guelph, Ontario N1G 2W1, Canada
bauschke@cecm.sfu.ca*

JONATHAN M. BORWEIN†

*Centre for Experimental & Constructive Mathematics,
Simon Fraser University, Burnaby,
British Columbia V5A 1S6, Canada
jborwein@cecm.sfu.ca*

PATRICK L. COMBETTES‡

*Laboratoire d'Analyse Numérique,
Université Pierre et Marie Curie — Paris 6,
4 Place Jussieu, 75005 Paris, France
plc@ann.jussieu.fr*

Received 21 December 2000

The classical notions of essential smoothness, essential strict convexity, and Legendreness for convex functions are extended from Euclidean to Banach spaces. A pertinent duality theory is developed and several useful characterizations are given. The proofs rely on new results on the more subtle behavior of subdifferentials and directional derivatives at boundary points of the domain. In weak Asplund spaces, a new formula allows the recovery of the subdifferential from nearby gradients. Finally, it is shown that every Legendre function on a reflexive Banach space is zone consistent, a fundamental property in the analysis of optimization algorithms based on Bregman distances. Numerous illustrating examples are provided.

Keywords: Bregman distance, Bregman projection; coercive; cofinite function; convex function of Legendre type; essentially smooth; essentially strictly convex; Legendre function; Schur property; Schur space; subdifferential; supercoercive; weak Asplund space; zone consistent.

Mathematics Subject Classification 2000: Primary 52A41; Secondary 46G05, 46N10, 49J50, 90C25

*Research supported by NSERC.

†Research supported by NSERC and by the Shrum Endowment.

‡Research supported in part by the National Science Foundation under grant MIP-9705504.

1. Introduction

1.1. Classical Legendre functions (in Euclidean spaces)

We start by reviewing some of Rockafellar's classical results on Legendre functions [41, Sec. 26]: Suppose $f: \mathbb{R}^M \rightarrow]-\infty, +\infty]$ is convex, lower semicontinuous, and proper. Then f is called:

- *essentially smooth*, if it is differentiable on $\text{int dom } f \neq \emptyset$, and $\|\nabla f(x_n)\| \rightarrow +\infty$ whenever $x_n \rightarrow x \in \text{bdry dom } f$;
- *essentially strictly convex*, if it is strictly convex on every convex subset of $\text{dom } \partial f$;
- *Legendre*, if it is both essentially smooth and essentially strictly convex.

The corresponding theory is both very elegant and powerful: f is essentially smooth if and only if its conjugate f^* is essentially strictly convex. Consequently, f is Legendre if and only if f^* is, in which case ∇f is an isomorphism between $\text{int dom } f$ and $\text{int dom } f^*$. Many functions in convex optimization are Legendre [5]; perhaps most notably, the log barrier in Interior Point Methods [34].

1.2. An application: the method of cyclic Bregman projections

We now demonstrate the power of Legendre functions by studying a specific optimization problem. Suppose C_1, \dots, C_N are closed convex sets ("the constraints") in \mathbb{R}^M with $C = \bigcap_{i=1}^N C_i \neq \emptyset$. The *convex feasibility problem* consists of finding a point ("a solution") in C . Suppose further that the orthogonal projection onto each set C_i , which we denote by P_i , is readily computable. Then *the method of cyclic (orthogonal) projections* operates as follows:

Given a starting point y_0 , generate a sequence (y_n) by projecting cyclically onto the constraints:

$$y_0 \xrightarrow{P_1} y_1 \xrightarrow{P_2} y_2 \xrightarrow{P_3} \dots \xrightarrow{P_N} y_N \xrightarrow{P_1} y_{N+1} \xrightarrow{P_2} \dots$$

The sequence (y_n) does indeed converge to a solution of the convex feasibility problem [4].

In some applications, however, it is desirable to employ the method of cyclic projections with (*nonorthogonal*) *Bregman projections* [16]. These are constructed as follows. Given a "sufficiently nice" convex function f , the *Bregman distance* between x and y is

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle,$$

where $y \in \text{int dom } f$ is a point of differentiability of f . Then the *Bregman projection* of y onto the i th constraint C_i with respect to f is defined by

$$\text{arginf}_{x \in C_i} D_f(x, y).$$

Here we have implicitly assumed that y is a point of differentiability so that $D_f(x, y)$ is well defined. More importantly, to define the sequence of cyclic projections unambiguously, the following is required: • the arginf is nonempty ("existence of

nearest points”), • the arginf is a singleton (“no selection necessary”), • the arginf is contained in $\text{int dom } f$ (in order to project the arginf onto the next constraint C_{i+1}).

The punch-line is that if f is Legendre, then all these good properties hold [5] — in the terminology of Censor and Lent [19], “every Legendre function is zone consistent”. Moreover, the Legendre property is the most general condition known to date [5] that guarantees zone consistency.

1.3. Objective in this paper

The objective in this paper is to extend the classical notions of essential smoothness, essential strict convexity, and Legendreness from Euclidean to Banach spaces, to furnish an elegant and effective concomitant theory, and to demonstrate the applicability of these new notions.

1.4. Standing assumptions

Throughout the paper, we assume that

$$X \text{ is a real Banach space with norm } \|\cdot\|$$

and that

$$f: X \rightarrow]-\infty, +\infty] \text{ is a proper convex lower semicontinuous function.}$$

1.5. Summary of the main results

We say that f is:

- *essentially smooth*, if ∂f is both locally bounded and single-valued on its domain;
- *essentially strictly convex*, if $(\partial f)^{-1}$ is locally bounded on its domain and f is strictly convex on every convex subset of $\text{dom } \partial f$;
- *Legendre*, if it is both essentially smooth and essentially strictly convex.

The most important results are the following: (**compatibility**) the new notions agree with the classical ones in Euclidean spaces (Theorem 5.11); (**duality**) in reflexive spaces, f is Legendre if and only if f^* is (Corollary 5.4); various **characterizations of essential smoothness** in Banach spaces (Theorem 5.6); a **subdifferential formula** that is particularly useful in weak Asplund spaces (Theorem 4.5); **Legendre functions are zone consistent** in reflexive spaces (Corollary 7.8).

Furthermore, we believe that the results gathered and refined during the course of this study comprise a part of the theory on convex functions in Banach (especially reflexive) spaces that is not only of great utility in optimization — as illustrated by the applications in the later sections and in [6] — but also of significant value in its own right.

1.6. Organization of the paper

In Sec. 2, we review (and sometimes extend) basic facts from convex analysis. Coercivity, supercoercivity, and cofiniteness are discussed in Sec. 3, where we also characterize spaces in which every cofinite function is necessarily supercoercive (Theorem 3.6). Section 4 contains crucial results on the more subtle properties of directional derivatives and subgradients at boundary points of the domain. We obtain a powerful subdifferential formula (Theorem 4.5) that becomes particularly useful in weak Asplund spaces.

Essential smoothness, essential strict convexity, and Legendreanness are introduced in the fifth section. We present basic duality results, very useful characterizations, and some examples. It is also shown that the new notions coincide with the classical ones in Euclidean spaces. Section 6 is devoted to the discussion of further examples. The final Sec. 7 contains useful properties of Bregman distances and Bregman projections in reflexive spaces. We conclude with the striking result that Legendre functions are zone consistent.

1.7. Notation

The notation we employ is for the most part standard (see [1] for details).

The topological dual of X is denoted by X^* . $B_X = \{x \in X: \|x\| \leq 1\}$ (respectively $S_X = \{x \in X: \|x\| = 1\}$) is the *unit ball* (respectively *unit sphere*); if $x \in X$ and $r \in \mathbb{R}$, then $B(x; r) = x + rB_X$.

A function $g: X \rightarrow [-\infty, +\infty]$ is *convex* if its *epigraph* $\text{epi } g = \{(x, r) \in X \times \mathbb{R}: g(x) \leq r\}$ is a convex set.

Suppose S is a set in X , and $x \in X$. The *interior* (*closure*, *boundary*, *convex hull*, *recession cone* respectively) of S is denoted by $\text{int } S$ ($\text{cl } S$, $\text{bdry } S$, $\text{conv } S$, $\text{rec } S$ respectively). The *indicator function* of S is defined by $\iota_S(x) = 0$, if $x \in S$; $+\infty$, otherwise. The *normal cone* (respectively *tangent cone*) to S at a point $x \in S$ is denoted by $N_S(x)$ (respectively $T_S(x)$). For convenience, we set $N_S(x) = \emptyset$ if $x \in X \setminus S$. Given $x, y \in X$ the set $[x, y] = \{(1-t)x + ty: 0 \leq t \leq 1\}$ (respectively $]x, y[= \{(1-t)x + ty: 0 < t < 1\}$) is the *closed* (respectively *open*) *line segment* between x and y ; half-open segments are defined analogously.

The *domain* (respectively *range*) of a set-valued map T from X to 2^{X^*} is $\{x \in X: Tx \neq \emptyset\}$ (respectively $\bigcup_{x \in X} Tx$), written $\text{dom } T$ (respectively $\text{ran } T$).

Finally, convergence with respect to the norm (respectively weak, weak*) topology of a sequence/net is indicated through \rightarrow (respectively \rightharpoonup , $\overset{w^*}{\rightarrow}$).

2. Facts

Most of the following results are known, and their proofs can be pieced together from various sources such as [1, 3, 27, 29, 36, 37, 43]. We restate them here for the reader's convenience.

2.1. Convex sets

Fact 2.1 (Accessibility Lemma). Suppose C is a convex set in X , and $0 < \lambda \leq 1$. Then $\lambda \operatorname{int}(C) + (1 - \lambda) \operatorname{cl}(C) \subseteq \operatorname{int}(C)$. Consequently, if $\operatorname{int} C \neq \emptyset$, then $\operatorname{cl} C = \operatorname{cl} \operatorname{int} C$.

Proof. See [29, Eq. 11.1 and Lemma 11.A on p. 59]. □

Fact 2.2. Suppose C is a convex set in X with $\operatorname{int} C \neq \emptyset$ and $x \in C$. Then

- (i) $\operatorname{int} T_C(x) = \bigcup_{r>0} r(\operatorname{int} C - x)$.
- (ii) $T_C(x) + \operatorname{int} T_C(x) = \operatorname{int} T_C(x)$.

Proof. (i) See [1, Proposition 4.1.7]. (ii) Follows easily from (i) and Fact 2.0. □

2.2. Continuity and properness

Fact 2.3. Suppose $g: X \rightarrow]-\infty, +\infty]$ is proper and convex and let $x \in \operatorname{dom} g$. Then the following are equivalent.

- (i) g is Lipschitz in a neighbourhood of x .
- (ii) g is continuous at x .
- (iii) g maps some neighbourhood of x to a bounded set.
- (iv) g maps some neighbourhood x to a set that is bounded above.

If one of these conditions holds, then g is continuous throughout $\operatorname{int} \operatorname{dom} g$. Finally, if g is also lower semicontinuous, then the above conditions are equivalent to

- (v) $x \in \operatorname{int} \operatorname{dom} g$.

Proof. Combine [29, Theorem 14.A], [37, Proposition 1.6], and [37, Proposition 3.3]. □

Fact 2.4. Suppose $g: X \rightarrow]-\infty, +\infty]$ is convex. If g is finite at some point in $\operatorname{int} \operatorname{dom} g$, then g is proper.

Proof. The proof of [11, Lemma 3.2.6] works in our setting without change. □

2.3. The directional derivative and subgradients

The following notions are fundamental in convex analysis. Fix $x \in \operatorname{dom} f$. The *directional derivative* of f at x in direction $h \in X$ is defined by $f'(x; h) = \lim_{t \rightarrow 0^+} (f(x + th) - f(x))/t$. The function $f'(x; \cdot)$ is convex and positively homogeneous, i.e. sub-linear. The set $\partial f(x) = \{x^* \in X^*: \langle x^*, h \rangle \leq f(x + h) - f(x), \forall h \in X\}$ is the *subdifferential* of f at x and its elements are called *subgradients*.

We start with a simple yet useful observation.

Fact 2.5. f' is finite and upper semicontinuous on $\text{int dom } f \times X$. If $x \in \text{int dom } f$, then $f'(x; \cdot)$ is continuous on X .

Proof. Fix $x \in \text{int dom } f$ and $h \in X$. We clearly have $f'(x; h) < +\infty$. On the other hand, using local Lipschitzness of f at x (Fact 2.3), it follows readily that $f'(x; h) > -\infty$. Altogether, f' is finite on $\text{int dom } f \times X$. Finally, for all $\epsilon > 0$ sufficiently small, $f'(x; h) = \inf_{0 < t < \epsilon} (f(x + th) - f(x))/t$ is the pointwise infimum of continuous functions on $\text{int dom } f \times X$; therefore, f' is upper semicontinuous on this set. For the “If” statement, see [37, Corollary 1.7]. \square

The next few results are classical.

Fact 2.6. $\text{int dom } f \subseteq \text{dom } \partial f$.

Proof. Combine [37, Proposition 2.24] with [37, Proposition 3.3]. \square

Fact 2.7 (Brøndsted–Rockafellar). $\text{cl dom } \partial f = \text{cl dom } f$.

Proof. See [37, Theorem 3.17]. \square

Fact 2.8 (Rockafellar). ∂f is a maximal monotone operator from X to 2^{X^*} .

Proof. See, for instance, [37, Theorem 3.24]. \square

Fact 2.9 (Moreau’s Classical Max Formula). Suppose $x \in \text{int dom } f$, and $h \in X$. Then $f'(x; h) = \max\langle \partial f(x), h \rangle$.

Proof. See [36, Sec. 10, p. 65] or [37, Proposition 2.24]. \square

Subsequently, we shall require the following generalization of the classical Max Formula. (Note that Fact 2.5 shows that this result indeed generalizes Fact 2.0.)

