Vietnam Journal of Mathematics **40**:2&3(2012) 181-199

Vietnam Journal
of
MATHEMATICS

The Moreau Envelope Function and Proximal Mapping with Respect to the Bregman Distances in Banach Spaces

Ying Ying Chen[1], Chao Kan[2] and Wen Song[1*]

[1]*School of Mathematical and Sciences, Harbin Normal University, Harbin, 150025, P. R. China*

[2]*Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, P. R. China*

Dedicated to Professor Phan Quoc Khanh on the occasion of his 65th birthday

Received December 11, 2011
Revised April 24, 2012

Abstract. In this paper, we explore some properties of the Moreau envelope function $e_\lambda f(x)$ of f and the associated proximal mapping $P_\lambda f(x)$ with respect to the Bregman distance induced by a convex function in Banach spaces. Precisely, we study the continuity, locally Lipschitz property and differentiability of the Moreau envelope function and the upper semicontinuity and single-valuedness of the proximal mapping in Banach spaces.

2000 Mathematics Subject Classification. 90C25, 90C48, 47H05, 41A65.

Key words. Bregman distance, Moreau envelope, proximal mapping, continuity, single-valuedness.

1. Introduction

Let X be a Banach space and let $f\colon X \to (-\infty, +\infty]$ be a extended real-valued function. The *Moreau envelope function* $e_\lambda f\colon X \to [-\infty, +\infty]$ and the associated

[*] Corresponding author, the research of the third author was partly supported by the National Natural Sciences Grant (No. 11071052)

proximal mapping $P_\lambda f\colon X \rightrightarrows X$ are defined respectively by

$$e_\lambda f(x) = \inf_w \left\{ f(w) + \frac{1}{2\lambda}\|x-w\|^2 \right\},$$

$$P_\lambda f(x) = \arg\min_w \left\{ f(w) + \frac{1}{2\lambda}\|x-w\|^2 \right\},$$

where λ is a positive parameter. The Moreau envelope function and the associated proximal mapping play important roles in optimization, nonlinear analysis and signal recovery both theoretically and computationally [2, 3, 9, 11, 12, 18, 21, 27, 29–31]. We know that the set of minimizers of f over X coincides with the set of minimizers of $e_\lambda f$ over X. We also know (see [30]) that, when $X = \mathbb{R}^n$ is the n-dimensional Euclidean space, $P_\lambda f$ is maximal monotone if and only if $f + \frac{1}{2\lambda}\|\cdot\|^2$ is convex, and $P_\lambda f$ is single-valued and continuous if and only if $e_\lambda f$ is convex and continuously differentiable whenever f is convex. Recently, Wang [31] proved, under some basic assumptions, that $P_\lambda f$ is single-valued everywhere (in this case f is referred as λ-*Chebyshev function*), if and only if $f + \frac{1}{2\lambda}\|\cdot\|^2$ is essentially strictly convex, if and only if $e_\lambda f$ is continuously differentiable when f is a proper lower semicontinuous (not necessarily convex) function on $\mathbb{R}^n$. For the differentiability of $e_\lambda f$ in Hilbert and Banach spaces in the case when f is convex or prox-regular, we refer to [3, 9, 27].

When the function f is an indicator function ι_C of a nonempty set C of X, we have $P_\lambda f = P_C$, and $e_\lambda f = \frac{1}{2\lambda} d_C^2$, where P_C is the metric projection operator on $C \subset X$, and d_C is the distance function to C. We recall that C is said to be Chebyshev if $P_C(x)$ is a singleton for each $x \in X$. The Chebyshev problem as posed by Klee [24] asks: "Is a Chebyshev set necessarily convex?". Both Bunt and Motzkin gave an affirmative answer to this problem for the Euclidean space $\mathbb{R}^n$ independently in 1934 and in 1935. In the past fifty years many works have been devoted to solve it in more general settings, but it is still unanswered completely even in the case of infinite dimensional Hilbert spaces [4, 10, 19, 20, 22, 24]. The convexity problem of Chebyshev sets in general Banach spaces was also studied extensively. A comprehensive survey is given in [4].

In 1976, Bregman [13] discovered an elegant and effective technique for the use of the function D_g (which is called Bregman distance now) instead of the usual distance function in the process of designing and analyzing feasibility and optimization algorithms. Then, the Bregman's techniques have been applied, in various ways, in order to design and analyze iterative algorithms for solving feasibility, optimization and variational inequality problems, computing fixed points of nonlinear mappings, etc., see [5–7, 17, 28]. Let $g\colon X \to (-\infty, +\infty]$ be a proper convex function with its domain dom$g$. The Bregman distance function $D_g\colon X \times X \to \mathbb{R}_+$ is defined by

$$D_g(y,x) = g(y) - g(x) - g'_+(x, y-x) \quad \text{for all } x, y \in \operatorname{dom} g,$$

where $g'_+(x,h)$ is a right hand side derivative of g at $x \in \mathrm{dom} g$ in the direction h, defined by

$$g'_+(x,h) = \lim_{t \to 0^+} \frac{g(x+th) - g(x)}{t}.$$

Let $C \subset \mathrm{dom}\, g$ be a nonempty set. The Bregman projection on C with respect to g, denoted by Π_C^g, is defined as the set of the solutions of the optimization problem $\min_{y \in C} D_g(y,x)$, i.e.,

$$\Pi_C^g(x) = \arg\min_{y \in C} D_g(y,x) \text{ for any } x \in \mathrm{dom}\, g.$$

C is said to be D-Chebyshev if $\Pi_C^g(x)$ is nonempty and single-valued for each $x \in \mathrm{dom}\, g$.

Bauschke et al. started in [8] to consider the convexity problem of D-Chebyshev sets in the Euclidean space $\mathbb{R}^n$. They proved that each D-Chebyshev subset of $\mathbb{R}^n$ is convex provided g is a function of Legendre-type and 1-coercive. Li, Song and Yao [25] presented certain sufficient conditions and equivalent conditions for the convexity of a D-Chebyshev subset in Banach spaces as well as some sufficient conditions ensuring the upper semicontinuity and the continuity of the Bregman projection operator Π_C^g.

The Moreau envelope function and the associated proximal mapping with respect to Bregman distance in $\mathbb{R}^n$, namely D-Moreau envelope function $e_\lambda f$ and D-proximal mapping $P_\lambda f$, are defined by

$$e_\lambda f(x) = \inf_w \left\{ f(w) + \frac{1}{\lambda} D_g(w,x) \right\},$$

$$P_\lambda f(x) = \arg\min_w \left\{ f(w) + \frac{1}{\lambda} D_g(w,x) \right\}.$$

Obviously, if $f = \iota_C$, then $P_\lambda \iota_C = \Pi_C^g$ for all $\lambda > 0$. Therefore, a set C is D-Chebyshev if and only if the proximal mapping $P_\lambda \iota_C(x)$ is nonempty and single-valued for each $x \in \mathrm{dom} g$. Bauschke et al. [6, 7] explored some properties of $P_\lambda f$ (where it is called D-prox operator), such as the effective domain, range, set of fixed points, single-valuedness and continuity as well as the differentiability of $e_\lambda f$ in the case when f is convex. Kan and Song [23] presented some sufficient conditions for the continuity and upper semicontinuity of $P_\lambda f$ and showed that $\lim_{\lambda \to 0} e_\lambda f(x) = f(x)$ in the case when f is proper lower semicontinuous (not necessarily convex) defined on $\mathbb{R}^n$. They also explored the Clarke regularity and differentiability of $e_\lambda f$ and established a relationship between the single-valuedness of $P_\lambda f$ and the convexity of $\lambda f + g$. In this paper, we shall extend some results of [23] from $\mathbb{R}^n$ to reflexive Banach spaces which include the corresponding results from [25] as a special case when $f = \iota_C$ is the indicator function of a closed set C.

