Nonlinear Analysis 75 (2012) 1385–1399

Contents lists available at SciVerse ScienceDirect

Nonlinear Analysis

journal homepage: www.elsevier.com/locate/na

The Moreau envelope function and proximal mapping in the sense of the Bregman distance

Chao Kan [a], Wen Song [a,b,*]

[a] *Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, PR China*
[b] *School of Mathematical and Sciences, Harbin Normal University, Harbin, 150025, PR China*

ARTICLE INFO

Communicated by Ravi Agarwal

Keywords:
Bregman distance
Moreau envelope
Proximal mapping
Convexity
Continuity
Clarke regularity

ABSTRACT

In this paper, we explore some properties of the Moreau envelope function $e_\lambda f(x)$ of f and the associated proximal mapping $P_\lambda f(x)$ in the sense of the Bregman distance induced by a convex function g. Precisely, we study the continuity, differentiability, and Clarke regularity of the Moreau envelope function and the upper semicontinuity and single-valuedness of the proximal mapping as well as its relation to the convexity of $\lambda f+g$, where λ is a positive parameter.

1. Introduction

Let X be a Banach space, and let $f:X \to \bar{\mathbb{R}}$ be a generalized real-valued function. The *Moreau envelope* function $e_\lambda f:X \to [-\infty,+\infty]$, and the associated proximal mapping $P_\lambda f:X \rightrightarrows X$, are defined respectively by

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

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

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 [1–11]. It is well known that the set of minimizers of f over X coincides with the set of minimizers of $e_\lambda f$ over X. When $X=\mathbb{R}^n$ is a finite-dimensional space, it is well known that (see [5]) $P_\lambda f$ is maximal monotone if and only if $f+\frac{1}{2\lambda}\|\cdot\|^2$ is convex; $P_\lambda f$ is single valued and continuous if and only if $e_\lambda f$ is convex and continuously differentiable for the case when f is convex. Recently, Wang [10] proved, under some basic assumptions, that $P_\lambda f$ is single valued everywhere; in this case, f is referred to as a λ-*Chebyshev function*, if and only if $f+\frac{1}{2\lambda}\|\cdot\|^2$ is essentially strictly convex and if and only if $e_\lambda f$ is continuously differentiable 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 [1,6,11].

When the function f is an indicator function ι_C of a nonempty set C of X, we have $P_\lambda f=P_C$, where P_C is the metric projection operator on $C\subset X$, while $e_\lambda f$ coincides with $\frac{1}{2\lambda}d_C^2$ for the distance function d_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 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,

* Corresponding author at: Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, PR China.
E-mail address: wsong218@yahoo.com.cn (W. Song).

0362-546X/$ – see front matter
doi:10.1016/j.na.2011.07.031

and many others followed. The problem is still open in the case of infinite-dimensional Hilbert spaces [12–17]. The convexity problem of Chebyshev sets in general Banach spaces has also been studied extensively, and many sufficient conditions for a Chebyshev set to be convex have been obtained. In particular, Busemann [18] pointed out that each Chebyshev set in a smooth, strictly convex finite-dimensional space is convex. Klee [12] showed that any weakly closed Chebyshev set in a uniformly convex and uniformly smooth Banach space is convex, and more general results can be found in [14].

In 1976, Bregman [19] discovered an elegant and effective technique for the use of the function D_g (which is now called the Bregman distance) instead of the usual distance function in the process of designing and analyzing feasibility and optimization algorithms. This opened a growing area of research in which Bregman's technique is applied in various ways in order to design and analyze iterative algorithms not only for solving feasibility and optimization problems, but also for solving variational inequalities and computing fixed points of nonlinear mappings and more [20–24]. Let $g: X \to (-\infty, +\infty]$ be a proper convex function with its domain $\mathrm{dom}\, g$. The Bregman distance D_g is defined by

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

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

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

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.,

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

Bauschke et al. started in [25] to consider the convexity problem of Chebyshev sets in the sense of the Bregman distance in the Euclidean space $\mathbb{R}^n$. Under the assumption that g is a convex function of Legendre type and 1-coercive, they proved that each Chebyshev subset of $\mathbb{R}^n$ (in the sense of the Bregman distance) is convex. Li et al. [26] presented certain sufficient conditions and equivalent conditions for the convexity of a Chebyshev subset of a Banach space as well as some sufficient conditions ensuring the upper semicontinuity and the continuity of the Bregman projection operator Π_C^g.

Likewise, we shall consider the Moreau envelope function and the associated proximal mapping in the sense of the Bregman distance, namely the D-Moreau envelope $e_\lambda f$ and D-proximal mapping $P_\lambda f$, defined by

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

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

In particular, when the function $f = \iota_C$, we also have $P_\lambda f = \Pi_C^g$ [25,23]. Bauschke et al. [22,24] explored some properties of $P_\lambda f$ (where it is called D-prox operator), such as the effective domain, range, the set of fixed points, single-valuedness and continuity, as well as the differentiability of $e_\lambda f$ in the case when f is convex, which plays an important role in various Bregman monotone algorithms. It is natural to ask: if $P_\lambda f$ is single valued everywhere on its domain, what can we say about the function f? In this paper, we mainly study some basic properties of $e_\lambda f$ and $P_\lambda f$. In Section 2, we shall present some sufficient conditions for the continuity and locally Lipschitz continuity of $e_\lambda f$ and upper semicontinuity of $P_\lambda f$ and show that $\lim_{\lambda \searrow 0} e_\lambda f(x) = f(x)$; in Section 3, we shall explore the Clarke regularity and differentiability of $e_\lambda f$; in Section 4, we obtain a relationship between the single-valuedness of $P_\lambda f$ and the convexity of $\lambda f + g$.

2. Continuity of function $e_\lambda f$ and set-valued mapping $P_\lambda f$

In the following, we assume that $X = \mathbb{R}^n$ and $C \subset \mathbb{R}^n$ is a nonempty subset. As usual, we denote by $\bar{C} = \mathrm{cl}\, C$ the closure of C, $\mathrm{int}\, C$ the interior of C, $\mathrm{rint}\, C$ the relatively interior of C, and $\mathrm{conv}\, C$ the convex hull of C. For a function $f: \mathbb{R}^n \to \bar{\mathbb{R}}$, denote by $\mathrm{dom}\, f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$ the domain of f. In addition, we always denote by $B(x, \delta)$ the open ball with center at $x \in \mathbb{R}^n$ and radius $\delta \in (0, +\infty]$ and $\mathrm{lev}_{\leq \alpha} f = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ the (lower) level set of f for some $\alpha \in \mathbb{R}^n$. If T is a set-valued mapping on $\mathbb{R}^n$, we denote the domain and image of T, respectively, by

$$\mathrm{dom}\, T := \{x \in \mathbb{R}^n : T(x) \neq \emptyset\} \quad \text{and} \quad \mathrm{Im}\, T := \{x^* \in \mathbb{R}^n : x^* \in T(x), x \in \mathrm{dom}\, T\},$$

and $\mathrm{Fix}\, T := \{x \in \mathbb{R}^n : x \in T(x)\}$ is the set of fixed points of T.