Theorem 2.10 (Generalized Max Formula). Let $x \in \text{dom } f$ and $f'(x; \cdot)$ be continuous at $h \in X$. Then $f'(x; h) = \max\langle \partial f(x), h \rangle$. In particular, $\partial f(x) \neq \emptyset$.

Proof. Let $p = f'(x; \cdot)$. Then p is convex and, since $h \in \text{int dom } p$, p must be proper (Fact 2.4). By the classical max formula (Fact 2.0), $p'(h; k) = \max\langle \partial p(h), k \rangle$, for every $k \in X$. Fix $x^* \in \partial p(h)$ and $k \in X$. Then $\langle x^*, k \rangle \leq p(h + k) - p(h) \leq p(k) = f'(x; k)$. Hence $x^* \in \partial f(x)$ and thus $\partial p(h) \subseteq \partial f(x)$. It follows that $\partial f(x)$ is nonempty and $\sup\langle \partial f(x), h \rangle \leq f'(x; h)$. Pick $x^* \in \partial p(h)$ such that $p'(h; h) = \langle x^*, h \rangle$. Then $x^* \in \partial f(x)$ and $\langle x^*, h \rangle = p'(h; h) = \lim_{t \rightarrow 0^+} (p(h + th) - p(h))/t = p(h) = f'(x; h)$. \square

The next example demonstrates that continuity at h in Theorem 2.10 is important.

Example 2.11. Let X be the Euclidean plane, $f = \iota_{B_X}$, and $x = (0, -1)$. It is not hard to check that $f'(x; \cdot)$ is the indicator function of the open upper halfplane, whereas $\sup\langle \partial f(x), \cdot \rangle$ is the indicator function of the closed upper halfplane. Hence the generalized max formula holds precisely at points where $f'(x; \cdot)$ is continuous.

2.4. Conjugates

The *conjugate* of a function $g: X \rightarrow [-\infty, +\infty]$ is the (lower semicontinuous convex) function $g^*: X^* \rightarrow [-\infty, +\infty]: x^* \mapsto \sup_{x \in X} \langle x^*, x \rangle - g(x)$.

Fact 2.12. Suppose $x \in X$ and $x^* \in X^*$. Then $f(x) + f^*(x^*) \geq \langle x^*, x \rangle$, and equality holds if and only if $x^* \in \partial f(x)$.

Proof. See [1, Proposition 4.4.3] or [36, Sec. 10]. □

Fact 2.13. $f^{**}|_X = f$.

Proof. See [3, Theorem 2.1.4, pp. 97–99] or [43, Remark 6.3]. □

2.5. Closures

The next three lemmata relate closure operations to conjugates.

Lemma 2.14. *Suppose $g: X^* \rightarrow [-\infty, +\infty]$ is convex, lower semicontinuous, and proper. Then $(g^*|_X)^* = g$ if and only if g is weak* lower semicontinuous.*

Proof. Modify the standard proof of Fact 2.13, for instance, [3, Proof of Theorem 2.1.4, pp. 97–99]. □

Lemma 2.15. *Suppose S is a set in X^* . Then $(\iota_S^*|_X)^* = \iota_{\text{cl}_{w^*}\text{conv } S}$.*

Proof. Fix $x^* \in X^*$. We discuss two cases. If $x^* \in \text{cl}_{w^*}\text{conv } S$, then both functions evaluate to 0. Otherwise, $x^* \notin \text{cl}_{w^*}\text{conv } S$, in which $\iota_{\text{cl}_{w^*}\text{conv } S}(x^*) = +\infty$. Separating weak* yields $x_1 \in X \setminus \{0\}$ such that $\langle x^*, x_1 \rangle > \sup\langle S, x_1 \rangle = (\iota_S^*|_X)(x_1)$. By evaluating $\iota_S^*|_X$ at nx_1 , where $n \rightarrow +\infty$, we readily deduce that $(\iota_S^*|_X)^*(x^*) = +\infty$. □

We define the *closure* of a convex function g via its epigraph, namely, $\text{epi}(\text{cl } g) = \text{cl}(\text{epi } g)$. This is in accordance with [11, Sec. 4.2], but it differs (for improper functions) from Rockafellar’s definition in [41, Sec. 7].

Lemma 2.16. *Suppose $g: X \rightarrow [-\infty, +\infty]$ is convex, positively homogeneous, and $g \not\equiv +\infty$. Set $S = \{x^* \in X^*: g^*(x^*) \leq 0\}$. Then $(\text{cl } g)^* = g^* = \iota_S$ and $g^{**} = \iota_S^*$. Consequently, if g is lower semicontinuous at some point where it is finite, then $\text{cl } g = g^{**}|_X = \iota_S^*|_X$.*

Proof. This follows along familiar steps which we only sketch: (1) $\text{cl}g$ is a well-defined function, convex and lower semicontinuous. (2) The definitions easily yield $(\text{cl}g)^* = g^* = \iota_S$. (3) Hence $g^{**} = (\text{cl}g)^{**} = \iota_S^*$ and so $g^{**}|_X = (\text{cl}g)^{**}|_X = \iota_S^*|_X$. (4) If $g(x) \in \mathbb{R}$, then: g is lower semicontinuous at x if and only if $g(x) = (\text{cl}g)(x)$; in this case, $\text{cl}g$ is proper. (5) Assume now in addition that g is finite and lower semicontinuous at some point. By (1) and (4), $\text{cl}g$ is convex, lower semicontinuous, and proper. It follows that $\text{cl}g = (\text{cl}g)^{**}|_X$ by Fact 2.13, and we are done. \square

2.6. Local boundedness

Recall that a set-valued map T from X to 2^{X^*} is *locally bounded* at a point $x \in X$ if there exists $\epsilon > 0$ such that $\sup \|T(B(x; \epsilon))\| < +\infty$. (See [43, Sec. 17]. This differs slightly from Phelps’s definition [37, Chap. 2] which requires $x \in \text{dom}T$.)

Fact 2.17 (Rockafellar–Vesely’). Suppose T is a maximal monotone operator from X to 2^{X^*} , and $x \in \text{cl dom}T$.

- (i) If $x \in \text{int dom}T$, then T is locally bounded at x .
- (ii) If $\text{cl dom}T$ is convex and T is locally bounded at x , then $x \in \text{int dom}T$.

Proof. (i) this is Rockafellar’s [40, Theorem 1]; see also [37, Theorem 2.28]. (ii) is due to Vesely’ — see [37, Remarks on Chap. 2]. \square

Remark 2.18. It is unknown whether $\text{cl dom}T$ can fail to be convex [43, Problem 18.9]. However, $\text{cl dom}T$ is convex in any of the following cases: (i) T is a subdifferential (Fact 2.0); (ii) X is reflexive [43, Theorem 18.6]; (iii) T is of type (FPV) [43, Theorem 26.3]; (iv) $\text{int conv dom}T \neq \emptyset$ [40, Theorem 1] and also [43, Theorems 18.3 and 18.4]. In fact, (iii) generalizes both (i) and (ii); see Simons’s [43] for background and references.

Corollary 2.19. Suppose $x \in \text{dom } \partial f$. Then $x \in \text{int dom } f$ if and only if ∂f is locally bounded at x .

Proof. By Fact 2.0, $\text{cl dom } \partial f = \text{cl dom } f$. But the latter set is clearly convex. The result now follows from Fact 2.10. \square

We conclude this section with a result which complements Corollary 2.19 in the sense that it will tell us when to expect an unbounded subdifferential.

Lemma 2.20. Suppose T is a maximal monotone operator map from X to 2^{X^*} . Then $N_{\text{dom}T} = \text{rec}T$. In particular:

- (i) $N_{\text{dom}f}(x) = \text{rec } \partial f(x)$, for every $x \in \text{dom } \partial f$.
- (ii) If $\text{int dom } f \neq \emptyset$ and $x \in \text{dom } \partial f \setminus \text{int dom } f$, then $\partial f(x)$ is unbounded.

Proof. Fix $x \in \text{dom } T$. “ \subseteq ”: Fix $z^* \in N_{\text{dom } T}(x)$. Then $\langle z^*, y-x \rangle \leq 0, \forall y \in \text{dom } T$. Pick $x^* \in Tx$ and $p \geq 0$. Then $\langle (x^* + pz^*) - y^*, x-y \rangle = \langle x^* - y^*, x-y \rangle + p\langle z^*, x-y \rangle \geq 0, \forall y \in \text{dom } T, \forall y^* \in Ty$. By maximal monotonicity, $x^* + pz^* \in Tx, \forall p \geq 0$. Hence $z^* \in \text{rec } Tx$. “ \supseteq ”: Now let $z^* \in \text{rec } Tx, y \in \text{dom } T, y^* \in Ty$, and $p \geq 0$. Then $x^* + pz^* \in Tx$ and so $0 \leq \langle (x^* + pz^*) - y^*, x-y \rangle = \langle x^* - y^*, x-y \rangle + p\langle z^*, x-y \rangle$. This is true for all $p \geq 0$; thus, letting tend p to $+\infty$, we learn that $\langle z^*, x-y \rangle \geq 0$. Since y is arbitrary, we have $z^* \in N_{\text{dom } T}(x)$, as required.

(i) is now clear: $T = \partial f$ is maximal monotone (Fact 2.0), and $N_{\text{dom } f}(x) = N_{\text{dom } \partial f}(x), \forall x \in \text{dom } \partial f$ (by Fact 2.0).

(ii) we separate x from $\text{int dom } f$ by, say, $z^* \neq 0: \langle z^*, x \rangle \geq \sup \langle z^*, \text{dom } f \rangle$. It follows that $z^* \in N_{\text{dom } f}(x)$. By (i), $\partial f(x)$ is (even linearly) unbounded. □

3. Coercivities and the Schur Property

3.1. Coercivity

Recall that f is *coercive*, if $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$.

The following result is not as well known as it should be.

Fact 3.1 (Moreau–Rockafellar). Suppose $y^* \in X^*$. Then $f - y^*$ is coercive if and only if $y^* \in \text{int dom } f^*$.

Proof. [35] and [39, Theorem 7A(a)]. □

3.2. Supercoercivity

Lemma 3.2. Suppose $\alpha > 0$. Consider the following conditions:

- (i) $\underline{\lim}_{\|x\| \rightarrow +\infty} f(x)/\|x\| > \alpha$.
- (ii) There exists $\beta \in \mathbb{R}$ such that $f \geq \alpha\|\cdot\| + \beta$.
- (iii) There exists $\gamma \in \mathbb{R}$ such that $f^* \leq \gamma$ on αB_{X^*} .
- (iv) $\underline{\lim}_{\|x\| \rightarrow +\infty} f(x)/\|x\| \geq \alpha$.

Then: (i) \Rightarrow (ii) \Leftrightarrow (iii) \Rightarrow (iv).

Proof. (i) \Rightarrow (ii): There exists $\eta > 0$ such that:

$$\|x\| > \eta \quad \Rightarrow \quad f(x) \geq \alpha\|x\|.$$

On the other hand, the existence of subgradients (guaranteed by Fact 2.0) readily yields $-\infty < \mu = \inf f(\eta B_X)$. Thus if $x \in \eta B_X$, then $\alpha\|x\| \leq \alpha\eta \leq (\alpha\eta - \mu) + f(x)$. Hence:

$$\|x\| \leq \eta \quad \Rightarrow \quad f(x) \geq \alpha\|x\| + (\mu - \alpha\eta).$$

Altogether, (ii) holds with $\beta = \min\{0, \mu - \alpha\eta\} \in \mathbb{R}$.

$$\begin{aligned}
 \text{(ii)} \Leftrightarrow \text{(iii)}: \alpha \|\cdot\| + \beta \leq f &\Leftrightarrow f^* \leq (\alpha \|\cdot\| + \beta)^* \Leftrightarrow f^* \leq \iota_{\alpha B_{X^*}} - \beta. \\
 \text{(ii)} \Rightarrow \text{(iv)}: \underline{\lim}_{\|x\| \rightarrow +\infty} f(x)/\|x\| &\geq \alpha + \underline{\lim}_{\|x\| \rightarrow +\infty} \beta/\|x\| = \alpha. \quad \square
 \end{aligned}$$

The next result defines and characterizes supercoercivity, a condition much more restrictive than coercivity.

Theorem 3.3 (Supercoercivity). *The following are equivalent:*

- (i) f is supercoercive: $\lim_{\|x\| \rightarrow +\infty} f(x)/\|x\| = +\infty$.
- (ii) f^* is bounded above on bounded sets.
- (iii) $\text{dom } \partial f^* = X^*$ and ∂f^* maps bounded sets to bounded sets.

Proof. (i) \Leftrightarrow (ii): See Lemma 3.2. (ii) \Leftrightarrow (iii): The proof of [4, Proposition 7.8] works in our setting without change. □

Theorem 3.4. *Consider the following conditions:*

- (i) f is supercoercive.
- (ii) $f - y^*$ is coercive, for every $y^* \in X^*$.
- (iii) f is cofinite: $\text{dom } f^* = X^*$.

Then: (i) \Rightarrow (ii) \Leftrightarrow (iii). If X is finite-dimensional, then (i) \Leftarrow (ii).

Proof. (i) \Rightarrow (iii): See Theorem 3.3. (ii) \Leftrightarrow (iii): See Fact 3.0.