2. Preliminaries

Let X be a Banach space with the topological dual space X^* and $g\colon X \to \overline{\mathbb{R}}$ be a proper convex function. We use $\operatorname{dom} g$ to denote the domain of g. Let $x \in \operatorname{dom} g$. The *subdifferential* of g at x is the convex set defined by

$$\partial g(x) := \{x^* \in X^* : \ g(x) + \langle x^*, y - x\rangle \le g(y) \quad \text{for each } y \in X\}.$$

The domain and the range of ∂g are denoted by $\operatorname{dom}(\partial g)$ and $\operatorname{range}(\partial g)$, respectively, which are defined by

$$\operatorname{dom}(\partial g) := \{x \in \operatorname{dom} g : \partial g(x) \neq \emptyset\}$$

and

$$\operatorname{range}(\partial g) := \{x^* \in X^* : x^* \in \partial g(x),\, x \in \operatorname{dom}(\partial g)\}.$$

The *conjugate function* of g is the function $g^* : X^* \to \overline{\mathbb{R}}$ defined by

$$g^*(x^*) := \sup\{\langle x^*, x\rangle - g(x) : \ x \in X\}.$$

Then, by [32, Theorem 2.4.2(iii)], the Young-Fenchel inequality holds

$$\langle x^*, x\rangle \le g(x) + g^*(x^*) \quad \text{for each pair } (x^*, x) \in X^* \times X,$$

and the equality holds if and only if $x^* \in \partial g(x)$, i.e.,

$$\langle x^*, x\rangle = g(x) + g^*(x^*) \Longleftrightarrow x^* \in \partial g(x) \quad \text{for each pair } (x^*, x) \in X^* \times X. \tag{1}$$

Recall that the Bregman distance with respect to g is defined by

$$D_g(y, x) := g(y) - g(x) - g'_+(x, y - x) \quad \text{for any } x, y \in \operatorname{dom} g.$$

According to [14], we define the modulus of total convexity at x by

$$\nu_g(x, t) := \inf\{D_g(y, x) : \ y \in \operatorname{dom} g, \|y - x\| = t\} \quad \text{for each } t \ge 0.$$

We recall the following definitions which were introduced in [17] and [6], respectively:

Definition 2.1. Let $g\colon X \to \mathbb{R}$ be a function and $x \in \operatorname{dom} g$. g is said to be

(a) totally convex at x if its modulus is positive on $(0, \infty)$, i.e., $\nu_g(x, t) > 0$ for each $t > 0$, and

(b) essentially strictly convex if $(\partial g)^{-1}$ is locally bounded on its domain and g is strictly convex on every convex subset of $\operatorname{dom}(\partial g)$.

Remark 2.2. By [28, Proposition 2.2], the function g is totally convex at $x \in \operatorname{dom} g$ if and only if for any sequence $\{y_n\} \subset \operatorname{dom} g$,

$$\lim_{n\to\infty} D_g(y_n, x) = 0 \Longrightarrow \lim_{n\to\infty} \|y_n - x\| = 0.$$

It was observed in [15, Theorem 2.14] that strict convexity, essentially strict convexity, and total convexity for a real-valued convex function on the Euclidean space $\mathbb{R}^n$ are equivalent.

The following notions of essentially smooth convex functions and Legendre convex functions were extensively studied in [6].

Definition 2.3. A function g is said to be

(a) essentially smooth if ∂g is both locally bounded and single-valued on its domain, and

(b) Legendre if g is both essentially strictly convex and essentially smooth.

Proposition 2.4. [6, Theorems 5.4 and 5.6] *The following assertions hold.*

(i) *The function g is essentially smooth if and only if* $\mathrm{dom}(\partial g) = \mathrm{int}(\mathrm{dom}\, g) \neq \emptyset$ *and ∂g is single-valued on its domain.*

(ii) *If X is reflexive, then g is essentially smooth if and only if g^* is essentially strictly convex.*

The following propositions will be frequently used in the sequel, see [1, Corollary 3.1, Corollary 3.2], [16, Proposition 3.4] and [11, Corollary 4.5.2] for correspondingly.

Proposition 2.5. *Suppose that $g\colon X \to \bar{\mathbb{R}}$ is a lower semicontinuous proper convex function which is Gâteaux differentiable* (*resp. Fréchet differentiable*) *on* $\mathrm{int}(\mathrm{dom}\, g)$. *Then g is continuous and its Gâteaux derivative ∇g is norm-weak* continuous* (*resp. continuous*) *on* $\mathrm{int}(\mathrm{dom}\, g)$.

Proposition 2.6. *Suppose that $g\colon X \to \overline{\mathbb{R}}$ is a lower semicontinuous proper convex function. Let $x \in \mathrm{dom}\, g$ and suppose that g is totally convex at x. Then* $\partial g(x) \subseteq \mathrm{int}(\mathrm{dom}\, g^*)$ *and g^* is Fréchet differentiable at each point $x^* \in \partial g(x)$.*

Proposition 2.7. *Let $\phi\colon X \to (-\infty, +\infty]$ be a function such that ϕ^{**} is proper.*

(i) *Suppose ϕ^* is Fréchet differentiable at all $x^* \in \mathrm{dom}(\partial\phi^*)$ and ϕ is lower semicontinuous. Then ϕ is convex;*

(ii) *Suppose ϕ^* is Gâteaux differentiable at all $x^* \in \mathrm{dom}(\partial\phi^*)$ and ϕ is sequentially weakly lower semicontinuous. Then ϕ is convex.*

3. Continuity of the Moreau envelope functions and proximal mappings

Let X, Y be Banach spaces, and let $g\colon X \to \overline{\mathbb{R}}$ be a lower semicontinuous proper convex function which is Gâteaux differentiable on $\mathrm{int}(\mathrm{dom}\, g)$ with its Gâteaux derivative denoted by ∇g. In this case, the Bregman distance D_g becomes

$D_g(y,x) = g(y) - g(x) - \langle \nabla g(x), y - x\rangle \quad \text{for each pair } (y,x) \in X \times \operatorname{int}(\operatorname{dom} g).$

Given $f\colon X \to (-\infty, +\infty]$ be a function. We always assume that $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. Associated to f, we now define the *D-Moreau envelope* function $e_\lambda f$ and the D-proximal mapping $P_\lambda f$.