Let $g: \mathbb{R}^n \to \bar{\mathbb{R}}$ be a proper convex function, and let $x \in \mathrm{dom}\, g$. The *subdifferential* of g at x is the convex set defined by

$$\partial g(x) := \{x^* \in \mathbb{R}^n : g(x) + \langle x^*, y - x \rangle \leq g(y) \text{ for each } y \in \mathbb{R}^n\},$$

while the *conjugate function* of g (not necessarily convex) is the function $g^*: \mathbb{R}^n \to \bar{\mathbb{R}}$ defined by

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

Then, by [5], the Young–Fenchel inequality holds:

$$\langle x^*, x\rangle \le g(x) + g^*(x^*) \quad \text{for each pair } (x^*, x) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{2.1}$$

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

$$\langle x^*, x\rangle = g(x) + g^*(x^*) \Longleftrightarrow x^* \in \partial g(x) \quad \text{for each pair } (x^*, x) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{2.2}$$

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

$$\mathrm{dom}(\partial g) := \{x \in \mathrm{dom}\, g \colon \partial g(x) \neq \emptyset\}$$

and

$$\mathrm{Im}(\partial g) := \{x^* \in \mathbb{R}^n \colon x^* \in \partial g(x),\ x \in \mathrm{dom}(\partial g)\}.$$

Let $g\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ be a lower semicontinuous proper convex function with $\mathrm{int}(\mathrm{dom}\, g) \neq \emptyset$. Recall that (see [27,21]) g is said to be essentially smooth if ∂g contains at most one element for every x; g is said to be essentially strictly convex if g is strictly convex on every convex subset of $\mathrm{dom}(\partial g)$; g is said to be Legendre if g is both essentially smooth and essentially strictly convex; g is 1-coercive, if $\lim_{\|x\|\to+\infty} \frac{g(x)}{\|x\|} = +\infty$, which is equivalent to $\mathrm{dom}\, g^* = \mathbb{R}^n$. Some examples of functions with the above properties can be found in [27,21,24]. It is well known that (see [27]), if g is essentially strictly convex, then $\mathrm{argmin}\, g$ is at most a singleton, and g is of Legendre type if and only if g^* is. In this case, the gradient mapping

$$\nabla g \colon \mathrm{int}(\mathrm{dom}\, g) \to \mathrm{int}(\mathrm{dom}\, g^*) \colon x \to \nabla g(x)$$

is a topological isomorphism with inverse mapping $(\nabla g)^{-1} = \nabla g^*$.

Definition 2.1. Let $g\colon \mathbb{R}^n \to (-\infty, +\infty]$ be a proper convex lower semicontinuous function which is differentiable on $\mathrm{int}(\mathrm{dom}\, g)$. The Bregman distance function $D_g(w, x)\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty]$ is defined by

$$D_g(w, x) = \begin{cases} g(w) - g(x) - \langle \nabla g(x), w - x\rangle & \text{if } x \in \mathrm{int}(\mathrm{dom}\, g), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\nabla g(x)$ denotes the gradient operator of g at x.

Next, we define the *D-Moreau envelope* function $e_\lambda f$ and the associated D-proximal mapping $P_\lambda f$.

Definition 2.2. Let $f\colon \mathbb{R}^n \to (-\infty, +\infty]$. For some index $\lambda \in (0, +\infty]$, the *D-Moreau envelope* function $e_\lambda f$ and the associated 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

$$P_\lambda f(x) = \left\{ w \in \mathrm{dom}\, f \cap \mathrm{dom}\, g : f(w) + \frac{1}{\lambda} D_g(w, x) = \inf\left(f + \frac{1}{\lambda} D_g(\cdot, x) \right)(\mathbb{R}^n) < +\infty \right\}.$$

Therefore, we have $\mathrm{dom}\, P_\lambda f \subset \mathrm{int}(\mathrm{dom}\, g)$ and $\mathrm{ran}\, P_\lambda f \subset \mathrm{dom}\, f \cap \mathrm{dom}\, g$.

Definition 2.3 (*Prox-Boundedness*). A function $f\colon \mathbb{R}^n \to (-\infty, +\infty]$ is prox bounded if there exists $\lambda > 0$ such that $e_\lambda f(x) > -\infty$ for some $x \in \mathbb{R}^n$. The supremum of the set of all such λ is the threshold λ_f of the prox-boundedness, i.e.,

$$\lambda_f = \sup\{\lambda > 0\colon \text{ there is } x \in \mathbb{R}^n \text{ such that } e_\lambda f(x) > -\infty\}.$$

We observe that if there exists $\bar{\lambda}$ such that $e_{\bar{\lambda}} f(\bar{x}) > -\infty$ for some $\bar{x} \in \mathbb{R}^n$, we then have $e_\lambda f(\bar{x}) \ge e_{\bar{\lambda}} f(\bar{x}) > -\infty$ for all $\lambda \in (0, \bar{\lambda})$. This implies that, for any $\lambda \in (0, \lambda_f)$, there is $x \in \mathbb{R}^n$ such that $e_\lambda f(x) > -\infty$.

Definition 2.4 (*Level Boundedness*). A function $f\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ is (lower) level bounded if, for every $\alpha \in \mathbb{R}$, the set $\mathrm{lev}_{\le\alpha} f$ is bounded (possibly empty).

Definition 2.5 (*Uniform Level Boundedness*). A function $\varphi\colon \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ is level bounded in the first variable w locally uniformly at $\bar{x}$ if, for every $\alpha \in \mathbb{R}$, there exists $\delta > 0$ such that

$$\bigcup_{x \in B(\bar{x}, \delta)} \{w \in \mathbb{R}^n \colon \varphi(w, x) \le \alpha\}$$

is bounded.

In order to discuss the continuity of $e_\lambda f(x) = \inf_w\{f(w) + \frac{1}{\lambda}D_g(w, x)\}$ and $P_\lambda f$, we need to establish the uniform level boundedness of $\psi(w, x, \lambda) := f(w) + \frac{1}{\lambda}D_g(w, x)$. For that purpose, we first recall a known result concerning the continuity of the gradient operator of a convex function.

Lemma 2.1 (*See [28]*). *Suppose that $g: X \to \bar{\mathbb{R}}$ is a lower semicontinuous proper convex function defined a Banach space which is Gâteaux differentiable (respectively, Fréchet differentiable) on* $\mathrm{int}(\mathrm{dom}\, g)$. *Then g is continuous and its Gâteaux derivative ∇g is norm-weak * continuous (respectively, continuous) on* $\mathrm{int}(\mathrm{dom}\, g)$.

Since the underlying space is the Euclidean space $\mathbb{R}^n$ in this paper, by Lemma 2.1, the function g and its gradient ∇g defined in the definition of the Bregman distance function are both continuous on $\mathrm{int}(\mathrm{dom}\, g)$.