(i) \Leftarrow (ii) when X is finite-dimensional: We argue by contradiction. Let (x_n) be a sequence in X and $\eta > 0$ such that $0 < \|x_n\| \rightarrow +\infty$ and $f(x_n)/\|x_n\| \leq \eta$, for every n . Abbreviate $x_n/\|x_n\|$ by q_n . Passing to a subsequence if necessary, we may assume that (q_n) converges to some point $q \in S_X$. Now pick $q^* \in J(q)$, where J is the normalized duality map, and let $y^* = r q^*$, where $r = 2\eta > 0$. Since $f - y^*$ is coercive, we have $f(x_n) - r \langle q^*, x_n \rangle \rightarrow +\infty$. On the other hand, $\langle q^*, q_n \rangle \rightarrow \langle q^*, q \rangle = \|q\|^2 = 1$. Hence, for n sufficiently large, $\langle q^*, q_n \rangle \geq 1/2$ and therefore

$$+\infty \leftarrow f(x_n) - r \langle q^*, x_n \rangle = \|x_n\| \left(\frac{f(x_n)}{\|x_n\|} - r \langle q^*, q_n \rangle \right) \leq \|x_n\| (\eta - r/2).$$

Thus necessarily $2\eta > r = 2\eta$, which is absurd. □

Remark 3.5. In Example 7.0 below, we present an explicit function that is cofinite but not supercoercive.

3.3. The Schur property

In finite-dimensional spaces, every cofinite convex function is necessarily supercoercive — this is essentially due to Rockafellar; see [5, Proposition 2.16]. Clearly, it is interesting and useful to know in which spaces cofinite functions are necessarily supercoercive. The following theorem provides a complete answer.

Theorem 3.6. *The following are equivalent:*

- (i) *X has the Schur property: every weakly compact set in X is compact.*
- (ii) *Every convex continuous everywhere finite weak* lower semicontinuous function on X* maps bounded sets to bounded sets.*
- (iii) *Every proper convex lower semicontinuous cofinite function on X is supercoercive.*

Proof. (i) \Leftrightarrow (ii): See [7, Theorem 4.1].

(ii) \Rightarrow (iii): Suppose g is proper, convex, lower semicontinuous, and cofinite on X . Then g^* is convex continuous and weak* lower semicontinuous on X^* . Hence g^* is bounded (above) on bounded sets. By Theorem 3.3, g is supercoercive.

(ii) \Leftarrow (iii): Suppose g is convex everywhere continuous and weak* lower semicontinuous on X^* . Set $h = g^*|_X$. By Lemma 2.14, $h^* = g$. So h is cofinite, hence supercoercive. By Theorem 3.3, g maps bounded sets to bounded sets. \square

We now digress briefly to discuss examples of spaces possessing the Schur property. We require below the following:

Fact 3.7. Let $X = C(K)$, where K is a compact Hausdorff space. Then the following are equivalent.

- (i) X does not contain an isomorphic copy of ℓ_1 .
- (ii) X is an Asplund space [37, Definition 1.22].
- (iii) K is *scattered* [29, Sec. 25.C]: every closed nonempty subset of K contains an isolated point.
- (iv) X does not contain an isometric copy of $C[0, 1]$.

Proof. (See also [8, Theorem 4.3].)

(ii) \Leftrightarrow (iii): [24, Lemma VI.8.3 on p. 258].

(ii) \Rightarrow (i): Otherwise, X does contain an isomorphic copy Y of ℓ_1 . Then Y^* is an isomorphic copy of ℓ_∞ . Since ℓ_∞ is not separable, we have contradicted that assumption that X is Asplund.

(i) \Rightarrow (iv): Otherwise X does contain an isometric copy of $C[0, 1]$. Now $C[0, 1]$ is universal for separable spaces [29, Theorem 25.B] and ℓ_1 is separable. Altogether, X contains an isometric copy of ℓ_1 , which is absurd.

(iv) \Rightarrow (iii): We prove the contrapositive, and thus assume that K is not scattered. By [31, Theorem 2 on p. 29] (see also [29, Lemma 25.C.2] when K is actually a metric space), there exists a continuous map x from K onto $[0, 1]$. It is now straightforward to verify that $T: C[0, 1] \rightarrow C(K): y \mapsto (y \circ x)$ is an isometry. So $X = C(K)$ contains an isometric copy of $C[0, 1]$, hence (iv) does not hold. \square

Remark 3.8. Asplund $C(K)$ spaces are well understood in the sense that they are characterized by K being scattered (Fact 3.7) — the most basic example is the Alexandrov compactification of a countable discrete metric space. (See also [24,

Sec. VI.8].) However, they are sufficiently complex to host remarkable constructions due to Haydon. In [28], he constructed a scattered set K so that $\bullet C(K)$ is Asplund, but $\bullet C(K)$ has no smooth renorm, and $\bullet C(K)$ has no rotund renorm!

We are now ready to record examples of spaces with the Schur property.

Example 3.9. The following spaces possess the Schur property:

- (i) finite-dimensional spaces,
- (ii) ℓ_1 , and
- (iii) the dual of $C(K)$, where K is compact, Hausdorff, and scattered.

Proof. (i) trivial. (ii) is well-known; see [2, Chap. 9, p. 137] or [26, Chap. 7, p. 85]. (iii): This is proven by combining the following results. $\bullet X^*$ has the Schur property $\Leftrightarrow X$ has the Dunford–Pettis property and X does not contain ℓ_1 [26, Exercise 4(ii) on p. 212]. \bullet every $C(K)$ space has the Dunford–Pettis property [26, Exercise 1(ii) on p. 113]. $\bullet C(K)$ does not contain $\ell_1 \Leftrightarrow K$ is scattered (Fact 3.7). □

4. On Directional Derivatives and (sub) Gradients

The results in this section make the proofs in the next section considerably easier; moreover, they are also of independent interest: for instance, the next theorem extends [41, Theorem 23.3] to infinite-dimensional spaces and sharpens results in [29, Sec. 14.C].

Theorem 4.1 (Dichotomy). *Suppose $\text{int dom } f \neq \emptyset$ and $x \in \text{dom } f$. Set $U = \text{int dom } f$. Then exactly one of the following two alternatives holds:*

- (i) $\partial f(x) = \emptyset$ and $f'(x; u - x) = -\infty$, for every $u \in U$.
- (ii) $\partial f(x) \neq \emptyset$, the function $y \mapsto f'(x; y - x)$ is continuous on U , and

$$f'(x; h) = \max\{\partial f(x), h\}, \quad \text{for every } h \in \text{cone}(U - x).$$

Proof. Case 1. $f'(x; u - x) = -\infty$, for every $u \in U$. We claim that $\partial f(x) = \emptyset$. Otherwise, we fix $\bar{u} \in \text{int dom } f$ and $x^* \in \partial f(x)$. Set $u = (1 - t)x + t\bar{u}$, for $t > 0$. Then $t\langle x^*, \bar{u} - x \rangle \leq f(x + t(\bar{u} - x)) - f(x)$. Divide by t and let t tend to 0 from above to deduce the absurdity $\langle x^*, \bar{u} - x \rangle \leq -\infty$. Hence we have $\partial f(x) = \emptyset$, and Case 1 is thus dealt with.

Case 2. $f'(x; u_0 - x) > -\infty$, for some $u_0 \in U$.

Consider the sequence of functions $(p_n)_{n \geq 1}$ defined by

$$p_n: X \rightarrow]-\infty, +\infty]: y \mapsto \frac{f(x + \frac{1}{n}(y - x)) - f(x)}{\frac{1}{n}}.$$

Clearly, each p_n is convex, lower semicontinuous, proper, and continuous on U . Let $p = X \rightarrow]-\infty, +\infty]: y \mapsto f'(x; y - x)$. Then $p(y) = \inf_n p_n(y) = \lim_n p_n(y)$,

$\forall y \in X$. Hence p is convex. Since $\text{dom } p \supseteq \text{dom } p_1$, we deduce that $U \subseteq \text{dom } p$. Also, $p(u_0) \in \mathbb{R}$. Fact 2.4 implies that p is proper. In particular, p is finite on U . Moreover, $p|_U$ is the pointwise limit of $(p_n|_U)$, a sequence of continuous functions. It follows (see, e.g. [15, Exercice IX.5.20(b)]) that the set of points where $p|_U$ is not continuous is meagre. Since U is a Baire space [15, Théorème IX.5.1 and Proposition IX.5.3], we conclude that the set of points where $p|_U$ is continuous is dense and hence nonempty. (This follows from [27, Theorem 3.1.7 on p. 109], too.) Consequently, by Fact 2.3 (or by [27, Theorem 3.1.8 on p. 110]), p is continuous on U . Equivalently, $f'(x; \cdot)$ is continuous on $U - x$. The result now follows from the positive homogeneity of $f'(x; \cdot)$ and Theorem 2.10. \square

Lemma 4.2. *Suppose $\{x, y\} \subseteq \text{dom } f$. Then $\lim_{t \rightarrow 0^+} f'(x + t(y - x); y - x) = f'(x; y - x)$.*

Proof. Let $g: \mathbb{R} \rightarrow]-\infty, +\infty]: t \mapsto f(x + t(y - x))$. Then g is convex, lower semicontinuous, proper on \mathbb{R} , $\text{dom } g \supseteq [0, 1]$, and $g'_+(t) = \lim_{h \rightarrow 0^+} \frac{g(t+h) - g(t)}{h} = f'(x + t(y - x); y - x)$, for all $t \in [0, 1[$. Hence

$$\begin{aligned} \lim_{t \rightarrow 0^+} f'(x + t(y - x); y - x) &= \lim_{t \rightarrow 0^+} g'_+(t) = g'_+(0) = \lim_{t \rightarrow 0^+} \frac{g(h) - g(0)}{h} \\ &= f'(x; y - x), \end{aligned}$$

where we have used [41, Theorem 24.1] to arrive at the second equality. \square

The next result involves the gradient map ∇f , which is always meant in the Gâteaux sense.

Lemma 4.3. *Suppose $\text{int dom } f \neq \emptyset$, $\text{dom } \nabla f$ is dense in $\text{dom } \partial f$, $x \in \text{int dom } f$, $h \in X$, and $\epsilon > 0$. Then there exists $y \in \text{int dom } f$ such that $\|y - x\| \leq \epsilon$, f is differentiable at y , and $|f'(x; h) - \langle \nabla f(y), h \rangle| \leq \epsilon$.*

Proof. By the classical Max Formula (Fact 2.0), choose $x^* \in \partial f(x)$ such that $\langle x^*, h \rangle = f'(x; h) \triangleq \alpha$. After decreasing ϵ if necessary, we may assume that $B(x; \epsilon) \subseteq \text{int dom } f$. By Fact 2.5, $\overline{\lim}_{\delta \rightarrow 0^+} f'(x + \delta h; h) \leq f'(x; h)$. Fix $\delta > 0$ sufficiently small so that

$$B(x + \delta h; \epsilon/2) \subseteq B(x; \epsilon) \quad \text{and} \quad f'(x + \delta h; h) < f'(x; h) + \epsilon.$$

Fix $y^* \in \partial f(x + \delta h)$. By monotonicity of ∂f , $0 \leq \langle y^* - x^*, (x + \delta h) - x \rangle$. Hence $\langle y^*, h \rangle \geq \langle x^*, h \rangle = f'(x; h) = \alpha$. Thus we obtain $\inf \langle \partial f(x + \delta h), h \rangle \geq \alpha$; equivalently, $-\alpha \geq \sup \langle \partial f(x + \delta h), -h \rangle = f'(x + \delta h; -h)$. By assumption, there exists a sequence (y_n) in $B(x + \delta h; \epsilon/2) \cap \text{dom } \nabla f$ such that $y_n \rightarrow x + \delta h$. The local boundedness of ∂f at $x + \delta h$ (Corollary 2.19) secures the boundedness of $(\nabla f(y_n))$. Passing to a subsequence if necessary, we assume that $(\langle \nabla f(y_n), h \rangle)$ converges to $\lambda \in \mathbb{R}$. Using

Fact 2.5, we obtain $-\alpha \geq f'(x + \delta h; -h) \geq \overline{\lim}_n f'(y_n; -h) = \overline{\lim}_n \langle \nabla f(y_n), -h \rangle = -\underline{\lim}_n \langle \nabla f(y_n), h \rangle = -\lambda$, so

$$\lambda \geq f'(x; h).$$

On the other hand, again by Fact 2.5, $\overline{\lim}_n f'(y_n; h) \leq f'(x + \delta h; h)$ and, therefore, $\lambda \leq f'(x + \delta h; h)$. Altogether,

$$f'(x; h) \leq \lambda = \lim_n \langle \nabla f(y_n), h \rangle \leq f'(x + \delta h; h) < f'(x; h) + \epsilon.$$

Hence y_n is as required, for all n sufficiently large. □

Lemma 4.4. *Suppose $\text{int dom } f \neq \emptyset$, $x_0 \in \text{dom } \partial f$, and $x_1 \in \text{int dom } f$. Then $\sup \|\partial f(\cdot|x_0, x_1)\| < +\infty$ and there exist $x_0^* \in \partial f(x_0)$ and bounded nets (x_α) in $]x_0, x_1] \subseteq \text{int dom } f$ and (x_α^*) in X^* , such that $x_\alpha^* \in \partial f(x_\alpha)$, $f'(x_\alpha; x_1 - x_0) = \langle x_\alpha^*, x_1 - x_0 \rangle$, $x_0 = \lim_\alpha x_\alpha$, $x_0^* = w^* \lim_\alpha x_\alpha^*$, $f'(x_0; x_1 - x_0) = \lim_\alpha f'(x_\alpha; x_1 - x_0)$, and $f(x_0) = \lim_\alpha f(x_\alpha)$.*