Definition 3.1. For a proper function $f\colon X \to (-\infty, +\infty]$ and some index $\lambda \in (0, +\infty]$, the *D-Moreau envelope* function $e_\lambda f$ and the D-proximal mapping $P_\lambda f$ are defined respectively as

$$e_\lambda f(x) = \inf_w \left\{ f(w) + \frac{1}{\lambda} D_g(w,x) \right\},$$

$$P_\lambda f(x) = \arg\min_w \left\{ f(w) + \frac{1}{\lambda} D_g(w,x) \right\}.$$

From the definition, $P_\lambda f$ can be rewritten as

$$P_\lambda f(x) = \left\{ w \in \operatorname{dom} f \cap \operatorname{dom} g\colon f(w) + \frac{1}{\lambda} D_g(w,x) = \inf\Big(f + \frac{1}{\lambda} D_g(\cdot, x)\Big)(X) < +\infty \right\},$$

then we have $\operatorname{dom} P_\lambda f \subset \operatorname{int}(\operatorname{dom} g)$ and $\operatorname{range} P_\lambda f \subset \operatorname{dom} f \cap \operatorname{dom} g$.

Definition 3.2 (Prox-boundedness). A function $f\colon X \to (-\infty, +\infty]$ is prox-bounded if there exists $\lambda > 0$ such that $e_\lambda f(x) > -\infty$ for some $x \in X$. The supremum of the set of all such λ is the threshold λ_f of the prox-boundedness, i.e.,

$$\lambda_f = \sup\{\lambda > 0 : \text{ there is } x \in X \text{ such that } e_\lambda f(x) > -\infty\}.$$

We observe that if there exists $\bar\lambda$ such that $e_{\bar\lambda} f(\bar x) > -\infty$ for some $\bar x \in X$, we then have $e_\lambda f(\bar x) \geq e_{\bar\lambda} f(\bar x) > -\infty$ for all $\lambda \in (0, \bar\lambda)$. This implies that for any $\lambda \in (0, \lambda_f)$, there is $x \in X$ such that $e_\lambda f(x) > -\infty$. If f is bounded from below, i.e., $\inf f > \infty$, then f is prox-bounded with threshold $\lambda_f = \infty$.

Definition 3.3 (Level boundedness and inf-compact). A function $\phi\colon X \to \bar{\mathbb{R}}$ is (lower) level-bounded (resp. inf-compact) if for every $\alpha \in \mathbb{R}$ the set $\operatorname{lev}_{\leq\alpha}\phi := \{x \in X\colon \phi(x) \leq \alpha\}$ is bounded (possibly empty) (resp. nonempty and compact).

The following definition was introduced in [8] which is a pointwise version of a concept due to Rockafellar and Wets [30, Definition 1.16].

Definition 3.4 (Uniform level boundedness). A function $\varphi\colon X \times Y \to \bar{\mathbb{R}}$ is level-bounded with respect to w locally uniformly in x at $\bar x$ if for every $\alpha \in \mathbb{R}$, there exists $\delta > 0$ such that

$$\bigcup_{x \in B(\bar x, \delta)} \{w \in X\colon \varphi(w,x) \leq \alpha\}.$$

is bounded.

We recall that a convex function $g\colon X \to \mathbb{R}$ is said to be 1-coercive, or super-coercive (cf. [6]) if

$$\lim_{\|y\|\to\infty} \frac{g(y)}{\|y\|} = \infty.$$

It is easy to see (cf. [6]) that g is 1-coercive if and only if

$$\operatorname{int}(\operatorname{dom} g^*) = \operatorname{dom} g^* = X^*.$$

Theorem 3.5. *Suppose that $f\colon X \to (-\infty, +\infty]$ is a proper prox-bounded function and g is 1-coercive. Then, the function $\psi(w,x,\lambda) := f(w) + \frac{1}{\lambda}D_g(w,x)$ is level-bounded with respect to w locally uniformly in (x,λ) at every $(x,\lambda) \in \operatorname{int}(\operatorname{dom} g) \times (0,\lambda_f)$.*

Proof. Suppose the conclusion does not hold. Then there exist $(\bar{x},\bar{\lambda}) \in \operatorname{int}(\operatorname{dom} g) \times (0,\lambda_f)$ and some $\alpha \in \mathbb{R}$ such that for each n, we can find $x_n \in \operatorname{int}(\operatorname{dom} g)$, $\lambda_n \in (0,\lambda_f)$, and $w_n \in X$ with $x_n \to \bar{x}$, $\lambda_n \to \bar{\lambda}$ as $n \to \infty$, and $\|w_n\| \geq n$ such that $\psi(w_n,x_n,\lambda_n) \leq \alpha$, i.e.,

$$f(w_n) + \frac{1}{\lambda_n}D_g(w_n,x_n) \leq \alpha. \tag{2}$$

It is clear that $w_n \neq x_n$ for n large enough. Since $\lambda_n \to \bar{\lambda} < \lambda_f$ there is $\lambda_0 \in (0,\lambda_f)$ such that $\lambda_n \in (0,\lambda_0)$ for sufficiently large n.

By the definition of λ_f, for any $\lambda^{'} \in (\lambda_0,\lambda_f)$, there exist some $x' \in \operatorname{int}(\operatorname{dom} g)$ and $\beta \in \mathbb{R}$ such that

$$f(w_n) + \frac{1}{\lambda'}D_g(w_n,x') \geq e_{\lambda'}f(x') \geq \beta > -\infty. \tag{3}$$

Combining (2) and (3), we get

$$f(w_n) + \frac{1}{\lambda_n}D_g(w_n,x_n) - \left(f(w_n) + \frac{1}{\lambda_0}D_g(w_n,x_0)\right) \leq \alpha - \beta.$$

Hence, for sufficiently large n, we have

$$\frac{1}{\lambda_0}D_g(w_n,x_n) - \frac{1}{\lambda'}D_g(w_n,x_0) \leq \alpha - \beta.$$

That is

$$\begin{aligned}\left(\frac{1}{\lambda_0} - \frac{1}{\lambda'}\right) g(w_n) + \frac{1}{\lambda'}g(x_0) - \frac{1}{\lambda_0}g(x_n) + \frac{1}{\lambda'}\langle \nabla g(x_0), w_n - x_0\rangle \\ - \frac{1}{\lambda_0}\langle \nabla g(x_n), w_n - x_n\rangle \leq \alpha - \beta.\end{aligned} \tag{4}$$

By Proposition 2.5, g is continuous at $\bar{x}$ and ∇g is norm-weak* continuous at $\bar{x}$. It follows that $\{g(x_n)\}$ and $\{\nabla g(x_n)\}$ are bounded. Hence $\frac{g(x_0)}{\|w_n\|} \to 0$, $\frac{g(x_n)}{\|w_n\|} \to 0$, as $n \to \infty$, and both $\langle \nabla g(x_0), \frac{w_n - x_0}{\|w_n\|} \rangle$ and $\langle \nabla g(x_n), \frac{w_n - x_n}{\|w_n\|} \rangle$ are bounded. Divide by $\|w_n\|$ from both sides of (4), and then since g is 1-coercive and $\frac{1}{\lambda_0} - \frac{1}{\lambda'} > 0$, we get $+\infty \leq \alpha - \beta$, which is a contradiction. ∎

Corollary 3.6. *Suppose that $f\colon X \to (-\infty, +\infty]$ is proper prox-bounded and that g is 1-coercive. Then, for any $\lambda \in (0, \lambda_f)$, the function $\psi_\lambda(w, x) := \psi(w, x, \lambda)$ is level-bounded with respect to w locally uniformly in x at every $x \in \operatorname{int}(\operatorname{dom} g)$. Particularly, for any fixed $(x, \lambda) \in \operatorname{int}(\operatorname{dom} g) \times (0, \lambda_f)$, the function $\varphi(w) := \psi(w, x, \lambda)$ is level-bounded on* $\operatorname{dom} f \cap \operatorname{dom} g$.