Theorem 2.1. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous prox-bounded function and that g is a lower semicontinuous proper convex function which is differentiable on* $\mathrm{int}(\mathrm{dom}\, g)$ *and* 1*-coercive. Then the function $\psi(w, x, \lambda)$ is level bounded in w locally uniformly at every $(x, \lambda) \in \mathrm{int}(\mathrm{dom}\, g) \times (0, \lambda_f)$.*

Proof. Suppose that the conclusion does not hold. Then there exist $(\bar{x}, \bar{\lambda}) \in \mathrm{int}(\mathrm{dom}\, g) \times (0, \lambda_f)$ and some $\alpha \in \mathbb{R}$ such that, for each n, we can find $(x_n, \lambda_n) \in B((\bar{x}, \bar{\lambda}), \frac{1}{n})$, $w_n \in \mathbb{R}^n$ with $\|w_n\| \geq n$ such that $\psi(w_n, x_n, \lambda_n) \leq \alpha$, i.e.,

$$f(w_n) + \frac{1}{\lambda_n}D_g(w_n, x_n) \leq \alpha. \tag{2.3}$$

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 fixed $\lambda' \in (\lambda_0, \lambda_f)$, there exist some $x_0 \in \mathrm{int}(\mathrm{dom}\, g)$ and $\beta \in \mathbb{R}$ such that

$$f(w_n) + \frac{1}{\lambda'}D_g(w_n, x_0) \geq e_{\lambda'}f(x_0) \geq \beta > -\infty. \tag{2.4}$$

Combining (2.3) and (2.4), we get

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

That is,

$$\left(\frac{1}{\lambda_n} - \frac{1}{\lambda'}\right)g(w_n) + \frac{1}{\lambda'}g(x_0) - \frac{1}{\lambda_n}g(x_n) + \frac{1}{\lambda'}\langle\nabla g(x_0), w_n - x_0\rangle - \frac{1}{\lambda_n}\langle\nabla g(x_n), w_n - x_n\rangle \leq \alpha - \beta. \tag{2.5}$$

Divide both sides of the last inequality by $\|w_n\|$, and then since g is 1-coercive and $\nabla g(x_n) \to \nabla g(\bar{x})$ by Lemma 2.1, we have that $\lim_{n\to+\infty} \frac{g(w_n)}{\|w_n\|} = +\infty$, $\frac{g(x_0)}{\|w_n\|} \to 0$, $\frac{1}{\lambda_n}\frac{g(x_n)}{\|w_n\|} \to 0$, as $n \to \infty$, and both $\frac{1}{\lambda'}\langle\nabla g(x_0), \frac{w_n - x_0}{\|w_n\|}\rangle$ and $\frac{1}{\lambda_n}\langle\nabla g(x_n), \frac{w_n - x_n}{\|w_n\|}\rangle$ are bounded. Since $0 < \lambda_n < \lambda_0 < \lambda'$, we have $\frac{1}{\lambda_n} - \frac{1}{\lambda'} > 0$ for sufficiently large n. Hence, we get $+\infty \leq \alpha - \beta$, which is a contradiction. □

Remark 1. In the remainder of this paper, we always assume that $g: \mathbb{R}^n \to (-\infty, +\infty]$ is a lower semicontinuous proper convex function which is differentiable on $\mathrm{int}(\mathrm{dom}\, g)$ and 1-coercive, and that $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \emptyset$.

Corollary 2.1. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. Then, for any fixed $\lambda \in (0, \lambda_f)$, the function $\psi_\lambda(w, x) := \psi(w, x, \lambda)$ is level bounded in w locally uniformly at every $x \in (\mathrm{dom}\, f \cap \mathrm{dom}\, g) \times$* $\mathrm{int}(\mathrm{dom}\, g)$. *In particular, for any fixed $(x, \lambda) \in \mathrm{int}(\mathrm{dom}\, g) \times (0, \lambda_f)$, the function $\varphi(w) := \psi(w, x, \lambda)$ is level bounded on* $\mathrm{dom}\, f \cap \mathrm{dom}\, g$.

Lemma 2.2 (*See [5]*). *Suppose that $\varphi: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and level-bounded function. Then* $\inf \varphi$ *is finite, and the set* $\mathrm{argmin}\varphi$ *is nonempty and compact.*

Theorem 2.2. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. Then, for any fixed $\lambda \in (0, \lambda_f)$, we have*

(i) *$P_\lambda f(x)$ is nonempty and compact and $e_\lambda f(x)$ is finite for all $x \in$* $\mathrm{int}(\mathrm{dom}\, g)$;
(ii) $\mathrm{dom}\, e_\lambda f = \mathrm{dom}\, P_\lambda f = \mathrm{int}(\mathrm{dom}\, g)$;
(iii) *$e_\lambda f$ is proper and lower semicontinuous.*

Proof. (i) For any fixed $\lambda \in (0, \lambda_f)$ and any $x \in \mathrm{int}(\mathrm{dom}\, g)$, clearly, $\varphi(w) = \psi(w, x, \lambda)$ is lower semicontinuous in w. It follows from Corollary 2.1 and Lemma 2.2 that $P_\lambda f(x) = \mathrm{argmin}_w \varphi(w)$ is nonempty compact and that $e_\lambda f(x) = \inf_w \varphi(w) > -\infty$ is finite.

(ii) It follows from (i) that $\mathrm{int}(\mathrm{dom}\, g) \subset \mathrm{dom}\, P_\lambda f$ and $\mathrm{int}(\mathrm{dom}\, g) \subset \mathrm{dom}\, e_\lambda f$. The converse inclusions are obviously true by the definitions of $P_\lambda f$ and $e_\lambda f$. Hence, $\mathrm{dom}\, P_\lambda f = \mathrm{dom}\, e_\lambda f = \mathrm{int}(\mathrm{dom}\, g)$.

(iii) Clearly, $e_\lambda f$ is proper by (i). To prove that $e_\lambda f$ is lower semicontinuous, it is sufficient to prove that $\mathrm{lev}_{\le\alpha} e_\lambda f$ is closed for every $\alpha \in \mathbb{R}$. Let $\{x_n\} \subset \mathrm{lev}_{\le\alpha} e_\lambda f$ be a sequence such that $x_n \to \bar{x}$. Since f is lower semicontinuous, and g and its gradient ∇g are continuous on int(dom g), we can easily show that $\psi_\lambda(w, x) = \psi(w, x, \lambda)$ is lower semicontinuous in (w, x), and consequently the level set

$$\mathrm{lev}_{\le\alpha}\psi_\lambda = \{(w, x) \colon \psi_\lambda(w, x) \le \alpha\}$$

is closed. Since ψ_λ is level bounded in w locally uniformly at $\bar{x}$ by Corollary 2.1, there exists $\delta > 0$ such that

$$\bigcup_{x\in\bar{B}(\bar{x},\delta)} \{w \colon \psi_\lambda(w, x) \le \alpha\}$$

is bounded. Consequently,

$$\Omega := \{(w, x) \colon \psi_\lambda(w, x) \le \alpha\} \cap (\mathbb{R}^n \times \bar{B}(\bar{x}, \delta))$$

is bounded and closed, and so is compact.

For $x_n \in \mathrm{lev}_{\le\alpha} e_\lambda f$, i.e., $e_\lambda f(x_n) \le \alpha$, there exists $w_n \in P_\lambda f(x_n)$ such that

$$e_\lambda f(x_n) = \psi_\lambda(w_n, x_n) \le \alpha.$$

Hence, $(w_n, x_n) \in \mathrm{lev}_{\le\alpha}\psi_\lambda$, and hence $(w_n, x_n) \in \Omega$ for n large enough. Since Ω is compact and $x_n \to \bar{x}$, there exists a subsequence $\{(w_{n_k}, x_{n_k})\}$ of $\{(w_n, x_n)\}$ converging to $(\bar{w}, \bar{x}) \in \Omega$, i.e., $\psi_\lambda(\bar{w}, \bar{x}) \le \alpha$. Thus, we get

$$e_\lambda f(\bar{x}) = \inf_w \psi_\lambda(w, \bar{x}) \le \psi_\lambda(\bar{w}, \bar{x}) \le \alpha.$$

This shows that $\mathrm{lev}_{\le\alpha} e_\lambda f$ is closed; equivalently, $e_\lambda f$ is lower semicontinuous. □

Corollary 2.2. *Suppose that $f \colon \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. Then, for any fixed $\lambda \in (0, \lambda_f)$, $e_\lambda f(x)$ is continuous on* int(dom g).