Furthermore, if $\text{dom } \nabla f$ is dense in $\text{dom } \partial f$, then there exist $y_0^ \in \partial f(x_0)$ and a bounded net (y_α) in $\text{int dom } f$ such that $(\nabla f(y_\alpha))$ is bounded, $x_0 = \lim_\alpha y_\alpha$, $y_0^* = w^* \lim_\alpha \nabla f(y_\alpha)$, $f'(x_0; x_1 - x_0) = \lim_\alpha \langle \nabla f(y_\alpha), x_1 - x_0 \rangle$, and $f(x_0) = \lim_\alpha f(y_\alpha)$.*

Proof. In view of Fact 2.3, obtain $\delta > 0$ such that $\eta = \sup f(B(x_1; \delta)) < +\infty$. Set $x_t = (1 - t)x_0 + tx_1$, for every $t \in]0, 1[$. Convexity of f yields

$$(\forall t \in]0, 1[)(\forall b \in B_X) f(x_t + t\delta b) \leq (1 - t)f(x_0) + t\eta.$$

Now fix $t \in]0, 1[$, $x_t^* \in \partial f(x_t)$, and $b \in B_X$. Then $t\delta \langle x_t^*, b \rangle = \langle x_t^*, (x_t + t\delta b) - x_t \rangle \leq f(x_t + t\delta b) - f(x_t) \leq (1 - t)f(x_0) + t\eta - f(x_t)$. Taking the supremum over $b \in B_X$, we conclude $t\delta \|x_t^*\| \leq (1 - t)f(x_0) + t\eta - f(x_t)$. Dividing by t and then employing Theorem 4.1 yields

$$\begin{aligned} \delta \|x_t^*\| &\leq \eta - f(x_0) - \frac{f(x_t) - f(x_0)}{t} \\ &\leq \eta - f(x_0) - f'(x_0; x_1 - x_0) < +\infty. \end{aligned} \tag{4.1}$$

Also, $f(x_0) \leq \underline{\lim}_{t \rightarrow 0^+} f(x_t) \leq \overline{\lim}_{t \rightarrow 0^+} f(x_t) \leq \overline{\lim}_{t \rightarrow 0^+} (1 - t)f(x_0) + tf(x_1) = f(x_0)$, so that

$$\lim_{t \rightarrow 0^+} f(x_t) = f(x_0).$$

Recall that $x_t \in \text{int dom } f$ (using Fact 2.0). By the classical Max Formula, we are able to pick $x_t^* \in \partial f(x_t)$ such that $f'(x_t; x_1 - x_0) = \langle x_t^*, x_1 - x_0 \rangle$, for all $0 < t \leq 1$. Lemma 4.2 implies $f'(x_0; x_1 - x_0) = \lim_{t \rightarrow 0^+} f'(x_t; x_1 - x_0)$. In view of (4.1), the net (x_t^*) is bounded. Hence we can extract suitable convergent subnets, i.e. $x_\alpha \rightarrow x_0$, $x_\alpha^* \xrightarrow{w^*} x_0^*$, for some $x_0^* \in \partial f(x_0)$.

“Furthermore”: We keep (x_t) and (x_t^*) as just found. For every $t \in]0, 1]$, there exists (Lemma 4.3) $y_t \in X$ such that $\|y_t - x_t\| \leq t\delta/2$ and $|f'(x_t, x_1 - x_0) - \langle \nabla f(y_t), x_1 - x_0 \rangle| \leq t$. Hence

$$\lim_{t \rightarrow 0^+} y_t = \lim_{t \rightarrow 0^+} x_t = x_0$$

and

$$\lim_{t \rightarrow 0^+} \langle \nabla f(y_t), x_1 - x_0 \rangle = \lim_{t \rightarrow 0^+} f'(x_t, x_1 - x_0) = f'(x_0; x_1 - x_0).$$

For every $t \in]0, 1]$, $y_t \in B(x_t; t)$ and therefore $f(y_t) \leq (1 - t)f(x_0) + t\eta$. Thus $f(x_0) \leq \underline{\lim}_{t \rightarrow 0^+} f(y_t) \leq \overline{\lim}_{t \rightarrow 0^+} f(y_t) \leq \overline{\lim}_{t \rightarrow 0^+} (1 - t)f(x_0) + t\eta = f(x_0)$. Thus

$$\lim_{t \rightarrow 0^+} f(y_t) = f(x_0).$$

We now tackle the boundedness of $(\nabla f(y_t))_{t \in]0, 1]}$. Write $y_t = x_t + (t\delta/2)b_t$, where $b_t \in B_X$, for all $t \in]0, 1]$. Fix $b \in B_X$ and set $z_t = y_t + (t\delta/2)b \in x_t + t\delta B_X$. Then

$$\begin{aligned} \frac{1}{2}t\delta \langle \nabla f(y_t), b \rangle &= \langle \nabla f(y_t), z_t - y_t \rangle \leq f(z_t) - f(y_t) \\ &\leq (1 - t)f(x_0) + t\eta - f(y_t) \\ &\leq t(\eta - f(x_0)) - (f(x_t) - f(x_0)) - (f(y_t) - f(x_t)). \end{aligned}$$

Dividing by t and taking the supremum over $b \in B_X$ results in

$$\begin{aligned} \frac{1}{2}\delta \|\nabla f(y_t)\| &\leq \eta - f(x_0) - \frac{f(x_t) - f(x_0)}{t} - \frac{f(y_t) - f(x_t)}{t} \\ &\leq \eta - f(x_0) - f'(x_0, x_1 - x_0) - \frac{f(y_t) - f(x_t)}{t}. \end{aligned}$$

On the other hand, $(t\delta/2)\langle x_t^*, b_t \rangle = \langle x_t^*, y_t - x_t \rangle \leq f(y_t) - f(x_t)$, so that

$$-\frac{f(y_t) - f(x_t)}{t} \leq \frac{1}{2}\delta \langle x_t^*, -b_t \rangle \leq \frac{1}{2}\delta \|x_t^*\|.$$

Using (*) to estimate $\|x_t^*\|$, we therefore conclude that

$$\begin{aligned} \delta \|\nabla f(y_t)\| &\leq 2\eta - 2f(x_0) - 2f'(x_0; x_1 - x_0) + \delta \|x_t^*\| \\ &\leq 3(\eta - f(x_0) - f'(x_0; x_1 - x_0)) \\ &< +\infty. \end{aligned}$$

We conclude by passing to a subnet (y_α) of (y_t) such that $(\nabla f(y_\alpha))$ is weak* convergent. □

We now derive a powerful subdifferential formula. Note that the assumption on denseness in the “Furthermore” part is always satisfied in weak Asplund spaces and thus in Euclidean spaces; see also Observation 4.10 and Observation 4.13 below.

Theorem 4.5 (Subdifferential Formula). *Suppose $\text{int dom } f \neq \emptyset$ and $x \in X$. Define a set $S(x)$ in X^* by requiring $x^* \in S(x)$ if and only if there exist bounded nets (x_α) in $\text{int dom } f$ and (x_α^*) in X^* such that for every α , $x_\alpha^* \in \partial f(x_\alpha)$, $x_\alpha \rightarrow x$, $x_\alpha^* \xrightarrow{w^*} x^*$, and $f(x_\alpha) \rightarrow f(x)$. Let $N(x) = N_{\text{dom } f}(x)$. Then:*

$$\partial f(x) = \text{cl}_{w^*}(N(x) + \text{cl}_{w^*}\text{conv } S(x)).$$

Furthermore, if $\text{dom } \nabla f$ is dense in $\text{dom } \partial f$, then define $G(x)$ by $y^ \in G(x)$ precisely when there exists a bounded net (y_α) in $\text{dom } \nabla f$ such that $(\nabla f(y_\alpha))$ is bounded, $y_\alpha \rightarrow x$, $\nabla f(y_\alpha) \xrightarrow{w^*} y^*$, $f(y_\alpha) \rightarrow f(x)$. In this case,*

$$\partial f(x) = \text{cl}_{w^*}(N(x) + \text{cl}_{w^*}\text{conv } G(x)).$$

Proof. Clearly, $S(x) \subseteq \partial f(x)$. For brevity, set $C = \text{dom } f$.

Case 1. $x \notin \text{dom } \partial f$.

Then $S(x) = \emptyset$ and the formula holds trivially.

Case 2. $x \in \text{int dom } f$.

Fix $x^* \in \partial f(x)$ and set $x_\alpha \equiv x$ and $x_\alpha^* \equiv x^*$. Then $\partial f(x) \subseteq S(x)$, so that $\partial f(x) = S(x) = \{0\} + \text{cl}_{w^*}\text{conv } S(x) = \text{cl}_{w^*}(N(x) + \text{cl}_{w^*}\text{conv } S(x))$, as announced.

Case 2 is isolated because it provides a very short proof when $\text{dom } f$ is open. In fact, the proof of Case 3 below requires only $x \in \text{dom } \partial f$ (and this is important when proving the ‘‘Furthermore’’ part).

Case 3. $x \in (\text{dom } \partial f) \setminus (\text{int dom } f)$.

Lemma 4.4 results in $S(x) \neq \emptyset$. Let p be the closure of $f'(x; \cdot)$, i.e. $\text{epi } p = \text{cl epi } f'(x; \cdot)$. Then p is lower semicontinuous, convex, positively homogeneous, and $p \leq f'(x; \cdot)$. Since $f'(x; \cdot)$ is continuous on $\text{int } C - x$ (Theorem 4.1), we have $p = f'(x; \cdot)$ on $\text{int } C - x$. By Fact 2.4, p is proper and so is $f'(x; \cdot)$. Hence Lemma 2.16 yields

$$p = \iota_{\partial f(x)}^*|_X = \sup\langle \partial f(x), \cdot \rangle.$$

Set

$$q = \sup\langle N(x) + S(x), \cdot \rangle = \iota_{N(x)+S(x)}^*|_X.$$

We always have $N(x) + S(x) \subseteq N(x) + \partial f(x)$. Lemma 2.20(i) yields $N(x) + \partial f(x) = \partial f(x)$. Altogether, this implies

$$q \leq p.$$

We now show that $p(h) \leq q(h)$, for every $h \in X$, by discussing cases.

Case (i). $h \notin T_C(x)$.

Since $T_C(x)$ is the negative polar cone of $N_C(x)$ intersected with X , we obtain $\sup\langle N_C(x), h \rangle = +\infty$. But $S(x)$ is nonempty and hence $p(h) \leq q(h) = +\infty$.

Case (ii). $h \in \text{int } T_C(x)$.

By Fact 2.2(i) and positive homogeneity of both p and q , we may assume that $h = x_1 - x_0$, where $x_0 = x$ and $x_1 \in \text{int } C$. Obtain nets (x_α) , (x_α^*) , and x_0^* as

in Lemma 4.4. Hence $x_0^* \in S(x) \subseteq N(x) + S(x)$, so that $\langle x_0^*, h \rangle \leq \sup \langle N(x) + S(x), h \rangle = q(h)$. Thus

$$p(h) \leq f'(x; h) = \lim_{\alpha} f'(x_{\alpha}; h) = \lim_{\alpha} \langle x_{\alpha}^*, x_1 - x_0 \rangle = \langle x_0^*, h \rangle \leq q(h).$$

Case (iii). $h \in \text{bdry } T_C(x)$.

Fix $k \in \text{int } T_C(x)$. By Fact 2.2(ii), $h + tk \in \text{int } T_C(x)$, for all $t > 0$. Now the already verified Case (ii) results in $p(h + tk) \leq q(h + tk) \leq q(h) + tq(k)$, for all $t > 0$. Thus $p(h) \leq \underline{\lim}_{t \rightarrow 0^+} p(h + tk) \leq \underline{\lim}_{t \rightarrow 0^+} q(h) + tq(k) = q(h)$.

Altogether,

$$\iota_{\partial f(x)}^*|_X = p = q = \iota_{N(x)+S(x)}^*|_X.$$

By Lemma 2.15, $\partial f(x) = \text{cl}_{w^*} \text{conv } \partial f(x) = \text{cl}_{w^*} \text{conv}(N(x) + S(x))$. It is not hard to see that the last set equals $\text{cl}_{w^*}(N(x) + \text{cl}_{w^*} \text{conv } S(x))$. The proof of the main conclusion is complete.

The ‘‘Furthermore’’ part follows exactly the same lines — the only difference is that we appeal to the ‘‘Furthermore’’ part of Lemma 4.4. □

4.1. Sharper versions in Banach spaces with additional structure

The results proved in this section hold in general Banach spaces. The spaces encountered in applications, however, possess additional structure which sometimes allows us to give precise answers to the following questions:

- When can we replace nets by sequences?
- When is $\text{dom } \nabla f$ dense in $\text{dom } f$?
- What can we say in finite dimensions?

Let us now review the notions required to answer these questions.

Remark 4.6 (Weak Asplund Spaces). Recall that X is a *weak Asplund space* if every continuous convex function defined on a convex nonempty open set is differentiable at each point of some dense G_{δ} subset of its domain [37]. It is known that X is a weak Asplund space if any of the following conditions holds:

- X is a quotient of a weak Asplund space [37, Theorem 4.24];
- X has a smooth renorm [37, Theorem 4.31];
- X is a subspace of a *weakly compactly generated space* (there is a weakly compact subset in the space with norm dense span) [37, Theorem 2.45];
- X is reflexive or separable (hence weakly compactly generated) [37, Example 2.42(a)].

Remark 4.7 (Gâteaux Differentiability Spaces). X is a *Gâteaux differentiability space*, if every continuous convex function defined on a convex nonempty open set is differentiable at each point of some dense subset of its domain. Clearly,

each weak Asplund space is a Gâteaux differentiability space.

It is unknown whether the class of weak Asplund spaces actually coincides with the class of Gâteaux differentiability spaces [24, Problem I.1 on p. 34].