In order to investigate the continuity of $e_\lambda f$ and the upper semicontinuity of the proximal mapping $P_\lambda f$, we need the following compactness type assumption.

Definition 3.7. Let $f\colon X \to (-\infty, +\infty]$ be a proper lower semicontinuous and prox-bounded function. f is said to be D-approximately compact (D-approximately weakly compact) if, for any $x \in \operatorname{int}(\operatorname{dom} g)$ and any $\lambda \in (0, \lambda_f)$, each minimizing sequence of $e_\lambda f(x)$ has a subsequence converging (weakly converging) to an element of $\operatorname{dom} f \cap \operatorname{dom} g$, i.e., for any sequence $\{w_n\}$ with

$$\lim_{n\to\infty} \left[f(w_n) + \frac{1}{\lambda} D_g(w_n, x) \right] = e_\lambda f(x),$$

there exist a subsequence $\{w_{n_k}\}$ of $\{w_n\}$ and $w \in \operatorname{dom} f \cap \operatorname{dom} g$ such that $w_{n_k} \to w$ ($w_{n_k} \overset{w}{\to} w$) as $k \to \infty$.

Proposition 3.8. (i) *Suppose that $f\colon X \to (-\infty, +\infty]$ is proper lower semicontinuous, prox-bounded and inf-compact. Then f is D-approximately compact;*

(ii) *Let X be a reflexive Banach space. Suppose that $f\colon X \to (-\infty, +\infty]$ is proper weakly lower semicontinuous and prox-bounded and g is 1-coercive. Then f is D-approximately weakly compact.*

Proof. Let $x \in \operatorname{int}(\operatorname{dom} g)$ and $\{w_n\} \subset \operatorname{dom} f \cap \operatorname{dom} g$ be such that

$$\lim_{n\to\infty} \left[f(w_n) + \frac{1}{\lambda} D_g(w_n, x) \right] = e_\lambda f(x).$$

Then, for any $\alpha > e_\lambda f(x)$, we have that

$$f(w_n) \leq f(w_n) + \frac{1}{\lambda} D_g(w_n, x) \leq \alpha$$

for sufficiently large n.

For the first part (i), since f is inf-compact, there exist a subsequence $\{w_{n_k}\}$ of $\{w_n\}$ and $w \in \operatorname{dom} f \cap \operatorname{dom} g$ such that $\{w_{n_k}\}$ converges to w. For the second part (ii), by Theorem 3.5, $\{w_n\}$ is bounded. Since X is reflexive, there exist a subsequence $\{w_{n_k}\}$ of $\{w_n\}$ and $w \in \operatorname{dom} f \cap \operatorname{dom} g$ such that $\{w_{n_k}\}$ converges weakly to w.

Hence, in both cases, we have

$$\lim_{k\to\infty} \langle \nabla g(x), w_{n_k} - x\rangle = \langle \nabla g(x), w - x\rangle.$$

By the lower semicontinuity (or weak lower semicontinuity) of f and g, we have that

$$f(w) \le \liminf_{k\to\infty} f(w_{n_k}), \quad g(w) \le \liminf_{k\to\infty} g(w_{n_k}).$$

It follows that

$$f(w) + \frac{1}{\lambda} D_g(w, x) \le \liminf_{k\to\infty} \left[f(w_{n_k}) + \frac{1}{\lambda} D_g(w_{n_k}, x)\right] = e_\lambda f(x).$$

Consequently,

$$f(w) + \frac{1}{\lambda} D_g(w, x) = e_\lambda f(x),$$

and $w \in \operatorname{dom} f \cap \operatorname{dom} g$, and so the assertion follows. ■

Theorem 3.9. *Suppose that one of the following conditions holds:*

(a) $f\colon X \to (-\infty, +\infty]$ *is proper lower semicontinuous, prox-bounded and D-approximately compact;*

(b) X *is a reflexive Banach space and f is proper weakly lower semicontinuous, prox-bounded, and g is 1-coercive.*

Then the proximal mapping $P_\lambda f$ is nonempty on $\operatorname{int}(\operatorname{dom} g)$.

Proof. Let $x \in \operatorname{int}(\operatorname{dom} g)$ and $\{w_n\} \subset \operatorname{dom} f \cap \operatorname{dom} g$ be such that

$$\lim_{n\to\infty} \left[f(w_n) + \frac{1}{\lambda} D_g(w_n, x)\right] = e_\lambda f(x).$$

By the definition of D-approximate (weak) compactness and the proof of Proposition 3.8, we know that there exist a subsequence $\{w_{n_k}\}$ of $\{w_n\}$ and $w \in \operatorname{dom} f \cap \operatorname{dom} g$ such that $\{w_{n_k}\}$ converges (weakly) to w and

$$f(w) + \frac{1}{\lambda} D_g(w, x) = e_\lambda f(x).$$

That is, $w \in P_\lambda f(x)$. ■

Theorem 3.10. *Suppose that g is Fréchet differentiable on* $\operatorname{int}(\operatorname{dom} g)$ *and 1-coercive and that one of the following conditions holds:*

(a) $f\colon X \to (-\infty, +\infty]$ *is proper lower semicontinuous, prox-bounded and D-approximately compact;*

(b) *X is a reflexive Banach space and f is proper weakly lower semicontinuous and prox-bounded.*

Then, for any fixed $\lambda \in (0, \lambda_f)$, $e_\lambda f$ is continuous on $\mathrm{int}(\mathrm{dom}\, g)$.

Proof. Let $x \in \mathrm{int}(\mathrm{dom}\, g)$ and $\{x_n\} \subset \mathrm{dom}\, g$ be such that $x_n \to x$. Without loss of generality, we may assume that $\{x_n\} \subset \mathrm{int}(\mathrm{dom}\, g)$. Hence, by Proposition 2.5, g is continuous and ∇g is norm-norm continuous at x. Consequently,

$$e_\lambda f(x_n) \le f(w) + \frac{1}{\lambda} D_g(w, x_n) \to f(w) + \frac{1}{\lambda} D_g(w, x) \text{ for each } w \in \mathrm{dom}\, f \cap \mathrm{dom}\, g.$$

It follows that

$$\limsup_{n\to\infty} e_\lambda f(x_n) \le e_\lambda f(x). \tag{5}$$

Observe that if

$$e_\lambda f(x) \le \liminf_{n\to\infty} e_\lambda f(x_n) \tag{6}$$

holds, then it together with (5) completes the proof of the theorem. Therefore, it remains to show (6). By Theorem 3.9, there exists $\{w_n\} \subset \mathrm{dom}\, f \cap \mathrm{dom}\, g$ such that

$$f(w_n) + \frac{1}{\lambda} D_g(w_n, x_n) = e_\lambda f(x_n) \quad \text{for each } n = 1, 2, \ldots. \tag{7}$$

For any fixed $w_0 \in \mathrm{dom}\, f \cap \mathrm{dom}\, g$, we have that, for each $n \in \mathbb{N}$,

$$e_\lambda f(x_n) \le f(w_0) + \frac{1}{\lambda} D_g(w_0, x_n) \to f(w_0) + \frac{1}{\lambda} D_g(w_0, x).$$