Proof. According to Theorem 2.2, it is sufficient to prove that $e_\lambda f$ is upper semicontinuous on int(dom g). Since g and ∇g are continuous on int(dom g), for any fixed $\lambda \in (0, \lambda_f)$ and $w \in \mathrm{dom} f \cap \mathrm{dom}\, g$, the function $\psi(w, x, \lambda)$ is continuous in x on int(dom g). For any fixed $\bar{x} \in$ int(dom g), since $P_\lambda f(\bar{x}) \ne \phi$, there is $\bar{w} \in P_\lambda f(\bar{x})$ such that $e_\lambda f(\bar{x}) = \psi(\bar{w}, \bar{x}, \lambda)$. Hence,

$$\limsup_{x\to\bar{x}} e_\lambda f(x) \le \limsup_{x\to\bar{x}} \psi(\bar{w}, x, \lambda) = \psi(\bar{w}, \bar{x}, \lambda) = e_\lambda f(\bar{x}).$$

This shows that $e_\lambda f$ is upper semicontinuous on int(dom g). □

Theorem 2.3. *Suppose that $f \colon \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. Then $e_\lambda f(x)$ is continuous in (x, λ) on* int(dom g) $\times (0, \lambda_f)$.

Proof. Since g and ∇g are continuous on int(dom g), we can easily see that, for any fixed $w \in \mathrm{dom} f \cap \mathrm{dom}\, g$, the function

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

is continuous in variables (x, λ) on int(dom g) $\times (0, \lambda_f)$. Hence, by analogous arguments to Theorem 2.2 and Corollary 2.2, we can prove that $e_\lambda f(x)$ is continuous in (x, λ) on int(dom g) $\times (0, \lambda_f)$. □

Theorem 2.4. *Suppose that $f \colon \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function, and that g is Legendre. Then, for any fixed $\lambda \in (0, \lambda_f)$, $e_\lambda f$ satisfies*

$$e_\lambda f = \left(\frac{1}{\lambda} g^* - \frac{1}{\lambda}(\lambda f + g)^*\right) \circ \nabla g. \tag{2.6}$$

and

$$(\lambda f + g)^* = -\lambda e_\lambda f \circ \nabla g^* + g^*, \tag{2.7}$$

and consequently $e_\lambda f \circ \nabla g^*$ is locally Lipschitz on $\mathbb{R}^n$. Moreover, if, in addition, ∇g is locally Lipschitz on int(dom g), *then $e_\lambda f$ is locally Lipschitz on* int(dom g).

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

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

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

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

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

This means that

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

Since, being convex functions, both g^* and $(\lambda f + g)^*$ are locally Lipchitz on $\mathbb{R}^n$, we have that $e_\lambda f \circ \nabla g^*$ is locally Lipschitz on $\mathbb{R}^n$. Moreover, if ∇g is locally Lipschitz on $\operatorname{int}(\operatorname{dom} g)$, then, from (2.6), we have that $e_\lambda f$ is locally Lipschitz on $\operatorname{int}(\operatorname{dom} g)$. $\square$

Note that if g is second-order continuously differentiable and $\nabla^2 g$ is positive definite, we can easily obtain that $e_\lambda f$ is locally Lipschitz.

It is a natural question to ask whether $\lim_{\lambda \to 0} e_\lambda f(x) = f(x)$ for all $x \in \operatorname{int}(\operatorname{dom} g)$. An affirmative answer to this problem for the case when $D_g(w, x) = \frac{1}{2} \|w - x\|^2$, or a more general distance-like functions have been given in [3,29]. Next, we shall present an example which shows this is not true in general.

Example 2.1. Let

$$f(x) = \begin{cases} 1 & \text{if } x \in (-\infty, 0), \\ -1 & \text{if } x \in [0, +\infty), \end{cases}$$

and

$$g(x) = \begin{cases} (x+1)^2 & \text{if } x \in (-\infty, -1], \\ 0 & \text{if } x \in (-1, 1), \\ (x-1)^2 & \text{if } x \in [1, +\infty). \end{cases}$$

Obviously, $g(x)$ is 1-coercive, continuously differentiable, proper, and convex. For any $x \in [-1, 1]$, $D_g(w, x) = 0$ as $w \in [-1, 1]$. Thus, for any $x \in [-1, 1]$, we have

$$\begin{aligned} e_\lambda f(x) &= \inf_{w \in R^n} \left\{ f(w) + \frac{1}{\lambda} D_g(w, x) \right\} \\ &\le \inf_{w \in [-1,1]} \left\{ f(w) + \frac{1}{\lambda} D_g(w, x) \right\} \\ &= \inf_{w \in [-1,1]} \{ f(w) \} \\ &= -1. \end{aligned}$$

If $x \in [-1, 0)$, then $e_\lambda f(x) \le -1 < 1 = f(x)$. This means that $\sup_{\lambda \in (0, \lambda_f)} e_\lambda f(x) \le -1 < f(x)$.

Proposition 2.1. *Suppose that $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. 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 $0 < \alpha < \beta < \lambda_f$, $\frac{1}{\alpha} - \frac{1}{\beta} > 0$, and 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 $x \in \operatorname{int}(\operatorname{dom} g)$,

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

and hence

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

Theorem 2.5. *Suppose that $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. If g is strictly convex, then, for any $x \in \operatorname{int}(\operatorname{dom} g) \cap \operatorname{dom} f$,*

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

Proof. Since $e_\lambda f(x)$ is nonincreasing in $\lambda \in (0, \lambda_f)$, for any fixed $x \in \operatorname{int}(\operatorname{dom} g) \cap \operatorname{dom} f$ by Proposition 2.1, 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. For every n, there is $w_n \in P_{\lambda_n} f(x)$ such that

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

and hence, by Proposition 2.1(ii),

$$D_g(w_{n+1}, x) \le D_g(w_n, x).$$

This implies that the limit of $\{D_g(w_n, x)\}$ exists, and we denote it by $d := \lim_{n\to\infty} D_g(w_n, x)$. Then $d \ge 0$, since the Bregman distance is nonnegative. Next, we will show that $d = 0$.

For any $\alpha > d = \lim_{n\to\infty} D_g(w_n, x)$, we have $D_g(w_n, x) < \alpha$ for sufficiently large n. Since g is 1-coercive, we can easily verify that $w \to D_g(w, x)$ is 1-coercive too. It follows that $\{w_n\}$ is bounded. We may assume, without loss of generality, that $w_n \to \bar{w}$. Since f is lower semicontinuous, $\lim_{n\to\infty} f(w_n) \ge f(\bar{w})$. Hence, for any $\varepsilon > 0$, we have $f(w_n) > f(\bar{w}) - \varepsilon$ for sufficiently large n. Furthermore, we have

$$f(\bar{w}) - \varepsilon + \frac{1}{\lambda_n} d \le f(w_n) + \frac{1}{\lambda_n} D_g(w_n, x) = e_{\lambda_n} f(x) \le f(x) \tag{2.8}$$

for sufficiently large n. If $d > 0$, by taking the limit in the above inequality as $n \to \infty$, we then get a contradiction. Therefore, $d = 0$.