The space ℓ_∞ is not a Gâteaux differentiability space and so not a weak Asplund space: the map $\ell_\infty \rightarrow \mathbb{R}: (x_n) \mapsto \overline{\lim}_n |x_n|$ is continuous but nowhere differentiable; see [37, Example 1.21].

It is known that if X is a Gâteaux differentiability space, then so is the closure of any continuous linear image of X : for this and further information, we refer the reader to [37, Sec. 6]. (In fact, corresponding dense single-valuedness results hold for maximal monotone operators [30] and USCOS [9]. See also [37, Sec. 7].)

Remark 4.8 (Weak* Sequential Compactness of the Dual Ball). A thorough discussion of this property can be found in Diestel's [26, Chap. XIII]. For our purpose, it is enough to state the following sufficient condition [26, Chap. XIII, Notes and Remarks, p. 239]:

if X is a weak Asplund space, then the dual ball B_{X^*} is weak* sequentially compact.

(In fact, the dual ball B_{X^*} is weak* sequentially compact whenever X is a Gâteaux differentiability space; see [32].)

We are now ready to formulate the announced sharpenings.

Observation 4.9. If X is a weak Asplund space, then “nets” can be replaced by “sequences” in Lemma 4.4 and Theorem 4.5.

Proof. Modify the proof of Lemma 4.4 as follows: instead of working with the net (x_t) , consider the sequence obtained by setting $t = 1/n$. Obtain the corresponding dual sequence. In view of Remark 4.5, we can extract a weak* convergent subsequence and then complete the proof as before. The sharpened version of Lemma 4.4 then results in the desired sharpening of Theorem 4.5. □

Observation 4.10. If X is a Gâteaux differentiability space, then $\text{dom } \nabla f$ is dense in $\text{dom } \partial f$. Consequently, the “Furthermore” parts of Lemma 4.4 and Theorem 4.5 apply.

Proof. This is clear from the definition of Gâteaux differentiability space; see Remark 4.5. □

Observation 4.11. Suppose $\text{dom } \nabla f$ is dense in $\text{dom } f$ and let D be fixed dense subset of $\text{dom } \nabla f$. Then the following sharpenings hold true:

- (i) In Lemma 4.3, the point y can be taken from D ;
- (ii) In Lemma 4.4, the net (y_α) can be taken from D ;
- (iii) In Theorem 4.5, the set $G(x)$ can be defined by requiring that the net (y_α) lie in D as well.

Proof. The original proofs work without change. □

Remark 4.12. It is interesting to note that Observation 4.11 is similar to (but significantly stronger than) the well-known “blindness of the Clarke subdifferential to small sets”; see [23, p. 93 in Sec. 2.8] and [22, Theorem 2.5.1].

We conclude this section by discussing the finite-dimensional setting:

Observation 4.13. Suppose X is finite-dimensional. Then X is a weak Asplund space (Remark 4.5) and hence a Gâteaux differentiability space (Remark 4.5). It follows from Observation 4.10 that the “Furthermore” part of Theorem 4.5 applies. Moreover, by employing a recession argument and Carathéodory’s Theorem, we are able to “peel off” the outermost weak* closure. We skip the details, however, since the resulting formula

$$\partial f(x) = N(x) + \text{cl conv } G(x),$$

is well-known and due to Rockafellar [41, Theorem 25.6].

5. Legendre Functions: Basic Properties

We begin with

Lemma 5.1. (i) ∂f is single-valued on its domain $\Leftrightarrow f^*$ is strictly convex on line segments in $\text{ran } \partial f$.

(ii) $(\forall (x, y) \in X^2) x \neq y \Rightarrow \partial f(x) \cap \partial f(y) = \emptyset \Leftrightarrow f$ is strictly convex on line segments in $\text{dom } \partial f$.

Proof. (i) \Rightarrow : By contradiction. Thus we assume there exist y_1^*, y_2^* in $\text{ran } \partial f$ such that $y_1^* \neq y_2^*$, $[y_1^*, y_2^*] \subseteq \text{ran } \partial f$, and $\{\lambda_1, \lambda_2\} \subseteq]0, 1[$ with $\lambda_1 + \lambda_2 = 1$ and $f^*(\lambda_1 y_1^* + \lambda_2 y_2^*) = \lambda_1 f^*(y_1^*) + \lambda_2 f^*(y_2^*)$. Now let $y^* = \lambda_1 y_1^* + \lambda_2 y_2^*$. Then there exists $x \in X$ such that $y^* \in \partial f(x)$. Hence

$$0 = f(x) + f^*(y^*) - \langle y^*, x \rangle = \sum_{i=1}^2 \lambda_i (f(x) + f^*(y_i^*) - \langle y_i^*, x \rangle) \geq 0.$$

It follows that both y_1^* and y_2^* belong to $\partial f(x)$, which is absurd.

\Leftarrow : Now pick y_1^* and $y_2^* \in \partial f(x)$. Then $f(x) + f^*(y_i^*) = \langle y_i^*, x \rangle$, for $i = 1, 2$. For all nonnegative reals λ_1, λ_2 that add up to 1, we have:

$$\begin{aligned} f(x) + \lambda_1 f^*(y_1^*) + \lambda_2 f^*(y_2^*) &= \langle \lambda_1 y_1^* + \lambda_2 y_2^*, x \rangle \\ &\leq f(x) + f^*(\lambda_1 y_1^* + \lambda_2 y_2^*) \\ &\leq f(x) + \lambda_1 f^*(y_1^*) + \lambda_2 f^*(y_2^*). \end{aligned}$$

Hence equality holds throughout. It follows that $x \in \partial f^*([y_1^*, y_2^*])$ and that $f^*|_{[y_1^*, y_2^*]}$ is affine. Consequently, $y_1^* = y_2^*$.

(ii) is proved analogously. □

Definition 5.2. We say that f is

- (i) *essentially smooth*, if ∂f is both locally bounded and single-valued on its domain.
- (ii) *essentially strictly convex*, if $(\partial f)^{-1}$ is locally bounded on its domain and f is strictly convex on every convex subset of $\text{dom } \partial f$.
- (iii) *Legendre*, if it is both essentially smooth and essentially strictly convex.

Remark 5.3. In Euclidean spaces, these notions agree with their well-established classical counterparts; see the upcoming Theorem 5.11. However, as Example 5.11 will show, a strictly convex function may fail to be essentially strictly convex.

Theorem 5.4 (Duality). *Suppose X is reflexive. Then f is essentially smooth if and only if f^* is essentially strictly convex.*

Proof. Use the fact that $(\partial f)^{-1} = \partial f^*$ in reflexive spaces, and Lemma 5.1. □

Corollary 5.5 (Legendre Duality). *Suppose X is reflexive. Then f is Legendre if and only if f^* is.*

Proof. Clear from Theorem 5.4. □

Theorem 5.6 (Essential Smoothness). *The following are equivalent*

- (i) f is essentially smooth.
- (ii) $\text{int dom } f \neq \emptyset$ and ∂f is single-valued on its domain.
- (iii) $\text{dom } \partial f = \text{int dom } f \neq \emptyset$ and ∂f is single-valued on its domain.
- (iv) $\text{int dom } f \neq \emptyset$, f is differentiable on $\text{int dom } f$, and $\lim_{t \rightarrow 0^+} f'(x + t(y - x)); y - x) = -\infty$, for every $x \in (\text{dom } f) \setminus (\text{int dom } f)$, $y \in \text{int dom } f$.
- (v) $\text{int dom } f \neq \emptyset$, f is differentiable on $\text{int dom } f$, and $\|\nabla f(x_n)\| \rightarrow +\infty$, for every sequence (x_n) in $\text{int dom } f$ converging to some point in $\text{bdry dom } f$.

Proof. (i) \Rightarrow (ii): By Fact 2.0, $\text{dom } \partial f \neq \emptyset$. Pick $x \in \text{dom } \partial f$. By assumption, ∂f is locally bounded at x . Hence, by Corollary 2.19, $x \in \text{int dom } f$. Thus $\text{int dom } f \neq \emptyset$.

(ii) \Rightarrow (iii): We always have $\text{int dom } f \subseteq \text{dom } \partial f$ (Fact 2.6). Fix $x \in \text{dom } \partial f$. By Lemma 2.20(ii), x cannot be a boundary point of $\text{dom } f$. Hence $x \in \text{int dom } f$ and thus $\text{dom } \partial f = \text{int dom } f$.

(i) \Leftarrow (iii): Corollary 2.19.

(ii) \Rightarrow (iv): By Fact 2.6, $\emptyset \neq \text{int dom } f \subseteq \text{dom } \partial f$. (ii) implies that $\partial f(x)$ is a singleton, $\forall x \in \text{int dom } f$. Altogether f is differentiable on $\text{int dom } f$ (see, for instance, [27, Theorem 4 on p. 122]). Now fix $x \in (\text{dom } f) \setminus (\text{int dom } f)$ and $y \in \text{int dom } f$. By the just established (iii), $\partial f(x) = \emptyset$. Theorem 4.1 yields $f'(x; y - x) = -\infty$. (iv) thus follows from Lemma 4.2.

(ii) \Leftarrow (iv): Pick $x \in \text{dom } \partial f$. It suffices to show that $\partial f(x)$ is a singleton.

We claim that $x \in \text{int dom } f$. Suppose to the contrary that the claim is false: $x \in (\text{dom } \partial f) \setminus (\text{int dom } f) \subseteq (\text{dom } f) \setminus (\text{int dom } f)$. Fix $y \in \text{int dom } f$. Then

$\lim_{t \rightarrow 0^+} f'(x + t(y - x); y - x) = -\infty$. By Lemma 4.2, $f'(x; y - x) = -\infty$. Theorem 4.1 implies that $\partial f(x) = \emptyset$, the desired contradiction. The claim is verified.

As f is differentiable at $x \in \text{int dom } f$, the subdifferential $\partial f(x) = \{\nabla f(x)\}$ must be a singleton [27, Theorem 4 on p. 122]. (ii) thus holds.

(iv) \Rightarrow (v): (iv) implies the differentiability of f on $\text{int dom } f \neq \emptyset$. Now let $x \in \text{bdry dom } f$ and (x_n) in $\text{int dom } f$ such that $x_n \rightarrow x$. We need to show that $\|\nabla f(x_n)\| \rightarrow +\infty$. Assume to the contrary that $\underline{\lim}_n \|\nabla f(x_n)\| < +\infty$. Pass to a subnet (x_α) of (x_n) such that $\nabla f(x_\alpha) \xrightarrow{w^*} x^*$. By maximal monotonicity of ∂f , we conclude $x^* \in \partial f(x)$. Hence $x \in (\text{dom } \partial f) \cap (\text{bdry dom } f)$. This contradicts (iii), as well as the equivalent (iv). Consequently, (v) holds.

(ii) \Leftarrow (v): In view of Fact 2.6, it suffices to show that $\text{dom } \partial f \subseteq \text{int dom } f$. We prove this by assuming the opposite: select $x \in (\text{dom } \partial f) \setminus (\text{int dom } f) \subseteq \text{dom } f$. Pick $y \in \text{int dom } f$. By Lemma 4.4, $K = \sup \|\nabla f(\cdot|x, y)\| < +\infty$. Set $x_n = (1 - \frac{1}{n})x + \frac{1}{n}y$, for all $n \geq 1$. Then $x_n \rightarrow x$ and $\|\nabla f(x_n)\| \leq K$, contradicting (v). The entire theorem is proven. \square

Remark 5.7 (Convex Integral Functions). There is a very natural construction that takes us inevitably out of the class of essentially smooth functions: convex integral functions. Suppose (S, Σ, μ) is a complete finite measure space (with nonzero μ), and $\phi: \mathbb{R} \rightarrow]-\infty, +\infty]$ is convex, lower semicontinuous, proper, with $\text{dom } \phi$ containing more than one point. The mapping

$$I_\phi: L_1(S, \Sigma, \mu) \rightarrow]-\infty, +\infty]: x \mapsto \int_S (\phi \circ x) d\mu$$

is well-defined and well-behaved [42]: (i) $(I_\phi)^* = I_{\phi^*}$ so that (ii) $y \in \partial(I_\phi)(x)$ if and only if $y \in L_\infty(S, \Sigma, \mu)$ and $y(s) \in \partial\phi(x(s))$, for almost every $s \in S$. Moreover, if ϕ^* is everywhere differentiable on \mathbb{R} , then, by [10, Theorem 3.8], (iii) I_ϕ is *strongly rotund*: it is strictly convex, has weakly compact lower level sets, and $x_n \rightarrow x$ whenever $x_n \rightharpoonup x$ and $I_\phi(x_n) \rightarrow I_\phi(x)$. The prime example is the following. Let $S = [0, 1]$ with Lebesgue measure, and set

$$\psi(r) = \begin{cases} +\infty, & \text{if } r < 0; \\ 0, & \text{if } r = 0; \\ r \ln(r) - r, & \text{if } r > 0. \end{cases}$$

Then $\psi^* = \exp$ and so $\text{dom } \psi^* = \mathbb{R}$. Now $\text{dom } \psi = [0, +\infty[$; consequently, $\text{dom } I_\psi$ equals $L_1^+[0, 1]$, the nonnegative cone in $L_1[0, 1]$. But this cone has empty interior! Thus I_ψ is nowhere continuous let alone differentiable. Despite this, (ii) shows that a point $x \in L_1[0, 1]$ belongs to $\text{dom } \partial I_\psi$ precisely when $x \in L_\infty^+[0, 1]$ and x is essentially bounded away from 0; if this is the case, then $\partial I_\psi(x)$ has a *unique* subgradient, namely $\ln(x)$. Incorporating such convex integral functions in our corpus represents a significant challenge.

We now turn to essentially strictly convex functions.