Hence, for sufficiently large $n \in N$, we have that

$$f(w_n) + \frac{1}{\lambda} D_g(w_n, x_n) \le f(w_0) + \frac{1}{\lambda} D_g(w_0, x) + 1.$$

It follows from Theorem 3.5 that $\{w_n\}$ is bounded. Note that, for each $n \in \mathbb{N}$,

$$f(w_n) + \frac{1}{\lambda} D_g(w_n, x) = f(w_n) + \frac{1}{\lambda} D_g(w_n, x_n)$$
$$+\frac{1}{\lambda} \left[g(x_n) - g(x) + \langle \nabla g(x_n) - \nabla g(x), w_n \rangle + \langle \nabla g(x), x \rangle - \langle \nabla g(x_n), x_n \rangle \right]. \tag{8}$$

Letting $n \to +\infty$ in (8) and using (7) and the fact that $g(x_n) \to g(x)$ and $\nabla g(x_n) \to \nabla g(x)$, we get that

$$e_\lambda f(x) \le \liminf_{n\to\infty} \left\{ f(w_n) + \frac{1}{\lambda} D_g(w_n, x) \right\} \le \liminf_{n\to\infty} e_\lambda f(x_n).$$

This completes the proof of (6). ■

Corollary 3.11. *Suppose that g is Fréchet differentiable on* $\mathrm{int}(\mathrm{dom}\, g)$ *and 1-coercive and that one of the two conditions in Theorem 3.10 holds. Then, for* *any fixed* $\lambda \in (0, \lambda_f)$ *and* $w_n \in P_\lambda f(x_n)$ *with* $x_n \to x \in \operatorname{int}(\operatorname{dom} g)$*, we have that* $\{w_n\}$ *is bounded.*

Proof. It is a direct result of Theorem 3.5 and Theorem 3.10. ■

We recall the following facts:

(a) By (1), we have

$$x \in \partial g^*(x^*) \Longleftrightarrow x^* \in \partial g(x) \quad \text{for each pair } (x, x^*) \in X \times X^*. \tag{9}$$

Taking into account the differentiability of g, it yields

$$x \in (\partial g^* \circ \nabla g)(x) \quad \text{for each } x \in \operatorname{int}(\operatorname{dom} g).$$

(b) By (a) and Proposition 2.4(i), if g is essentially smooth, then

$$\operatorname{range}(\partial g^*) = \operatorname{dom}(\partial g) = \operatorname{int}(\operatorname{dom} g). \tag{10}$$

(c) If g is essentially smooth and 1-coercive, then

$$\nabla g(\operatorname{int}(\operatorname{dom} g)) = \operatorname{dom}(\partial g^*) = X^*. \tag{11}$$

In fact, by (10) and the 1-coercivity assumption, one has that $\operatorname{dom}(\partial g^*) = \operatorname{dom} g^* = X^*$ and $\operatorname{range}(\partial g^*) = \operatorname{int}(\operatorname{dom} g)$. Thus (11) follows from (9).

(d) If both g and g^* are essentially smooth (e.g., if X is reflexive and g is Legendre), then $\nabla g : \operatorname{int}(\operatorname{dom} g) \to \operatorname{int}(\operatorname{dom} g^*)$ is a bijection satisfying

$$(\nabla g)^{-1} = \nabla g^*.$$

Theorem 3.12. *Let* X *be a reflexive Banach space. Suppose that* $f : X \to (-\infty, +\infty]$ *is proper prox-bounded and that* g *is 1-coercive and Legendre. Then, for any fixed* $\lambda \in (0, \lambda_f)$*,* $e_\lambda f$ *satisfies*

$$e_\lambda f = \left(\frac{1}{\lambda} g^* - \frac{1}{\lambda}(\lambda f + g)^*\right) \circ \nabla g, \tag{12}$$

and

$$(\lambda f + g)^* = -\lambda e_\lambda f \circ \nabla g^* + g^*, \tag{13}$$

and consequently $e_\lambda f \circ \nabla g^*$ *is locally Lipschitz on* X^**. If, in addition,* ∇g *is locally Lipschitz on* $\operatorname{int}(\operatorname{dom} g)$*, then* $e_\lambda f$ *is locally Lipschitz on* $\operatorname{int}(\operatorname{dom} g)$*.*

Proof. For any $x \in \operatorname{int}(\operatorname{dom} g)$, we have

$$\begin{aligned} e_\lambda f(x) &= \inf_w \left\{ f(w) + \frac{1}{\lambda} D_g(w, x) \right\} \\ &= \inf_w \left\{ f(w) + \frac{1}{\lambda}(g(w) - g(x) - \langle \nabla g(x), w - x \rangle) \right\} \end{aligned}$$

$$= \inf_w \left\{ f(w) + \frac{1}{\lambda} g(w) - \frac{1}{\lambda} \langle \nabla g(x), w \rangle \right\} + \frac{1}{\lambda} \left(\langle \nabla g(x), x \rangle - g(x) \right)$$
$$= -\frac{1}{\lambda} \sup_w \{ \langle \nabla g(x), w \rangle - (\lambda f + g)(w) \} + \frac{1}{\lambda} \left(\langle \nabla g(x), x \rangle - g(x) \right)$$
$$= -\frac{1}{\lambda} (\lambda f + g)^*(\nabla g(x)) + \frac{1}{\lambda} g^*(\nabla g(x)).$$

Since g is Legendre and 1-coercive, $\nabla g\colon \operatorname{int}(\operatorname{dom} g) \to \operatorname{int}(\operatorname{dom} g^*) = X^*$ is a topological isomorphism with $(\nabla g)^{-1} = \nabla g^*$. Hence for any $x^* \in X^*$, $\nabla g^*(x^*) \in \operatorname{int}(\operatorname{dom} g)$ and hence

$$e_\lambda f(\nabla g^*(x^*)) = -\frac{1}{\lambda} (\lambda f + g)^*(x^*) + \frac{1}{\lambda} g^*(x^*).$$

This means that

$$(\lambda f + g)^* = -\lambda e_\lambda f \circ \nabla g^* + g^*.$$

It is clear that g^* and $(\lambda f + g)^*$ are convex functions with domain X^*. Since f and g are proper, we have that $g^*(x^*), (\lambda f + g)^*(x^*) > -\infty$ for all x^*. Hence g^* and $(\lambda f + g)^*$ are lower semicontinuous proper convex functions and hence they are locally Lipchitz on X^*. It follows that $e_\lambda f \circ \nabla g^*$ is locally Lipschitz on X^*. Moreover, if ∇g is locally Lipschitz on $\operatorname{int}(\operatorname{dom} g)$, then, from (12), we have that $e_\lambda f$ is locally Lipschitz on $\operatorname{int}(\operatorname{dom} g)$. ■

Proposition 3.13. *Suppose that one of the two conditions in Theorem 3.9 holds. Then, for any fixed $x \in \operatorname{int}(\operatorname{dom} g)$, we have*

(i) $e_\lambda f(x)$ *is a monotone nonincreasing function in λ on $(0, \lambda_f)$ and* $\sup_{\lambda \in (0, \lambda_f)} e_\lambda f(x) \le f(x)$*;*