Since g is strictly convex, we have that $\lim_{n\to\infty} D_g(w_n, x) = 0$ implies that $\|w_n - x\| \to 0$ as $n \to \infty$. Hence, $\bar{w} = x$, and (2.8) becomes the following:

$$f(x) - \varepsilon \le f(w_n) + \frac{1}{\lambda_n} D_g(w_n, x) = e_{\lambda_n} f(x) \le f(x). \tag{2.9}$$

Letting $n \to \infty$ in (2.9), and by noting that ε is arbitrary, we get that

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

From Theorem 2.5, we can see that, if $e_\lambda f$ is convex for all $\lambda > 0$, then f is convex.

Definition 2.6. Let U, X be normed spaces. A set-valued mapping $F : U \rightrightarrows X$ is said to be closed valued if $F(u)$ is a closed subset of X for every $u \in U$. F is said to be closed if its graph $\operatorname{gph}(F)$ is a closed subset of $U \times X$.

Definition 2.7. Let U, X be normed spaces. A set-valued mapping $F: U \rightrightarrows X$ is said to be upper semicontinuous at a point $u_0 \in U$ if, for any open set $V \supset F(u_0)$, there exists $\delta > 0$ such that, for every $u \in B(u_0, \delta)$, the inclusion $F(u) \subset V$ holds. F is said to be upper semicontinuous if it is upper semicontinuous at every $u_0 \in U$.

By Theorem 2.2, if f is a proper lower semicontinuous and prox-bounded function, for every $\lambda \in (0, \lambda_f)$ and every $x \in \operatorname{int}(\operatorname{dom} g)$, $P_\lambda f(x)$ is nonempty and compact and then closed valued.

Theorem 2.6. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. Then, for any $\lambda \in (0, \lambda_f)$, the set-valued mapping $P_\lambda f$ is upper semicontinuous.*

Proof. Suppose that 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 $w_n \in \mathbb{R}^n \setminus V$. Since $e_\lambda f(x)$ is continuous in x on $\operatorname{int}(\operatorname{dom} g)$ by Corollary 2.2 and $x_n \to x_0$, for any $\delta > 0$, $\alpha > e_\lambda f(x_0)$, there is a positive integer number n_0 such that

$$e_\lambda f(x_n) = \psi(w_n, x_n, \lambda) < \alpha \quad \text{for all } n \geq n_0,$$

and $\{x_n\}_{n \geq n_0} \subset B(x_0, \delta)$. From the proof of Theorem 4.1(iii), we can see that $\{(w_n, x_n)\}$ is bounded. We can assume, without loss of generality, that

$$(w_n, x_n) \to (\bar{w}, x_0).$$

Hence, $\psi(\bar{w}, x_0, \lambda) \leq \alpha$, and hence $\psi(\bar{w}, x_0, \lambda) \leq e_\lambda f(x_0)$, since α is an arbitrary number larger than $e_\lambda f(x_0)$. This implies that $\bar{w} \in P_\lambda f(x_0) \subset V$. However, for $w_n \in \mathbb{R}^n \setminus V$ and $\mathbb{R}^n \setminus V$ being closed, we also have $\bar{w} \in \mathbb{R}^n \setminus V$. This is a contradiction. Therefore, $P_\lambda f$ is upper semicontinuous. □

From Theorem 2.6, we can get the following corollaries.

Corollary 2.3. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. Then, for any fixed $\lambda \in (0, \lambda_f)$, $P_\lambda f$ is closed.*

Proof. Since each upper semicontinuous set-valued mapping with closed values must be closed (see [5]), the conclusion follows from Theorem 2.6 and the compactness of $P_\lambda f(x)$ for all $x \in \operatorname{int}(\operatorname{dom} g)$. □

Corollary 2.4. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. If, for any fixed $\lambda \in (0, \lambda_f)$, some sequence $w_n \in P_\lambda f(x_n)$ with $x_n \to \bar{x} \in \operatorname{int}(\operatorname{dom} g)$, then $\{w_n\}$ is bounded, and all its cluster points lie in $P_\lambda f(\bar{x})$.*

Corollary 2.5. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. If $P_\lambda f$ is single valued on* $\operatorname{int}(\operatorname{dom} g)$, *then $P_\lambda f$ is continuous.*

3. Clarke regularity of $e_\lambda f$

First, we recall some definitions and notations of derivatives and subdifferentials, which are taken from [30].

Definition 3.1 (*Derivatives and Subdifferentials*)**.** Suppose that $\phi: \mathbb{R}^n \to \bar{\mathbb{R}}$ is a function that is finite at a point $\bar{x}$. We say that f is *Fréchet differentiable* at $\bar{x}$ if there is a linear continuous operator $\nabla\phi(\bar{x}): \mathbb{R}^n \to \mathbb{R}$, called the *Fréchet derivative* of ϕ at $\bar{x}$, such that

$$\lim_{x \to \bar{x}} \frac{\phi(x) - \phi(\bar{x}) - \nabla\phi(\bar{x})(x - \bar{x})}{\|x - \bar{x}\|} = 0;$$

the *lower Dini (or Dini–Hadamard) directional derivative* of ϕ at $\bar{x}$ in the direction w is defined by

$$\mathrm{d}^-\phi(\bar{x}, w) = \liminf_{\substack{t \searrow 0 \\ w' \to w}} \frac{\phi(\bar{x} + tw') - \phi(\bar{x})}{t},$$

and if, in addition, ϕ is locally Lipschitz at $\bar{x}$, then

$$\mathrm{d}^-\phi(\bar{x}, w) = \liminf_{t \searrow 0} \frac{\phi(\bar{x} + tw) - \phi(\bar{x})}{t};$$

when ϕ is locally Lipschitz at $\bar{x}$, the *Clarke generalized derivative* of ϕ at $\bar{x}$ in the direction w is defined by

$$\phi^0(\bar{x}, w) = \limsup_{\substack{t \searrow 0 \\ x \to \bar{x}}} \frac{\phi(x + tw) - \phi(x)}{t}.$$

The corresponding *Fréchet subdifferential*, *lower Dini subdifferential*, and *Clarke subdifferential* of ϕ at the point $\bar{x}$ are defined by

$$\widehat{\partial}\phi(\bar{x}) = \left\{x^* \in \mathbb{R}^* \,|\, \liminf_{x\to\bar{x}} \frac{\phi(x) - \phi(\bar{x}) - \langle x^*, x - x^*\rangle}{\|x - \bar{x}\|} \geq 0\right\},$$
$$\partial^-\phi(\bar{x}) = \{x^* \in \mathbb{R}^n \mid \langle x^*, w\rangle \leq \mathrm{d}^-\phi(\bar{x}, w) \text{ for any } w \in \mathbb{R}^n\},$$

and

$$\partial_C\phi(\bar{x}) = \{x^* \in \mathbb{R}^n \mid \langle x^*, w\rangle \leq \phi^0(\bar{x}, w) \text{ for any } w \in \mathbb{R}^n\}.$$

Furthermore, the *basic subdifferential (Mordukhovich subdifferential)* is defined by

$$\partial\phi(\bar{x}) = \operatorname*{Limsup}_{x \xrightarrow{\phi} \bar{x}} \widehat{\partial}\phi(x).$$

We say that ϕ is *Clarke regular* at $\bar{x}$ if $\phi^0(\bar{x}, w) = \phi'(\bar{x}, w)$ for every $w \in \mathbb{R}^n$, where $\phi'(\bar{x}, w)$ is the classical *directional derivative*.