Lemma 5.8. *Suppose both $\text{dom } \partial f$ and $\text{dom } f^*$ are open. Then f is essentially strictly convex if and only if f is strictly convex on $\text{int dom } f$.*

Proof. By Fact 2.6, $\text{int dom } f^* \subseteq \text{dom } \partial f^*$. By openness of $\text{dom } f^*$ and Corollary 2.19, we deduce that ∂f^* is locally bounded on its domain. In particular, $(\partial f)^{-1}$ is locally bounded on its domain. Also, by Fact 2.6 and the assumption on $\text{dom } \partial f$, we observe that $\text{dom } \partial f = \text{int dom } f$, which is convex. The equivalence is now clear. □

Theorem 5.9 (Essential Strict Convexity). *Suppose X is reflexive and f is essentially strictly convex. Then:*

- (i) $(\forall (x, y) \in X^2) x \neq y \Rightarrow \partial f(x) \cap \partial f(y) = \emptyset$.
- (ii) $\text{ran } \partial f = \text{dom } \partial f^* = \text{int dom } f^* = \text{dom } \nabla f^*$.
- (iii) $(\forall y \in \text{dom } \partial f) \partial f^*(\partial f(y)) = \{y\}$.

Proof. (i) Clear from Lemma 5.1(ii). (ii) The first equality is trivial, the others follow with Theorem 5.4 and Theorem 5.6(iii) easy with (i). □

Theorem 5.10. *Suppose X is reflexive and f is Legendre. Then*

$$\nabla f: \text{int dom } f \rightarrow \text{int dom } f^*$$

is bijective, with inverse $(\nabla f)^{-1} = \nabla f^: \text{int dom } f^* \rightarrow \text{int dom } f$. Moreover, the gradient mappings $\nabla f, \nabla f^*$ are both norm-to-weak continuous and locally bounded on their respective domains.*

Proof. Since f is Legendre, it is both essentially smooth and essentially strictly convex. Hence f is differentiable on $\text{int dom } f \neq \emptyset$ (Theorem 5.6) and ∂f is a bijection between $\text{int dom } f$ and $\text{int dom } f^*$ (Theorem 5.9). It is known that ∂f is both norm-to-weak continuous [37, Proposition 2.8] and locally bounded on its domain [37, Theorem 2.28]. Now apply Corollary 5.4 to produce the dual statement regarding f^* . □

We now show, as previously announced, the compatibility of our new notions with their classical counterparts as defined in [41, Sec. 26]:

Theorem 5.11 (Compatibility). *Suppose X is a Euclidean space. Then*

- (i) *f is essentially smooth if and only if f is essentially smooth in the classical sense: f is differentiable on $\text{int dom } f \neq \emptyset$, and $\|\nabla f(x_n)\| \rightarrow +\infty$, for every sequence (x_n) in $\text{int dom } f$ converging to some point in $\text{bdry dom } f$.*
- (ii) *f is essentially strictly convex if and only if f is essentially strictly convex in the classical sense: f is strictly convex on every convex subset of $\text{dom } \partial f$.*
- (iii) *f is Legendre if and only if f is Legendre in the classical sense: f is both essentially smooth and essentially strictly convex in the classical sense.*

Proof. (i) follows from Theorem 5.6.

(ii) f is essentially strictly convex $\Leftrightarrow f^*$ is essentially smooth (Theorem 5.4) $\Leftrightarrow f^*$ is essentially smooth in the classical sense (by (i)) $\Leftrightarrow f$ is essentially strictly convex in the classical sense [41, Theorem 26.3].

(iii) clear from (i) and (ii). □

Remark 5.12. It is illuminating to consider the (sometimes subtle) difference between strict convexity and essential strict convexity with the help of the following three classical functions on \mathbb{R}^2 (see [41, Example before Theorem 23.5 and Examples before Theorem 26.3]): let

$$f_1(r, s) = \begin{cases} \max\{1 - r^{1/2}, |s|\}, & \text{if } r \geq 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $\text{dom } \partial f_1$ is not convex, and f_1 is not strictly convex on $\text{int dom } f$. Clearly, f_1 is not essentially strictly convex. Next, set

$$f_2(r, s) = \begin{cases} s^2/(2r) - 2s^{1/2}, & \text{if } r > 0 \text{ and } s \geq 0; \\ 0, & \text{if } r = s = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then f_2 is not strictly convex. However, $\text{dom } \partial f_2$ is convex, and f_2 is essentially strictly convex. Now define

$$f_3(r, s) = \begin{cases} s^2/(2r) + s^2, & \text{if } r > 0 \text{ and } s \geq 0; \\ 0, & \text{if } r = s = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $\text{dom } f_3 = \text{dom } \partial f_3$ is convex, f_3 is strictly convex on $\text{int dom } f_3$, but f_3 is not essentially strictly convex.

Finally, the function f_4 defined below is perhaps more borderline than any of the functions above:

$$f_4(r, s) = \begin{cases} \max\{(r - 2)^2 + s^2 - 1, -(rs)^{1/4}\}, & \text{if } r \geq 0 \text{ and } s \geq 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then f_4 is not strictly convex, $\text{dom } \partial f_4$ is not convex, yet f_4 is essentially strictly convex! (Note that the conjugates of f_1, \dots, f_4 are interesting with respect to essential smoothness.)

Remark 5.13. Another characterization of essential smoothness — *provided that X is Euclidean (or merely finite-dimensional)* — is this: f is essentially smooth $\Leftrightarrow f$ is differentiable on $\text{int dom } f \neq \emptyset$ and $\|\nabla f(x_n)\| \rightarrow +\infty$ whenever (x_n) is a bounded sequence in $\text{int dom } f$ with $d(x_n, \text{bdry dom } f) \rightarrow 0$. (This follows almost instantly from Theorem 5.11(i) and a compactness argument.) Moreover: (i) the boundedness of the sequence (x_n) in the new equivalent condition is important — consider the function $(r_1, r_2) \mapsto 1/(r_1 r_2)$ defined on the positive orthant in \mathbb{R}^2 .

(ii) the characterization fails in infinite-dimensional spaces; see [14, Example 2.7], which is based in c_0 .

Similar to Remark 5.13(ii), the last example in this section shows that the classical notions do differ from the new ones (outside finite-dimensional spaces):

Example 5.14 (Strictly Convex $\not\Rightarrow$ Essentially Strictly Convex). In $X = \ell_2$, let (p_n) be a sequence in $[2, +\infty[$ converging to $+\infty$. Define

$$f: X \rightarrow \mathbb{R}: x = (x_n) \mapsto \sum_n \frac{1}{p_n} |x_n|^{p_n}.$$

It is easy to check that f is everywhere differentiable and strictly convex. It is therefore Legendre in the classical sense. Hence, the function

f is essentially smooth.

Define the index conjugate to p_n through $\frac{1}{p_n} + \frac{1}{q_n} = 1$. Then $2 \geq q_n \rightarrow 1^+$, and $f^*(y) = \sum_n \frac{1}{q_n} |y_n|^{q_n}$.

We claim that $0 \notin \text{int dom } f^*$. Otherwise, obtain $\epsilon > 0$ such that $\epsilon B_{X^*} \subseteq \text{dom } f^*$. For $0 < \delta < \frac{1}{2}$ and $r > 0$ specified below, consider the sequence defined by $y_n = \frac{r}{n^{\frac{1}{2}+\delta}}$. Then $y = (y_n) \in \ell_2 = X^*$. Now choose $0 < r < 1$ small enough so that $\|y\| < \epsilon$. Since $1^+ \leftarrow q_n \leq 2$, we have $0 < n^{(\frac{1}{2}+\delta)q_n} < n$ eventually, say for $n \geq n_0$, and we obtain the absurdity:

$$+\infty > f^*(y) = \sum_n \frac{1}{q_n} \frac{r^{q_n}}{n^{(\frac{1}{2}+\delta)q_n}} \geq \sum_n \frac{1}{2} \frac{r^2}{n^{(\frac{1}{2}+\delta)q_n}} \geq \frac{r^2}{2} \sum_{n \geq n_0} \frac{1}{n} = +\infty.$$

The claim is thus proven. Since $\text{dom } f^*$ is symmetric, the Accessibility Lemma (Fact 2.0) implies $\text{int dom } f^* = \emptyset$. In particular, f^* is not essentially smooth in the classical sense. Moreover, $\partial f^*(y) = \{(\text{sign}(y_n)|y_n|^{q_n-1})\}$, provided this element lies in ℓ_2 . Consequently, ∂f^* is single-valued on its domain but not locally bounded (by Theorem 5.6). Thus the function

f is not essentially strictly convex.

The example thus shows that *the following three implications, which are always true in finite-dimensional spaces, each can fail in infinite dimensions:*

- “ f essentially strictly convex in the classical sense $\Rightarrow \text{int dom } f^* \neq \emptyset$ ”;
- “ ∂f^* is single-valued on its domain $\Rightarrow f^*$ is essentially smooth”.
- “ f is strictly convex $\Rightarrow f$ is essentially strictly convex (in our sense)”.

6. Legendre Functions: Further Examples

Example 6.1 (Spectral Functions). Suppose X is the real Hilbert space of $N \times N$ Hermitian matrices, with $\langle x, y \rangle = \text{trace}(xy)$, for all $x, y \in X$. Suppose $g: \mathbb{R}^J \rightarrow]-\infty, +\infty]$ is convex, lower semicontinuous, invariant under permutations,

and proper. Let $\lambda(x) \in \mathbb{R}^N$ denote the eigenvalues of $x \in X$ ordered decreasingly. Lewis [34] showed that

$$g \circ \lambda \text{ is Legendre if and only if } g \text{ is.}$$

(For extensions of this framework to compact operators, see [13] and [12].) This construction allows to build several interesting Legendre examples on X : for instance, the *log barrier* $x \mapsto -\ln \det x$ is a Legendre function on X (with the positive definite matrices as its domain) precisely because $-\ln$ is a Legendre function with domain $]0, +\infty[$.

Lemma 6.2. *Set $f = \frac{1}{p} \|\cdot\|^p$ for $1 < p < +\infty$. Let q be given by $\frac{1}{p} + \frac{1}{q} = 1$. Then $f^* = \frac{1}{q} \|\cdot\|^q$,*

$$\partial f(x) = \begin{cases} \|x\|^{p-2} Jx, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0, \end{cases}$$

and

$$\partial f^*(x^*) = \begin{cases} \|x^*\|^{q-2} J^* x^*, & \text{if } x^* \neq 0; \\ 0, & \text{if } x^* = 0. \end{cases}$$

Hence

- (i) X is smooth $\Leftrightarrow f$ is essentially smooth;
- (ii) X is rotund $\Leftrightarrow f$ is essentially strictly convex;
- (iii) X is smooth and rotund $\Leftrightarrow f$ is Legendre.

Proof. The formulae for the subdifferentials are immediate since $\partial \frac{1}{2} \|\cdot\|^2 = J$; see also [21, Sec. II.4].

- (i) X is smooth $\Leftrightarrow J$ is single-valued on X ([3, Proposition I.2.16 on p. 49]) $\Leftrightarrow \partial f$ is single-valued on $X \Leftrightarrow f$ is essentially smooth (Corollary 2.19).
- (ii) X is rotund $\Leftrightarrow \|\cdot\|^2$ is strictly convex ([3, Proposition I.2.13 on p. 43]) $\Leftrightarrow \frac{1}{p} \|\cdot\|^p$ is strictly convex (elementary) $\Leftrightarrow f$ is essentially strictly convex (Lemma 5.8).
- (iii) is clear from (i) and (ii). □

Example 6.3 (A Legendre Function with Bounded Closed Domain).

Suppose X is reflexive, smooth, and rotund so that $\frac{1}{2} \|\cdot\|^2$ is Legendre (Lemma 6.2). Define

$$f(x) = \begin{cases} -\sqrt{1 - \|x\|^2}, & \text{if } \|x\| \leq 1; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then f is strictly convex, $\text{dom } f = B_X$, and $f^*(x^*) = \sqrt{\|x^*\|^2 + 1}$. Moreover,

$$\nabla f(x) = \frac{Jx}{\sqrt{1 - \|x\|^2}} \quad \text{and} \quad \nabla f^*(x^*) = \frac{J^* x^*}{\sqrt{\|x^*\|^2 + 1}},$$

for every $x \in \text{dom } \nabla = \text{dom } \partial f = \text{int } B_X$, and every $x^* \in \text{dom } \nabla f^* = X^*$. It follows that f is Legendre with $\text{dom } f = B_X$.

Example 6.4 (A Legendre Function with Bounded Open Domain). Suppose X is reflexive, smooth, and rotund so that $\frac{1}{2}\|\cdot\|^2$ is Legendre (Lemma 6.2). Define

$$f(x) = \begin{cases} \frac{1}{1 - \|x\|^2}, & \text{if } \|x\| < 1; \\ +\infty, & \text{otherwise.} \end{cases}$$

Hence f is strictly convex, $\nabla f(x) = -(2Jx)/(1 - \|x\|^2)^2$, for every $x \in \text{dom } f = \text{dom } \nabla f = \text{int } B_X$, and f is essentially smooth (Theorem 5.6). Since $\text{dom } f \subseteq 1 \cdot B_X$, Rockafellar’s [39, Corollary 7.G and Remark on p. 62] implies that f^* is 1-Lipschitz on X^* . By Lemma 5.1(i) (applied to f^*), the function f^* is differentiable on the entire space X^* . Hence f^* is essentially smooth. By Theorem 5.4 (applied to f^*), the function f is essentially strictly convex. Altogether, f is Legendre with $\text{dom } f = \text{int } B_X$.