(ii) $D_g(P_\alpha f(x), x) \le D_g(P_\beta f(x), x)$ *for any $0 < \alpha < \beta < \lambda_f$.*

Proof. For any $0 < \alpha < \beta < \lambda_f$, $\frac{1}{\alpha} - \frac{1}{\beta} > 0$, by Theorem 3.9, there exist $w_\alpha \in P_\alpha f(x)$ and $w_\beta \in P_\beta f(x)$ such that

$$e_\alpha f(x) = f(w_\alpha) + \frac{1}{\alpha} D_g(w_\alpha, x) \le f(w_\beta) + \frac{1}{\alpha} D_g(w_\beta, x);$$
$$e_\beta f(x) = f(w_\beta) + \frac{1}{\beta} D_g(w_\beta, x) \le f(w_\alpha) + \frac{1}{\beta} D_g(w_\alpha, x).$$

Hence, we have

$$e_\alpha f(x) - e_\beta f(x) \ge \left(\frac{1}{\alpha} - \frac{1}{\beta} \right) D_g(w_\alpha, x) \ge 0.$$

This shows that $e_\lambda f(x)$ is monotone nonincreasing in λ on $(0, \lambda_f)$. For any fixed $\lambda \in (0, \lambda_f)$, it is obvious that $e_\lambda f(x) \le f(x)$. Hence $\sup_{\lambda \in (0, \lambda_f)} e_\lambda f(x) \le f(x)$.

On the other hand, we have

$$e_\alpha f(x) - e_\beta f(x) \le \left(\frac{1}{\alpha} - \frac{1}{\beta}\right) D_g(w_\beta, x),$$

and consequently

$$0 \le \left(\frac{1}{\alpha} - \frac{1}{\beta}\right) D_g(w_\alpha, x) \le e_\alpha f(x) - e_\beta f(x) \le \left(\frac{1}{\alpha} - \frac{1}{\beta}\right) D_g(w_\beta, x).$$

This implies that, for any fixed $x \in \operatorname{int}(\operatorname{dom} g)$,

$$D_g(w_\alpha, x) \le D_g(w_\beta, x) \text{ for any } w_\alpha \in P_\alpha f(x) \text{ and } w_\beta \in P_\beta f(x)$$

and hence

$$D_g(P_\alpha f(x), x) \le D_g(P_\beta f(x), x) \text{ for } \ 0 < \alpha < \beta < \lambda_f.$$

■

Theorem 3.14. *Suppose that g is totally convex and that one of the two conditions in Theorem 3.9 holds. Then*

$$\lim_{\lambda \to 0} e_\lambda f(x) = f(x) \quad \text{for all } x \in \operatorname{int}(\operatorname{dom} g).$$

Proof. Since $e_\lambda f(x)$ is nonincreasing in $\lambda \in (0, \lambda_f)$, for any fixed $x \in \operatorname{int}(\operatorname{dom} g)$ by Proposition 3.13, it is sufficient to show that there is some sequence $\{\lambda_n\} \subset (0, \lambda_f)$ with $\lambda_{n+1} < \lambda_n$ and $\lambda_n \to 0$ such that

$$\lim_{n \to \infty} e_{\lambda_n} f(x) = f(x).$$

Let $\{\lambda_n\}$ be such a sequence. By Theorem 3.9 and Proposition 3.13, there exist $w_n \in P_{\lambda_n} f(x)$ such that $D_g(w_{n+1}, x) \le D_g(w_n, x)$. This implies that the limit of a nonnegative sequence $\{D_g(w_n, x)\}$ exists. We denote by $d := \lim_{n \to \infty} D_g(w_n, x)$ and claim $d = 0$.

Since $w_n \in P_{\lambda_n} f(x)$, we have that

$$\begin{aligned} e_{\lambda_n} f(x) &= f(w_n) + \frac{1}{\lambda_n} D_g(w_n, x) \\ &\ge f(w_n) + \frac{1}{\lambda_n} d. \end{aligned}$$

Therefore, if $d > 0$, then

$$f(x) \ge \sup_{0 < \lambda < \lambda_f} e_\lambda f(x) \ge f(w_n) + d \sup_{n \ge 1} \frac{1}{\lambda_n} \ge +\infty,$$

and then $d = 0$.

Since g is totally convex, it follows from $\lim_{n\to\infty} D_g(w_n, x) = 0$ that $\|w_n - x\| \to 0$ as $n \to \infty$. By the lower semicontinuity of f, for any $\varepsilon > 0$, the following inequality holds for n sufficiently large:

$$f(x) - \varepsilon \le f(w_n) \le f(w_n) + \frac{1}{\lambda_n} D_g(w_n, x) = e_{\lambda_n} f(x) \le f(x).$$

Letting $n \to \infty$ in the above inequality, by the arbitrariness of ε, we obtain that

$$\lim_{\lambda\to 0} e_\lambda f(x) = f(x).$$

■

Definition 3.15. [26] Let Y, X be normed spaces. A set-valued mapping $F\colon Y \rightrightarrows X$ is said to be upper semicontinuous at a point $y_0 \in Y$, if for any open set $V \supset F(y_0)$ there exists $\delta > 0$ such that for every $y \in B(y_0, \delta)$, the inclusion $F(y) \subset V$ holds. F is said to be upper semicontinuous if it is upper semicontinuous at every $y_0 \in Y$.

Theorem 3.16. *Suppose that g is Fréchet differentiable on* $\operatorname{int}(\operatorname{dom} g)$ *and 1-coercive and that one of the two conditions in Theorem 3.10 holds. Then, for any $\lambda \in (0, \lambda_f)$, the set-valued mapping $P_\lambda f$ is upper semicontinuous on* $\operatorname{int}(\operatorname{dom} g)$.

Proof. Suppose the conclusion does not hold. Then there exist some $x_0 \in \operatorname{int}(\operatorname{dom} g)$ and an open set V with $V \supset P_\lambda f(x_0)$ and a sequence $w_n \in P_\lambda f(x_n)$ with $x_n \to x_0$, but $w_n \in X \setminus V$. By Corollary 3.11, we see that $\{w_n\}$ is bounded. Note that $e_\lambda f$ is continuous on $\operatorname{int}(\operatorname{dom} g)$, we have

$$\begin{aligned}
e_\lambda f(x_0) &= \lim_{n\to\infty} e_\lambda f(x_n)\\
&= \lim_{n\to\infty} \left\{ f(w_n) + \frac{1}{\lambda} D_g(w_n, x_n) \right\}\\
&= \lim_{n\to\infty} \Big\{ f(w_n) + \frac{1}{\lambda} D_g(w_n, x_0) + \frac{1}{\lambda}[g(x_0) - g(x_n) + \langle \nabla g(x_0)\\
&\quad - \nabla g(x_n), w_n\rangle - \langle \nabla g(x_0), x\rangle + \langle \nabla g(x_n), x_n\rangle] \Big\}\\
&= \lim_{n\to\infty} \{ f(w_n) + \frac{1}{\lambda} D_g(w_n, x_0)\}.
\end{aligned}$$

Hence, $\{w_n\}$ is a D-minimizing sequence of $e_\lambda f(x_0)$ and hence there exists a subsequence $\{w_{n_k}\}$ converging (or weakly converging) to some element w_0. Obviously, $w_0 \in P_\lambda f(x_0) \subset V$, but on the other hand, $w_0 \in X \setminus V$, since $X \setminus V$ is closed. Hence, we get a contradiction. ■

4. Characterization of single-valuedness of the proximal mapping

In this section, we assume that X is a reflexive Banach space and that $g\colon X \to \bar{\mathbb{R}}$ is a lower semicontinuous proper convex function which is Gâteaux differentiable on int(dom$g$) and 1-coercive, and that f is a proper function satisfying dom$f \cap$ dom$g \neq \emptyset$.