Remark 2. On the one hand, by the definitions, Clarke's directional derivative is a majorant of both the lower Dini directional derivative and its upper counterpart for locally Lipschitz functions, i.e.,

$$\mathrm{d}^-\phi(\bar{x}, w) \leq \mathrm{d}^+\phi(\bar{x}, w) \leq \phi^0(\bar{x}, w) \quad \text{for all } w \in \mathbb{R}^n,$$

where

$$\mathrm{d}^+\phi(\bar{x}, w) = \limsup_{t\searrow 0} \frac{\phi(\bar{x} + tw) - \phi(\bar{x})}{t}.$$

Consequently, if $\mathrm{d}^-\phi(\bar{x}, w) = \phi^0(\bar{x}, w)$ holds for all $w \in \mathbb{R}^n$, then

$$\mathrm{d}^-\phi(\bar{x}, w) = \mathrm{d}^+\phi(\bar{x}, w) = \phi^0(\bar{x}, w) \quad \text{for all } w \in \mathbb{R}^n,$$

which is equivalent to

$$\phi'(\bar{x}, w) = \phi^0(\bar{x}, w) \quad \text{for all } w \in \mathbb{R}^n.$$

Therefore, for a local Lipschitzian function ϕ, we can rewrite the Clarke regularity as $\phi^0(\bar{x}, w) = \mathrm{d}^-\phi(\bar{x}, w)$, which is equivalent to $\partial^-\phi(\bar{x}) = \partial_C\phi(\bar{x})$.

On the other hand, in the finite-dimensional space $\mathbb{R}^n$, the Fréchet subdifferential is just the lower Dini subdifferential, i.e., $\widehat{\partial}\phi(\bar{x}) = \partial^-\phi(\bar{x})$ (see [30, Page 144]). Hence, $\partial\phi(\bar{x}) = \operatorname{Limsup}_{x\to\bar{x}}\partial^-\phi(x)$. For basic properties and applications of these subdifferentials, we refer the reader to [30,5].

Theorem 3.1. *Suppose that $f: \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function, and that g is second-order continuously differentiable. Then $(-e_\lambda f)$ is Clarke regular, and, for any $x \in \operatorname{int}(\operatorname{dom} g)$ and $y \in \mathbb{R}^n$,*

$$\mathrm{d}^-(-e_\lambda f)(x, y) = (-e_\lambda f)^0(x, y) = \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x) - x), y\rangle; \tag{3.10}$$

$$\partial^-(-e_\lambda f)(x) = \partial_C(-e_\lambda f)(x) = \frac{1}{\lambda}\nabla^2 g(x)(\operatorname{conv}(P_\lambda f(x)) - x). \tag{3.11}$$

Proof. Let $\phi = -e_\lambda^f$. Since $e_\lambda f$ is locally Lipschitz on $\operatorname{int}(\operatorname{dom} g)$ by Theorem 2.4, we have

$$\mathrm{d}^-\phi(x, y) = \liminf_{t\searrow 0} \frac{\phi(x + ty) - \phi(x)}{t}.$$

For any fixed $x \in \operatorname{int}(\operatorname{dom} g)$, and for every $y \in \mathbb{R}^n$, there exists $t > 0$ sufficiently small such that $x + ty \in \operatorname{int}(\operatorname{dom} g)$. Taking any $w \in P_\lambda f(x)$, since

$$e_\lambda f(x) = f(w) + \frac{1}{\lambda}(g(w) - g(x) - \langle \nabla g(x), w - x\rangle),$$
$$e_\lambda f(x + ty) \leq f(w) + \frac{1}{\lambda}(g(w) - g(x + ty) - \langle \nabla g(x + ty), w - x - ty\rangle),$$

we have

$$\begin{aligned}\frac{\phi(x+ty)-\phi(x)}{t} &= \frac{-e_\lambda f(x+ty)-(-e_\lambda f(x))}{t}\\ &\geq \frac{1}{\lambda}\left[\frac{g(x+ty)-g(x)}{t}+\frac{\langle \nabla g(x+ty)-\nabla g(x), w-x\rangle}{t}-\langle \nabla g(x+ty), y\rangle\right]. \end{aligned} \tag{3.12}$$

Noting that g is second-order continuously differentiable, and by taking lim inf as $t \searrow 0$ in (3.12), we get

$$\liminf_{t\searrow 0}\frac{\phi(x+ty)-\phi(x)}{t} \geq \frac{1}{\lambda}\langle \nabla^2 g(x)y, w-x\rangle.$$

This implies that

$$\liminf_{t\searrow 0}\frac{\phi(x+ty)-\phi(x)}{t} \geq \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x)-x), y\rangle.$$

On the other hand, for any $w_t \in P_\lambda f(x+ty)$, we have

$$\begin{aligned} e_\lambda f(x+ty) &= f(w_t)+\frac{1}{\lambda}D_g(w_t, x+ty)\\ &= f(w_t)+\frac{1}{\lambda}[g(w_t)-g(x+ty)-\langle \nabla g(x+ty), w_t-x-ty\rangle],\\ e_\lambda f(x) &\leq f(w_t)+\frac{1}{\lambda}D_g(w_t, x)\\ &= f(w_t)+\frac{1}{\lambda}[g(w_t)-g(x)-\langle \nabla g(x), w_t-x\rangle]. \end{aligned}$$

Hence,

$$\begin{aligned}\frac{\phi(x+ty)-\phi(x)}{t} &= \frac{-e_\lambda f(x+ty)-(-e_\lambda f(x))}{t}\\ &\leq \frac{1}{\lambda}\left[\frac{g(x+ty)-g(x)}{t}+\frac{\langle \nabla g(x+ty)-\nabla g(x), w_t-x\rangle}{t}-\langle \nabla g(x+ty), y\rangle\right]. \end{aligned} \tag{3.13}$$

Take any positive sequence $t_n \to 0$ such that

$$d^-\phi(x, y) = \lim_{n\to 0}\frac{\phi(x+t_n y)-\phi(x)}{t_n}.$$

From Corollary 2.5, $\{w_{t_n}\}$ has at least one cluster point $\bar{w} \in P_\lambda f(x)$. Without loss of generality, we may assume that $w_{t_n} \to \bar{w} \in P_\lambda f(x)$. Then (3.13) implies that

$$\frac{\phi(x+t_n y)-\phi(x)}{t_n} \leq \frac{1}{\lambda}\left[\frac{g(x+t_n y)-g(x)}{t_n}+\frac{\langle \nabla g(x+t_n y)-\nabla g(x), w_{t_n}-x\rangle}{t_n}-\langle \nabla g(x+t_n y), y\rangle\right]. \tag{3.14}$$

Taking the limit as $n \to 0$ in (3.14), we have

$$\begin{aligned} d^-\phi(x, y) &\leq \frac{1}{\lambda}\langle \nabla^2 g(x)y, \bar{w}-x\rangle\\ &\leq \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x)-x), y\rangle. \end{aligned}$$