Example 6.5. Suppose X is uniformly rotund and uniformly smooth, and let $f = \|\cdot\|^s$, where $1 < s < +\infty$. Then f is Legendre, uniformly convex on closed balls, and totally convex.

Proof. It is well-known that X is both uniformly rotund and uniformly smooth, as is X^* . Lemma 6.2 yields that f is Legendre. By [45, Theorem 4.1(ii)], the function $f(x) = \|x\|^s = \int_0^{\|x\|} st^{s-1} dt$ is uniformly convex on closed balls, since $t \mapsto st^{s-1}$ is increasing (see also [44, Theorem 6] or [25, p. 54].) For total convexity, see [18]. \square

Example 6.6. In [38], Reich studies “the method of cyclic Bregman projections” in a reflexive Banach space X under the following assumptions: \bullet $\text{dom } f = \text{dom } \nabla f = X$ (hence f is essentially smooth by Theorem 5.6 and f^* is essentially strictly convex by Theorem 5.4); \bullet ∇f maps bounded sets to bounded sets, ∇f is uniformly continuous on bounded sets (hence f is Fréchet differentiable [37, Proposition 2.8], and f^* is supercoercive (Theorem 3.3)); \bullet f is uniformly convex (hence f is strictly convex on X).

These properties imply (see [33] and [44]) that $\bullet \lim_{\|x\| \rightarrow +\infty} \frac{f(x)}{\|x\|^2} > 0$ (hence f is supercoercive and so f^* is everywhere subdifferentiable and ∂f^* maps bounded sets to bounded sets). Altogether, f is Legendre and $\nabla f: X \rightarrow X^*$ is bijective and norm-to-norm continuous.

If X is reflexive and f is Legendre, then f^* is Legendre as well (Corollary 5.4). This is no longer true in general Banach spaces:

Example 6.7 (f Legendre $\not\Rightarrow f^*$ Legendre). $X = \ell_1$ is a weakly compactly generated space [25, Chap. 5, Sec. 2, p. 142] or [37, Examples 2.42]. Consequently, X admits an equivalent norm such that $(X, \|\cdot\|)$ is smooth and rotund, and

$(X^*, \|\cdot\|_*)$, where $\|\cdot\|_*$ denotes the dual norm of $\|\cdot\|$, is rotund [25, Chap. 5, Sec. 2, Corollary 2, p. 148]. Let $f = \frac{1}{2}\|\cdot\|^2$. Then, by Lemma 6.2,

f is Legendre and f^* is essentially strictly convex.

On the other hand, $X^* = \ell_\infty$ admits no smooth renorm [25, Chap. 4, Sec. 5, Proposition 2]; in particular, $(X^*, \|\cdot\|_*)$ is not smooth so that

f^* is not essentially smooth, hence f^* is not Legendre.

Finally, by James' theorem [25], $\text{int dom } f^* = \emptyset$.

Remark 6.8. In [14], Borwein and Vanderwerff discuss the construction of Legendre functions in terms of smoothness and rotundity of the underlying Banach space. For instance, they prove that in a weakly compactly generated space, every convex nonempty open subset is the domain of some Legendre function. In contrast, no Legendre function can exist on ℓ_∞/c_0 .

7. Legendre Functions are Zone Consistent

In this final section, we assume in addition that

$$\boxed{X \text{ is reflexive and } \text{int dom } f \neq \emptyset.}$$

Definition 7.1 (Bregman Distance). The *Bregman distance* corresponding to f is defined by

$$D = D_f: X \times \text{int dom } f \rightarrow [0, +\infty]: (x, y) \mapsto f(x) - f(y) + f'(y; y - x).$$

For more on Bregman distance and their fundamental importance in optimization and convex feasibility problems, see [16, 17, 20] and the references therein. We begin with a quite different example of a Legendre function:

Example 7.2 (Hilbert Space Projections). Suppose X is a Hilbert space, $\gamma > 0$, and

$$f(x) = \frac{1 + \gamma}{2}\|x\|^2 - \frac{1}{2}d^2(x, C),$$

where $d(x, C) = \min_{c \in C} \|x - c\| = \|x - Px\|$, P denotes the (orthogonal) projection map onto C , and $x \in X$. Then

$$\nabla f(x) = \gamma x + Px, D(x, y) = \frac{1}{2}(\gamma\|x - y\|^2 + \|x - Py\|^2 - \|x - Px\|^2),$$

and

$$f^*(y) = \frac{1}{2(1 + \gamma)}\|y\|^2 + \frac{1 + \gamma}{2\gamma}d^2\left(\frac{1}{1 + \gamma}y, C\right),$$

for all $x, y \in X$. Both f and f^* are supercoercive Legendre functions.

Proof. It is well-known (see, e.g. [37, Example 1.14(d)]) that $\frac{1}{2}\|\cdot\|^2 - \frac{1}{2}d^2(\cdot, C)$ is convex and Fréchet differentiable with gradient P . We thus readily obtain the formula for ∇f , and also conclude that f is strictly convex everywhere. The expression for the Bregman distance is a simple expansion. Now let $y = \nabla f(x) = \gamma x + Px$. Then $\frac{1}{1+\gamma}y$ is a convex combination of x and Px : $\frac{1}{1+\gamma}y = \frac{\gamma}{1+\gamma}x + \frac{1}{1+\gamma}Px$. It follows that $P(\frac{1}{1+\gamma}y) = Px$. Hence we can solve $y = \gamma x + Px = \gamma x + P(\frac{1}{1+\gamma}y)$ for x :

$$\nabla f^*(y) = x = \frac{1}{\gamma}y - \frac{1}{\gamma}P\left(\frac{1}{1+\gamma}y\right).$$

Thus $\text{dom } f^* = X$ is open and f is cofinite. By Lemma 5.8, f is a Legendre function. Hence f^* is a Legendre function, too (Corollary 5.4). In fact, since P is nonexpansive, the gradient mapping ∇f^* clearly maps bounded sets to bounded sets. Thus, by Theorem 3.3, f is supercoercive. The same argument shows that f^* is supercoercive. Integrating $\nabla f^*(y)$ with respect to y yields

$$f^*(y) = \frac{1}{2(1+\gamma)}\|y\|^2 + \frac{1+\gamma}{2\gamma}d^2\left(\frac{1}{1+\gamma}y, C\right) + k,$$

where k is constant that we shall determine from the equation $f(x) + f^*(\nabla f(x)) = \langle \nabla f(x), x \rangle$. Using the identity $d(\frac{1}{1+\gamma}y, C) = \frac{\gamma}{1+\gamma}d(x, C)$, we find $k = 0$. □

We next turn to basic properties of the Bregman distance.

Lemma 7.3. *Suppose $x \in X$ and $y \in \text{int dom } f$. Then*

- (i) $D(x, y) = f(x) - f(y) + \max\langle \partial f(y), y - x \rangle$.
- (ii) $D(\cdot, y)$ is convex, lower semicontinuous, proper with $\text{dom } D(\cdot, y) = \text{dom } f$.
- (iii) $D(x, y) = f(x) + f^*(y^*) - \langle y^*, x \rangle$, for every $y^* \in \partial f(y)$ with $\max\langle \partial f(y), y - x \rangle = \langle y^*, y - x \rangle$.
- (iv) If f is differentiable at y , then $D(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle = f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x \rangle$ and $\text{dom } \nabla D(\cdot, y) = \text{dom } \nabla f$.
- (v) If f is essentially strictly convex and differentiable at y , then $D(\cdot, y)$ is coercive.
- (vi) If f is essentially strictly convex, then: $D(x, y) = 0 \Leftrightarrow x = y$.
- (vii) If f is differentiable on $\text{int dom } f$ and essentially strictly convex, and $x \in \text{int dom } f$, then $D_f(x, y) = D_{f^*}(\nabla f(y), \nabla f(x))$.
- (viii) If f is supercoercive and $x \in \text{int dom } f$, then $D(x, \cdot)$ is coercive.
- (ix) If X is finite-dimensional, $\text{dom } f^*$ is open, and $x \in \text{int dom } f$, then $D(x, \cdot)$ is coercive.
- (x) If (y_n) is a sequence in $\text{int dom } f$ converging to y , then $D(y, y_n) \rightarrow 0$.

Proof. (i) See Fact 2.0.

(ii) Clear from (i).

(iii) Clear from Fact 2.12.

(iv) Follows from (i) and (iii).

(v) $\nabla f(y) \in \text{int dom } f^*$ by Theorem 5.9(ii). Fact 3.0 yields the coercivity of $f - \nabla f(y)$. Hence, by (iv), $D(\cdot, y)$ is coercive.

(vi) Pick y^* as in (iii) and assume $0 = D(x, y) = f(x) + f^*(y^*) - \langle y^*, x \rangle$. Then $x \in \partial f^*(y^*) \subseteq \partial f^*(\partial f(y)) = \{y\}$ (Theorem 5.9(iii)). The converse is trivial.

(vii) By using Theorem 5.9(ii), Theorem 5.9(iii), and item (iv) from above, we obtain the equalities $D_{f^*}(\nabla f(y), \nabla f(x)) = f^*(\nabla f(y)) + f(\nabla f^*(\nabla f(x))) - \langle \nabla f^*(\nabla f(x)), \nabla f(y) \rangle = f^*(\nabla f(y)) + f(x) - \langle x, \nabla f(y) \rangle = D_f(x, y)$.

(viii),(ix) Fix $x \in \text{int dom } f$ and let (y_n) be a sequence in $\text{int dom } f$ such that $(D(x, y_n))$ is bounded. Then it suffices to show that (y_n) is bounded.

Pick (see (iii)) $y_n^* \in \partial f(y_n)$ such that $D(x, y_n) = f(x) + f^*(y_n^*) - \langle y_n^*, x \rangle$, for every $n \geq 1$. Then (y_n^*) is bounded since $f^* - x$ is coercive by Fact 3.0. To prove (viii), note that supercoercivity of f implies (Theorem 3.3) that ∂f^* maps the bounded set $\{y_n^* : n \geq 1\}$ to a bounded set which contains $\{y_n : n \geq 1\}$. The coercivity of $D(x, \cdot)$ follows. It remains to prove (ix) in which X is finite-dimensional and $\text{dom } f^*$ is open. Assume to the contrary that (y_n) is unbounded. After passing to a subsequence if necessary, we may and do assume that $\|y_n\| \rightarrow +\infty$ and that (y_n^*) converges to point y^* . Then $(f^*(y_n^*)) = (D(x, y_n) - f(x) + \langle y_n^*, x \rangle)$ is bounded. Since f^* is lower semicontinuous, $y^* \in \text{dom } f^* = \text{int dom } f^*$. On the one hand, ∂f^* is locally bounded at y^* (Corollary 2.19). On the other hand, $y_n^* \rightarrow y^*$ and $y_n \in \partial f^*(y_n^*)$. Altogether, (y_n) is bounded — contradiction!

(x) By (iii), select $y_n^* \in \partial f(y_n)$ such that $D(y, y_n) = f(y) + f^*(y_n^*) - \langle y_n^*, y \rangle$, for every $n \geq 1$. The sequence (y_n^*) is bounded, since ∂f is locally bounded at y (Corollary 2.19). Assume to the contrary that $D(y, y_n) \not\rightarrow 0$. Again, after passing to a subsequence if necessary, we assume that there is some $\epsilon > 0$ such that $\epsilon \leq D(y, y_n) = f(y) + f^*(y_n^*) - \langle y_n^*, y \rangle$, for every n , and that (y_n^*) converges weakly to some $y^* \in \partial f(y)$. Since $y_n^* \in \partial f(y_n)$, Fact 2.12 and the assumption imply that $f^*(y_n^*) = \langle y_n^*, y_n \rangle - f(y_n) \rightarrow \langle y^*, y \rangle - f(y)$. Hence $f^*(y^*) \leq \underline{\lim}_n f^*(y_n^*) = \lim_n f^*(y_n^*) = \langle y^*, y \rangle - f(y) \leq f^*(y^*)$, which yields the absurdity $0 < \epsilon \leq D(y, y_n) = f(y) + f^*(y_n^*) - \langle y_n^*, y \rangle \rightarrow f(y) + f^*(y^*) - \langle y^*, y \rangle = 0$. \square

Remark 7.4. It is not possible to replace “ $y_n \rightarrow y$ ” in Lemma 7.3(x) by “ $y_n \rightharpoonup y$ ”: consider $f = \frac{1}{2}\|\cdot\|^2$ on $X = \ell_2$, let y_n denote the n th unit vector. Then $y_n \rightharpoonup 0$, but $D(0, y_n) = \frac{1}{2}\|0 - y_n\|^2 \equiv \frac{1}{2}$.