Theorem 4.1. *Suppose that f is a proper lower semicontinuous and prox-bounded function. Then the following inequality holds*

$$\langle P_\lambda f(y) - P_\lambda f(x), \nabla g(y) - \nabla g(x)\rangle \geq 0.$$

If, in addition, g is Legendre, then $P_\lambda f \circ \nabla g^*$ is monotone.

Proof. For any $x, y \in \operatorname{int}(\operatorname{dom} g)$ and any $u \in P_\lambda f(x)$, $v \in P_\lambda f(y)$, we have

$$f(u) + \frac{1}{\lambda} D_g(u, x) \leq f(v) + \frac{1}{\lambda} D_g(v, x);$$
$$f(v) + \frac{1}{\lambda} D_g(v, y) \leq f(u) + \frac{1}{\lambda} D_g(u, y),$$

that is

$$\lambda f(u) - \lambda f(v) + g(u) - g(v) + \langle \nabla g(x), v - u\rangle \leq 0;$$
$$\lambda f(v) - \lambda f(u) + g(v) - g(u) + \langle \nabla g(y), u - v\rangle \leq 0.$$

Adding the last two inequalities, we get

$$\langle \nabla g(x) - \nabla g(y), v - u\rangle \leq 0.$$

It follows that

$$\langle P_\lambda f(y) - P_\lambda f(x), \nabla g(y) - \nabla g(x)\rangle \geq 0.$$

If, in addition, g is Legendre, we have $(\nabla g)^{-1} = \nabla g^*$, and then

$$\langle (P_\lambda f \circ \nabla g^*)(y) - (P_\lambda f \circ \nabla g^*)(x), y - x\rangle \geq 0.$$

This shows that $P_\lambda f \circ \nabla g^*$ is monotone. ■

Corollary 4.2. *Suppose that g is Fréchet differentiable on* int(dom g) *and Legendre and that one of the two conditions in Theorem 3.10 holds. If $P_\lambda f$ is single valued, then $P_\lambda f$ is continuous and $P_\lambda f \circ \nabla g^*$ is maximal monotone.*

Proof. Since g is Legendre and 1-coercive, ∇g^* is continuous on X^*. It follows from Theorem 3.16 and Theorem 4.1 that $P_\lambda f \circ \nabla g^*$ is continuous and monotone on X, and consequently, maximal monotone. ■

Theorem 4.3. *Suppose that g is Fréchet differentiable on* int(dom g) *and Legendre and that one of the two conditions in Theorem 3.10 holds. Then, for any $\lambda \in (0, \lambda_f)$, $P_\lambda f \circ \nabla g^*$ is maximal monotone if and only if $P_\lambda f \circ \nabla g^* = \partial(\lambda f + g)^*$. Moreover, the above equivalent assertion holds if $\lambda f + g$ is convex.*

Proof. For every $\bar{x} \in \operatorname{int}(\operatorname{dom} g)$ and every $\bar{w} \in P_\lambda f(\bar{x})$, we have

$$f(\bar{w}) + \frac{1}{\lambda} D_g(\bar{w}, \bar{x}) \leq f(w) + \frac{1}{\lambda} D_g(w, \bar{x}) \text{ for any } w \in \ dom f \cap \operatorname{dom} g$$

which is equivalent to

$$(\lambda f + g)(w) \geq (\lambda f + g)(\bar{w}) + \langle \nabla g(\bar{x}), w - \bar{w} \rangle \text{ for any } w \in \operatorname{dom} f \cap \operatorname{dom} g.$$

This implies that

$$(\lambda f + g)^{**}(w) \geq (\lambda f + g)(\bar{w}) + \langle \nabla g(\bar{x}), w - \bar{w} \rangle \text{ for any } w \in \operatorname{dom} f \cap \operatorname{dom} g,$$

and that

$$(\lambda f + g)^{**}(\bar{w}) = (\lambda f + g)(\bar{w}),$$

and

$$\nabla g(\bar{x}) \in \partial (\lambda f + g)^{**}(\bar{w}). \tag{14}$$

It follows from (14) that

$$\bar{w} \in \partial (\lambda f + g)^* \circ \nabla g(\bar{x}).$$

Therefore

$$P_\lambda f(\bar{x}) \subset \partial (\lambda f + g)^* \circ \nabla g(\bar{x}).$$

This implies, since g is Legendre and 1-coercive, that

$$P_\lambda f \circ \nabla g^*(x^*) \subset \partial (\lambda f + g)^*(x^*) \ \text{ for all } \ x^* \in X^*. \tag{15}$$

It follows that

$$P_\lambda f \circ \nabla g^*(x^*) = \partial (\lambda f + g)^*(x^*) \text{ for all } x^* \in X^*,$$

if $P_\lambda f \circ \nabla g^*$ is maximal monotone, and the converse is obviously true.

Moreover, suppose that $\lambda f + g$ is convex. Then we have

$$\begin{aligned} \bar{w} \in P_\lambda f(x) &\Leftrightarrow (\lambda f + g)(\bar{w}) - \langle \nabla g(x), \bar{w} \rangle \leq (\lambda f + g)(w) - \langle \nabla g(x), w \rangle \text{ for all } w \in X^* \\ &\Leftrightarrow \nabla g(x) \in \partial (\lambda f + g)(\bar{w}) \\ &\Leftrightarrow \bar{w} \in \partial (\lambda f + g)^* \circ \nabla g(x), \end{aligned}$$

where the third equivalence holds because $\lambda f + g$ is proper convex lower semi-continuous. It follows that

$$P_\lambda f \circ \nabla g^*(x^*) = \partial (\lambda f + g)^*(x^*) \text{ for all } x^* \in X^*.$$

This implies that $P_\lambda f \circ \nabla g^*$ is maximal monotone. ∎

When $X = \mathbb{R}^n$, we proved in [23] that $P_\lambda f \circ \nabla g^*$ is maximal monotone if and only if $P_\lambda f \circ \nabla g^* = \partial(\lambda f + g)^*$, and if and only if $\lambda f + g$ is convex without assuming that f is D-approximately compact.

Theorem 4.4. *Suppose that g is Fréchet differentiable, totally convex on* int(dom$g$) *and Legendre and that one of the two conditions* (a) *and* (b) *in Theorem 3.10 holds. Then, for any fixed $\lambda \in (0, \lambda_f)$, the following statements are equivalent:*

(i) $P_\lambda f$ *is single valued on* $\mathrm{int}(\mathrm{dom}\, g)$*;*

(ii) $P_\lambda f \circ \nabla g^*$ *is single valued and maximal monotone;*

(iii) $(\lambda f + g)^*$ *is Fréchet differentiable on* X^*;

(iv) $e_\lambda f \circ \nabla g^*$ *is Fréchet differentiable on* X^*;

(v) $\lambda f + g$ *is essentially strictly convex.*

If, in addition, ∇g and ∇g^* are Fréchet differentiable at each point of $\mathrm{int}(\mathrm{dom}\, g)$ *and* $\mathrm{int}(\mathrm{dom}\, g^*)$*, then* (i)–(v) *are equivalent to the following assertion.*

(vi) *The function $e_\lambda f$ is Fréchet differentiable on* $\mathrm{int}(\mathrm{dom}\, g)$.