Therefore, we have

$$d^-\phi(x, y) = \liminf_{t\searrow 0}\frac{\phi(x+ty)-\phi(x)}{t} = \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x)-x), y\rangle,$$

i.e.,

$$d^-(-e_\lambda f)(x, y) = \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x)-x), y\rangle$$

and

$$\partial^-(-e_\lambda f)(x) = \left\{x^* \in \mathbb{R}^n \mid \langle x^*, y\rangle \leq \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x)-x), y\rangle,\ \forall\, y \in \mathbb{R}^n\right\}.$$

Clearly,

$$\partial^-(-e_\lambda f)(x) = \left\{x^* \in \mathbb{R}^n \mid \left\langle x^* + \frac{1}{\lambda}\nabla^2 g(x)x, y\right\rangle \le \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x)), y\rangle,\ \ \forall\, y \in \mathbb{R}^n\right\}.$$

By Theorem 8.24 in [5],

$$\partial^-(-e_\lambda f)(x) = \frac{1}{\lambda}\mathrm{cl}[\mathrm{conv}(\nabla^2 g(x)(P_\lambda f(x)))] - \frac{1}{\lambda}\nabla^2 g(x)x.$$

Since $P_\lambda f(x)$ is nonempty and compact, and $\nabla^2 g(x)$ is a linear continuous operator, we have that

$$\mathrm{conv}(\nabla^2 g(x)(P_\lambda f(x))) = \nabla^2 g(x)[\mathrm{conv}(P_\lambda f(x))]$$

is closed. Hence,

$$\partial^-(-e_\lambda f)(x) = \frac{1}{\lambda}\nabla^2 g(x)(\mathrm{conv}(P_\lambda f(x)) - x).$$

Since $P_\lambda f : \mathrm{int}(\mathrm{dom}\, g) \rightrightarrows \mathrm{dom}\, f \cap \mathrm{dom}\, g$ is upper semicontinuous and compact valued by Theorem 2.6, we see that $\mathrm{conv}\, P_\lambda f : \mathrm{int}(\mathrm{dom}\, g) \rightrightarrows \mathrm{conv}(\mathrm{dom}\, f \cap \mathrm{dom}\, g)$ is also upper semicontinuous, where $\mathrm{conv}\, P_\lambda f(x) = \mathrm{conv}(P_\lambda f(x))$, for all $x \in \mathrm{int}(\mathrm{dom}\, g)$ (see [31, Lemma 7.12]). Invoking now the continuity of $\nabla^2 g$, it follows that

$$\partial\phi(x) = \operatorname*{Limsup}_{z\to x} \partial^-\phi(x) = \frac{1}{\lambda}\nabla^2 g(x)(\mathrm{conv}(P_\lambda f(x)) - x).$$

We have shown that $e_\lambda f$ is locally Lipschitz on $\mathrm{int}(\mathrm{dom}\, g)$ in Theorem 2.4. Using [5, Theorem 8.49], we deduce that

$$\partial_C\phi(x) = \mathrm{conv}\partial\phi(x) = \partial\phi(x) = \frac{1}{\lambda}\nabla^2 g(x)(\mathrm{conv}(P_\lambda f(x)) - x) = \partial^-\phi(x),$$

and then

$$\phi^0(x, y) = \mathrm{d}^-\phi(x, y) = \max \frac{1}{\lambda}\langle \nabla^2 g(x)(P_\lambda f(x) - x), y\rangle,$$

which complete the proof. □

Corollary 3.1. *Suppose that $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function. If g is second-order continuously differentiable on $\mathrm{int}(\mathrm{dom}\, g)$ and $\nabla^2 g(x)$ is positive definite for each $x \in \mathrm{int}(\mathrm{dom}\, g)$, then, for any fixed $\lambda \in (0, \lambda_f)$, $e_\lambda f$ is continuously differentiable on $\mathrm{int}(\mathrm{dom}\, g)$ if and only if $P_\lambda f$ is single valued on $\mathrm{int}(\mathrm{dom}\, g)$.*

Proof. Suppose that $e_\lambda f$ is differentiable on $\mathrm{int}(\mathrm{dom}\, g)$. Then we have

$$\partial^-(-e_\lambda f)(x) = \{\nabla(-e_\lambda f)(x)\} = \frac{1}{\lambda}\nabla^2 g(x)(\mathrm{conv}(P_\lambda f(x)) - x).$$

Since g is second-order continuously differentiable on $\mathrm{int}(\mathrm{dom}\, g)$ and $\nabla^2 g(x)$ is positive definite, we have that $\nabla^2 g(x)$ is invertible. Hence,

$$\mathrm{conv}(P_\lambda f(x)) = x + \lambda(\nabla^2 g(x))^{-1}\nabla(-e_\lambda f)(x),$$

and hence $P_\lambda f$ is single valued.

On the other hand, if $P_\lambda f$ is single valued on $\mathrm{int}(\mathrm{dom}\, g)$, we have that

$$\partial^-(-e_\lambda f)(x) = \frac{1}{\lambda}\nabla^2 g(x)(P_\lambda f(x) - x)$$

is single valued. From Corollary 2.5, we have

$$\partial(-e_\lambda f)(x) = \mathrm{Limsup}_{z\to x}\partial^-(-e_\lambda f)(z) = \frac{1}{\lambda}\nabla^2 g(x)(P_\lambda f(x) - x),$$

which is a singleton. Noting that $e_\lambda f(x)$ is locally Lipschitz, by [5, Theorem 9.18(b)], we have that $(-e_\lambda f)$ is differentiable at x, and that $e_\lambda f$ is differentiable at x, with

$$\nabla e_\lambda f(x) = \frac{1}{\lambda}\nabla^2 g(x)(x - P_\lambda f(x)).$$

This, together with Corollary 2.5, implies that $\nabla e_\lambda f(x)$ is continuous. □

Corollary 3.2. *Suppose that $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper lower semicontinuous and prox-bounded function, that g is second-order continuously differentiable on* $\mathrm{int}(\mathrm{dom}\, g)$, *and that $\nabla^2 g(x)$ is positive definite for each $x \in \mathrm{int}(\mathrm{dom}\, g)$. If either f is strictly convex or f is convex and g is strictly convex, then, for any fixed $\lambda \in (0, \lambda_f)$, $P_\lambda f$ is single valued on* $\mathrm{int}(\mathrm{dom}\, g)$ *and $e_\lambda f$ is continuously differentiable on* $\mathrm{int}(\mathrm{dom}\, g)$.

Proof. If either f is strictly convex or f is convex and g is strictly convex, we can easily prove that $P_\lambda f$ is single valued. Hence, the result follows from Corollary 3.1. $\square$

4. Single-valuedness of proximal mapping $P_\lambda f$ and the convexity of $\lambda f + g$

In this section, we shall prove that the single-valuedness of proximal mapping $P_\lambda f$ is equivalent to the essential strict convexity of $\lambda f + g$ as well as the differentiability of $e_\lambda f \circ \nabla g^*$.