Example 7.5 (f Cofinite $\not\equiv f$ Supercoercive). Let $X = \ell_2$ and define (as in [4, Example 7.11]) $h(\mathbf{y}) = \sum_{n \geq 1} \frac{1}{2}y_n^{2n}$, for every $\mathbf{y} = (y_n) \in X^* = X$. Then h is strictly convex, proper, with $\text{dom } h = X^*$. Moreover, h is everywhere differentiable with $\nabla h(\mathbf{y}) = (ny_n^{2n-1})$. Now set $g = h + \frac{1}{2}\|\cdot\|^2$. Then g is strictly convex, proper, with $\text{dom } g = X^* = \text{int dom } g$, everywhere differentiable with $\nabla g = \nabla h + I$, and supercoercive. Since $\text{dom } \nabla g = X^*$, Corollary 2.19 yields that g is essentially smooth. Now let $f = g^*$. Then f is essentially strictly convex (Theorem 5.4), and $\text{dom } f = X$ (since $f = h^* \square \frac{1}{2}\|\cdot\|^2$ or by Theorem 3.4). The strict convexity of g together with Lemma 5.1 implies that ∂f is single-valued on its domain. Since

$X = \text{int dom } f \subseteq \text{dom } \partial f$ (Fact 2.6), f must be differentiable everywhere and hence (Corollary 2.19) f is essentially smooth. To sum up, by Corollary 5.4,

$$f \text{ is Legendre and cofinite with } \text{dom } f = \text{dom } \nabla f = X,$$

and

$$f^* \text{ is Legendre and supercoercive with } \text{dom } f^* = \text{dom } \nabla f^* = X^*.$$

Denote the standard unit vectors in X^* by \mathbf{e}_n and fix $\mathbf{x} \in X$ arbitrarily. Then

$$\mathbf{e}_n \rightarrow 0, \text{ but } \|\nabla f^*(\mathbf{e}_n)\| = n + 1 \rightarrow +\infty.$$

Now let $\mathbf{y}_n = \nabla f^*(\mathbf{e}_n) = (n + 1)\mathbf{e}_n$, for every $n \geq 1$. On the one hand, $\|\mathbf{y}_n\| \rightarrow +\infty$. On the other hand, by Lemma 7.3(iv), $D(\mathbf{x}, \mathbf{y}_n) = f(\mathbf{x}) + f^*(\nabla f(\mathbf{y}_n)) - \langle \nabla f(\mathbf{y}_n), \mathbf{x} \rangle = f(\mathbf{x}) + f^*(\mathbf{e}_n) - \langle \mathbf{e}_n, \mathbf{x} \rangle \leq f(\mathbf{x}) + g(\mathbf{e}_n) + \|\mathbf{e}_n\| \|\mathbf{x}\| \leq f(\mathbf{x}) + 1 + \|\mathbf{x}\|$. Altogether:

$$\text{there is no } \mathbf{x} \in X \text{ such that } D_f(\mathbf{x}, \cdot) \text{ is coercive.}$$

In view of Lemma 7.3(viii), f is not supercoercive.

Remark 7.6. It follows from the above example (together with Fact 3.0 and Theorem 3.4) that “ f is supercoercive” in Lemma 7.3(viii) cannot be replaced by “ f is cofinite”. Let us also observe that the existence of a cofinite, yet not supercoercive function is guaranteed by Theorem 3.6 (since ℓ_2 clearly does not have the Schur property). In Example 7.0, we have explicitly constructed such a function.

The following concept goes back to Bregman [16].

Definition 7.7 (Bregman Projection). Suppose C is a closed convex set in X . Given $y \in \text{int dom } f$, the set $P_C y = \{x \in C : D(x, y) = \inf_{c \in C} D(c, y)\}$ is called the *Bregman projection* of y onto C . Abusing notation slightly, we shall write $P_C y = x$, if $P_C y$ happens to be the singleton $P_C y = \{x\}$.

Theorem 7.8. *Suppose C is a closed convex set in X with $C \cap \text{dom } f \neq \emptyset$, and $y \in \text{int dom } f$. Then*

- (i) *If f is essentially strictly convex and differentiable at y , then $P_C y$ is nonempty and $P_C y \cap \text{int dom } f$ is at most a singleton.*
- (ii) *If f differentiable at y and strictly convex, then $P_C y$ is at most a singleton.*
- (iii) *If f is essentially smooth and $C \cap \text{int dom } f \neq \emptyset$, then $P_C y \subseteq \text{int dom } f$.*

Proof. (i) By Lemma 7.3(ii) and (v), $D(\cdot, y)$ is convex, lower semicontinuous, coercive, and $C \cap \text{dom } D(\cdot, y) \neq \emptyset$. Hence $P_C y = \text{arginf}_{x \in C} D(x, y) \neq \emptyset$. Since f and hence (Lemma 7.3(iii)) $D(\cdot, y)$ is strictly convex on $\text{int dom } f$, it follows that $P_C y \cap \text{int dom } f$ is at most a singleton.

(ii) By Lemma 7.3(iv), $D(x, y) = f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x \rangle$. Hence $D(\cdot, y)$ is strictly convex and the result follows.

(iii) Assume to the contrary that there exists $\bar{x} \in P_C y \cap (\text{dom } f \setminus (\text{int dom } f))$. Fix $c \in C \cap \text{int dom } f$ and define

$$\Phi: [0, 1] \rightarrow [0, +\infty[: t \mapsto D((1-t)\bar{x} + tc, y).$$

Then, using Lemma 7.3(ii), Φ is lower semicontinuous convex proper and $\Phi'(t) = \langle \nabla f(\bar{x} + t(c - \bar{x})), c - \bar{x} \rangle - \langle \nabla f(y), c - \bar{x} \rangle$, for all $0 < t < 1$. By Theorem 5.6, $\lim_{t \rightarrow 0^+} \Phi'(t) = -\infty$. This implies $\Phi(t) < \Phi(0)$, for all $t > 0$ sufficiently small (since $\Phi'(t)(0-t) \leq \Phi(0) - \Phi(t)$, i.e. $\Phi(t) \leq \Phi(0) + t\Phi'(t)$, for every $0 < t < 1$). It follows that for such t , $(1-t)\bar{x} + tc \in C \cap \text{int dom } f$ and $D((1-t)\bar{x} + tc, y) < D(\bar{x}, y)$, which contradicts $\bar{x} \in P_C y$. The entire theorem is proven. \square

In the terminology of Censor and Lent [19], the next result states that *every Legendre function is zone consistent*. This result is of crucial importance, since — as explained in the Introduction and carried out in Euclidean spaces in [5] — it makes the sequence generated by the method of cyclic Bregman projections well-defined under reasonable constraint qualifications. A detailed study of the central role played by Legendreness in the design and the analysis of this and various other algorithms in Banach spaces will appear in [6].

Corollary 7.9 (Legendre Functions are Zone Consistent). *Suppose f is a Legendre function, C is a closed convex set in X with $C \cap \text{int dom } f \neq \emptyset$, and $y \in \text{int dom } f$. Then:*

$$P_{C y} \text{ is a singleton and is contained in } \text{int dom } f.$$

Proof. Immediate from Theorem 7.8(i) and (iii). \square

Remark 7.10. Theorem 7.8 generalizes results in [5, Sec. 3]. We would like to point out an infelicity in the statement (not in the proof) of [5, Theorem 3.12(i)]: f should be essentially strictly convex rather than essentially smooth.

Note added in proof. After acceptance of this paper, Prof. Lionel Thibault kindly informed us that Theorem 4.5 is closely related to Proposition 3.4 and its corollaries in A. Jofré and L. Thibault, *D-representation of subdifferentials of directionally Lipschitz functions*, Proc. Amer. Math. Soc. **110**(1) (1990) 117–123 and to Theorem 3.3 in L. Thibault and D. Zagrodny, *Integration of subdifferentials of lower semicontinuous functions on Banach Spaces*, J. Math. Anal. Appl. **189**(1) (1995) 33–58. The proof methods are quite different.

References

- [1] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, John Wiley & Sons Inc., New York (1984).
- [2] S. Banach, *Théorie des Opérations Linéaires*, second edition, Chelsea, New York (1978). English translation: *Theory of Linear Operations*, North-Holland Publishing Co., Amsterdam (1987).

- [3] V. Barbu and Th. Precupanu, *Convexity and Optimization in Banach Spaces*, D. Reidel Publishing Co., second edition, Dordrecht (1986).
- [4] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev. **38**(3) (1996) 367–426.
- [5] H. H. Bauschke and J. M. Borwein, *Legendre functions and the method of random Bregman projections*, *J. Convex Anal.* **4**(1) (1997) 27–67.
- [6] H. H. Bauschke, J. M. Borwein and P. L. Combettes, *Bregman monotone optimization algorithms*, CECM preprint 01:171. Available at www.cecm.sfu.ca/preprints.
- [7] J. Borwein, M. Fabian and J. Vanderwerff, *Characterizations of Banach spaces via convex and other locally Lipschitz functions*, *Acta Math. Vietnam.* **22**(1) (1997) 53–69.
- [8] J. M. Borwein, *Asplund spaces are “sequentially reflexive”*, 1991. Combinatorics and Optimization Research Report CORR 91–14, University of Waterloo, Canada.
- [9] J. M. Borwein, *Minimal CUSCOS and subgradients of Lipschitz functions*, in *Fixed Point Theory and Applications (Marseille, 1989)*, Longman Sci. Tech., Harlow (1991), pp. 57–81.
- [10] J. M. Borwein and A. S. Lewis, *Strong rotundity and optimization*, *SIAM J. Optim.* **4**(1) (1994) 146–158.
- [11] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, Springer-Verlag, New York (2000).
- [12] J. M. Borwein, A. S. Lewis and Q. J. Zhu, *Convex spectral functions of compact operators, Part II: Lower semicontinuity and rearrangement invariance*, in *Optimization and Related Topics*, eds. A. Rubinov and B. Glover, Kluwer (2001), pp. 179–196.
- [13] J. M. Borwein, J. Read, A. S. Lewis and Q. J. Zhu, *Convex spectral functions of compact operators*, *J. Nonlinear Convex Anal.* **1**(1) (2000) 17–35.
- [14] J. M. Borwein and J. D. Vanderwerff, *Convex functions of Legendre type in general Banach spaces*, 2000. Preprint 00:151, Centre for Experimental and Constructive Mathematics, Simon Fraser University, Canada. Available at www.cecm.sfu.ca/preprints, to appear in *J. Convex Anal.*
- [15] N. Bourbaki, *Topologie Générale, Chapitres 5 à 10*, Hermann, Paris (1974). English Translation: *General Topology, Chapters 5–10*, Springer-Verlag, New York (1989).
- [16] L. M. Bregman, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming*, *Ž. Vyčisl. Mat. i Mat. Fiz.* **7** (1967) 620–631. English Translation in U.S.S.R. Comput. Math. Math. Phys. **7**(3) (1967) 200–217.
- [17] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer (2000).
- [18] D. Butnariu, A. N. Iusem and E. Resmerita, *Total convexity for powers of the norm in uniformly convex Banach spaces*, *J. Convex Anal.* **7**(2) (2000) 319–334.
- [19] Y. Censor and A. Lent, *An iterative row-action method for interval convex programming*, *J. Optim. Theory Appl.* **34**(3) (1981) 321–353.
- [20] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York (1997).
- [21] I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Kluwer, Dordrecht (1990).
- [22] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition (1990).
- [23] F. H. Clarke, Y. S. Ledyev, R. J. Stern and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag (1998).

- [24] R. Deville, G. Godefroy and V. Zizler, *Smoothness and Renormings in Banach Spaces*, Longman Scientific & Technical, Harlow (1993).
- [25] J. Diestel, *Geometry of Banach Spaces — Selected Topics*, Springer-Verlag, Berlin (1975). Lecture Notes in Mathematics, Vol. 485.
- [26] J. Diestel, *Sequences and Series in Banach spaces*, Springer-Verlag, New York (1984).
- [27] J. R. Giles, *Convex Analysis with Application in the Differentiation of Convex Functions*, Pitman, Boston, MA (1982).
- [28] R. Haydon, *A counterexample to several questions about scattered compact spaces*, Bull. London Math. Soc. **22**(3) (1990) 261–268.
- [29] R. B. Holmes, *Geometric Functional Analysis and its Applications*, Springer-Verlag, New York (1975). Graduate Texts in Mathematics, No. 24.
- [30] P. S. Kenderov, *The set-valued monotone mappings are almost everywhere single-valued*, C. R. Acad. Bulgare Sci. **27** (1974) 1173–1175.
- [31] H. E. Lacey, *The Isometric Theory of Classical Banach Spaces*, Springer-Verlag, New York (1974). Die Grundlehren der mathematischen Wissenschaften, Band 208.
- [32] D. G. Larman and R. R. Phelps, *Gâteaux differentiability of convex functions on Banach spaces*, J. London Math. Soc. (2) **20**(1) (1979) 115–127.
- [33] E. S. Levitin and B. T. Poljak, *Convergence of minimizing sequences in conditional extremum problems*, Dokl. Akad. Nauk SSSR **168** (1966) 997–1000. English translation in Soviet Math. Dokl. **7** (1966) 764–767.
- [34] A. S. Lewis, *Convex analysis on the Hermitian matrices*, SIAM J. Optim. **6**(1) (1996) 164–177.
- [35] J.-J. Moreau, *Sur la fonction polaire d'une fonction semi-continue supérieurement*, C. R. Acad. Sci. Paris **258** (1964) 1128–1130.
- [36] J.-J. Moreau, *Fonctionnelles convexes*, Collège de France, Paris, 1966–1967. Séminaire sur les Equations aux Dérivées Partielles II.
- [37] R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, Springer-Verlag, Berlin, second edition (1993).
- [38] S. Reich, *A weak convergence theorem for the alternating method with Bregman distances*, in *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, Dekker, New York (1996), pp. 313–318.
- [39] R. T. Rockafellar, *Level sets and continuity of conjugate convex functions*, Trans. Amer. Math. Soc. **123** (1966) 46–63.
- [40] R. T. Rockafellar, *Local boundedness of nonlinear, monotone operators*, Michigan Math. J. **16** (1969) 397–407.
- [41] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ (1970). Princeton Mathematical Series, No. 28.
- [42] R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, PA (1974).
- [43] Stephen Simons, *Minimax and Monotonicity*, Springer-Verlag, Berlin (1998).
- [44] A. A. Vladimirov, Ju. E. Nesterov and Ju. N. Čekanov, *Uniformly convex functionals*, Vestnik Moskov. Univ. Ser. XV Vychisl. Mat. Kibernet. **3** (1978) 12–23.
- [45] C. Zălinescu, *On uniformly convex functions*, J. Math. Anal. Appl. **95**(2) (1983) 344–374.