Proof. (i) $\Leftrightarrow$ (ii) follows from Corollary 4.2 and the fact that ∇g is a topological isomorphism from int(*domg*) to X^* with $\nabla g^* = (\nabla g)^{-1}$. Suppose that (ii) holds. By Theorem 4.3 and Corollary 4.2, $\partial(\lambda f + g)^* = P_\lambda f \circ \nabla g^*$ is single-valued and continuous on X^* by noticing that ∇g is a topological isomorphism from $\mathrm{int}(\mathrm{dom}\, g)$ to X^* with $\nabla g^* = (\nabla g)^{-1}$. Hence, $(\lambda f + g)^*$ is continuously differentiable, which implies (iii). Conversely, if (iii) is true, then, by (15), $P_\lambda f \circ \nabla g^*$ is single-valued and so is $P_\lambda f$ which implies that $P_\lambda f \circ \nabla g^*$ is maximal monotone by Corollary 4.2. Hence (ii) holds true. (iii) $\Leftrightarrow$ (iv) follows from (13) since g^* is Fréchet differentiable on X^* by Proposition 2.6.

Suppose that (iii) holds. Then $(\lambda f + g)^*$ is proper convex continuous on X^* and then $(\lambda f + g)^{**}$ is proper convex. By Proposition 2.7, $\lambda f + g$ is convex, and consequently, it is proper convex and lower semicontinuous by condition (a) or (b). Again since $(\lambda f + g)^*$ is Fréchet differentiable on X^*, by Proposition 2.4, $\lambda f + g$ is essentially strictly convex. If (v) holds, then, for any fixed $x \in \mathrm{int}(\mathrm{dom}\, g)$, the function

$$w \mapsto \psi(w, x, \lambda) = f(w) + \frac{1}{\lambda} g(w) - \langle \frac{1}{\lambda} \nabla g(x), w \rangle + \frac{1}{\lambda} g^*(\nabla g(x))$$

is essentially strictly convex too. Hence $P_\lambda f(x) = \mathrm{argmin}_w \psi(w, x, \lambda)$ is nonempty and single-valued by Theorem 3.9 and [31, Lemma 3.1].

It is obvious that (iv) is equivalent to (vi) under the condition that ∇g and ∇g^* are Fréchet differentiable at each point of $\mathrm{int}(\mathrm{dom}\, g)$ and $\mathrm{int}(\mathrm{dom}\, g^*)$. ■

Acknowledgments. We are grateful to both referees for their careful reading and helpful suggestions on the paper.

References

1. E. Asplund and R. T. Rockafellar, Gradient of Convex Functions, *Trans. Amer. Math. Soc.* **139** (1969), 443–467.
2. H. Attouch, *Variational Convergence for Functions and Operators*, Pitman, Boston, 1984.
3. M. Bačák, J. M. Borwein, A. Eberhard and B. S. Mordukhovich, Infimal convolutions and Lipschitzian properties of subdifferentials for prox-regular functions in Hilbert spaces, *J. Convex Anal.* **17** (2010), 732–763.
4. V. S. Balaganski and L. P. Vlasov, The problem of the convexity of Chebyshev sets, *Russ. Math. Surv.* **51** (1996), 1127–1190.
5. H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. Convex Anal.* **4** (1997), 27–67.
6. H. H. Bauschke, J. M. Borwein and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.* **42** (2003), 596–636.
7. H. H. Bauschke, P. L. Combettes and D. Noll, Joint minimization with alternating Bregman proximity operators, *Pacific J. Optim.* **2** (2006), 401–424.
8. H. H. Bauschke, X. Wang, J. Ye and X. Yuan, Bregman distances and Chebyshev sets, *J. Approx. Theory* **159** (2009), 3–25.
9. F. Bernard and L. Thibault, Prox-regularity of functions and sets in Banach spaces, *Set-valued Anal.* **12** (2004), 25–47.
10. J. M. Borwein, Proximality and Chebyshev sets, *Optim. Lett.* **1** (2007), 21–32.
11. J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010.
12. M. Bougeard, J.-P. Penot, and A. Pommellet, Towards minimal assumptions for the infimal convolution regularization, *J. Math. Anal. Appl.* **64** (1991), 245–270.
13. L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *U.S.S.R. Comput. Math. Math. Phys.* **7** (1967), 200–217.
14. D. Butnariu and A. N. Iusem, Local moduli of convexity and their applications to finding almost common fixed points of mesurable families of operators, in: *Recent Developments in Optimization and Nonlinear Analysis, Contemporary Mathematics*, Y. Censor and S. Reich (Eds.), Vol. 204, American Mathematical Society, Providence, Rhode Island, 1997, pp. 33–61.
15. D. Butnariu and E. Resmerita, Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces, *Abstr. Appl. Anal.* (2006), ID 84919, 39 pages.
16. D. Butnariu, A. N. Iusem and C. Zalinescu, On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces, *J. Convex Anal.* **10** (2003) 35–61.
17. D. Butnariu, Y. Censor and S. Reich, Iterative averaging of entropic projections for solving stochastic convex feasibility problems, *Comput. Optim. Appl.* **8** (1997), 21–39.
18. P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.* **4** (2005), 1168–1200.
19. F. Deutsch, Best Approximations in Inner Product Spaces, in: *CMS Books Math./Ouvrages Math., in: SMC*, Vol.7, Springer-Verlag, New York, 2001.
20. N. W. Efimow and S. B. Stechkin, Approximative compactness and Chebyshev sets, *Sov. Math.* **2** (1961), 1226–1228.
21. W. Hare and C. Sagastizábal, Computing proximal points of nonconvex functions, *Math. Program. Ser. B* **116** (2009), 221–258.
22. J.-B. Hiriart-Urruty, Potpourri of conjectures and open questions in nonlinear analysis and optimization, *SIAM Rev.* **49** (2007), 255–273.
23. C. Kan and W. Song, Moreau envelope function and the proximal mapping in the sense of Bregman distance, *Nonlinear Anal.* **75** (2012), 1385–1399.

24. V. Klee, Convexity of Chebyshev sets, *Math. Ann.* **142** (1961), 292–304.
25. C. Li, W. Song and J-C. Yao, The Bregman distance, approximate compactness and convexity of Chebyshev sets in Banach spaces, *J. Approx. Theory* **162** (2010), 1128–1149.
26. B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I. Basic theory*, Springer-Verlag, Berlin, 2006.
27. J.-J. Moreau, Proximité et dualité dans un espace Hilbertien, *Bull. Soc. Math. Fr.* **93** (1965), 273–299.
28. E. Resmerita, On total convexity, Bregman projections and stability in Banach spaces, *J. Convex Anal.* **11** (2004), 1–16.
29. R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* **14** (1976), 877–898.
30. R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, New York, 1998.
31. X. Wang, On Chebyshev functions and Klee functions, *J. Math. Anal. Appl.* **368** (2010), 293–310.
32. C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.