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

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

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

Proof. For any $x, y \in \mathrm{int}(\mathrm{dom}\, g)$ 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. $\square$

Corollary 4.1. *Suppose that f is a proper lower semicontinuous and prox-bounded function, and that g is Legendre. If $P_\lambda f$ is single valued, then $P_\lambda f \circ \nabla g^*$ is maximal monotone.*

Proof. Since g is Legendre and 1-coercive, ∇g^* is continuous on $\mathbb{R}^n$. It follows from Corollary 2.5 and Theorem 4.1 that $P_\lambda f \circ \nabla g^*$ is continuous and monotone on $\mathbb{R}^n$, and consequently maximal monotone. $\square$

Theorem 4.2. *Suppose that f is a proper lower semicontinuous and prox-bounded function, and that g is Legendre. Then, for any $\lambda \in (0, \lambda_f)$, $P_\lambda f \circ \nabla g^*$ is maximal monotone if and only if $P_\lambda f \circ \nabla g^* = \partial(\lambda f + g)^*(x^*)$, if and only if $\lambda f + g$ is convex.*

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

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

which is equivalent to

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

This implies that

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

and that

$$(\lambda f + g)^{**}(\bar{w}) = (\lambda f + g)(\bar{w}), \tag{4.15}$$

and

$$\nabla g(\bar{x}) \in \partial(\lambda f + g)^{**}(\bar{w}). \tag{4.16}$$

It follows from (4.16) 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

$$P_\lambda f \circ \nabla g^*(x^*) \subset \partial(\lambda f + g)^*(x^*) \quad \text{for all } x^* \in \mathbb{R}^n.$$

It follows that

$$P_\lambda f \circ \nabla g^*(x^*) = \partial(\lambda f + g)^*(x^*) \quad \text{for all } x^* \in \mathbb{R}^n, \tag{4.17}$$

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

It is clear that (4.17) is equivalent to

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

Hence,

$$\operatorname{ran} P_\lambda f = \operatorname{ran}\partial(\lambda f + g)^* = \operatorname{dom}\partial(\lambda f + g)^{**}.$$

This, together with (4.15), implies that

$$(\lambda f + g)^{**}(w) = (\lambda f + g)(w) \quad \text{for all } w \in \operatorname{dom}\partial(\lambda f + g)^{**}.$$

Noting that

$$\operatorname{dom}(\lambda f + g)^{**} \supset \operatorname{dom}\partial(\lambda f + g)^{**} \supset \operatorname{rint}(\operatorname{dom}(\lambda f + g)^{**}),$$

we have $(\lambda f + g)^{**}(w) = (\lambda f + g)(w)$ on $\operatorname{rint}(\operatorname{dom}(\lambda f + g)^{**})$. Since $(\lambda f + g)^{**}$ is proper convex lower semicontinuous on $\operatorname{rint}(\operatorname{dom}(\lambda f + g)^{**})$, $\lambda f + g$ is proper convex lower semicontinuous on $\operatorname{rint}(\operatorname{dom}(\lambda f + g)^{**})$. Since $\lambda f + g$ is completely determined by its values on $\operatorname{rint}(\operatorname{dom}(\lambda f + g)^{**})$, we obtain that $\lambda f + g$ is convex.

On the other hand, suppose that $\lambda f + g$ is convex. Then we have

$$\begin{aligned}
\bar{w} \in P_\lambda f(x) &\Leftrightarrow (\lambda f + g)(\bar{w}) - \langle \nabla g(x), \bar{w}\rangle \le (\lambda f + g)(w) - \langle \nabla g(x), w\rangle \quad \text{for all } w \in \mathbb{R}^n \\
&\Leftrightarrow \nabla g(x) \in \partial(\lambda f + g)(\bar{w}) \\
&\Leftrightarrow \bar{w} \in \partial(\lambda f + g)^* \circ \nabla g(x),
\end{aligned} \tag{4.18}$$

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

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

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

Theorem 4.3. *Suppose that f is a proper lower semicontinuous and prox-bounded function, and that g is Legendre. Then, for any fixed $\lambda \in (0, \lambda_f)$, the following statements are equivalent.*

(i) *$P_\lambda f$ is single valued on* $\operatorname{int}(\operatorname{dom} g)$.
(ii) *$P_\lambda f \circ \nabla g^*$ is single valued and maximal monotone.*
(iii) *$(\lambda f + g)^*$ is differentiable on $\mathbb{R}^n$.*
(iv) *$e_\lambda f \circ \nabla g^*$ is differentiable on $\mathbb{R}^n$.*
(v) *$\lambda f + g$ is essentially strictly convex.*

If, in addition, g is second-order continuously differentiable on $\operatorname{int}(\operatorname{dom} g)$ *and $\nabla^2 g(x)$ is positive definite for all $x \in$* $\operatorname{int}(\operatorname{dom} g)$, *then* (i)–(v)*are equivalent to the following assertion.*

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

Proof. (i) $\Leftrightarrow$ (ii) follows from Corollary 4.1, and ∇g is a topological isomorphism from $\operatorname{int}(\operatorname{dom} g)$ to $\mathbb{R}^n$ with $\nabla g^* = (\nabla g)^{-1}$. Theorem 4.2 implies that (ii) $\Leftrightarrow$ (iii). (iii) $\Leftrightarrow$ (iv) follows from (2.7), since g^* is differentiable on $\mathbb{R}^n$. By Theorem 4.2, we see that, if (ii) holds, then $\lambda f + g$ is proper convex lower semicontinuous and $(\lambda f + g)^*$ is differentiable on $\mathbb{R}^n$, which implies that $\lambda f + g$ is essentially strictly convex. If (v) holds, then, for any fixed $x \in \operatorname{int}(\operatorname{dom} g)$, the function

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

is essentially strictly convex too. Hence, $P_\lambda f(x) = \operatorname{argmin}_w \psi(w, x, \lambda)$ is nonempty and single valued by Theorem 2.2 and [10, Lemma 3.1]. (i) $\Leftrightarrow$ (vi) follows from Corollary 3.1. $\square$

Remark 3. The equivalent conditions (i)–(v) from Theorem 4.3 hold for all $\lambda \in (0, \lambda_f)$ if and only if $f + \frac{1}{\lambda_f} g$ (in the case $\lambda_f = +\infty$, reduces to f) is convex. Indeed, if (v) holds for all $\lambda \in (0, \lambda_f)$, then $f + \frac{1}{\lambda} g$ is convex, and then the result follows by taking limit as $\lambda \to \lambda_f$. Conversely, if $f + \frac{1}{\lambda_f} g$ is convex, then $f + \frac{1}{\lambda} g = (f + \frac{1}{\lambda_f} g) + (\frac{1}{\lambda} - \frac{1}{\lambda_f}) g$ is strictly convex for all $\lambda \in (0, \lambda_f)$, which implies that $P_\lambda f$ is a singleton. The equivalence among (i), (v) and (vi) has been proved by Wang [10] in the case when $g(x) = \frac{1}{2}\|x\|^2$.

Theorem 4.4. *Suppose that f is a proper lower semicontinuous and prox-bounded function. If $g + \lambda f$ is convex for some $\lambda \in (0, \lambda_f)$, then* $\operatorname{Fix} P_\lambda f = (\partial f)^{-1}(0) \cap \operatorname{int}(\operatorname{dom} g)$.

Proof. Since $g + \lambda f$ is convex, by (4.18), we have $P_\lambda f = (\partial(\lambda f + g))^{-1} \circ \nabla g$. Hence, $x \in P_\lambda f(x)$ if and only if $\nabla g(x) \in \partial(\lambda f + g)(x)$. Since $x \in \operatorname{int}(\operatorname{dom} g)$ and g is continuously differentiable on $\operatorname{int}(\operatorname{dom} g)$, by [30, Proposition 1.107], we have

$$\partial(\lambda f + g)(x) = \lambda \partial f(x) + \nabla g(x).$$

It follows that $x \in P_\lambda f(x)$ if and only if $0 \in \partial f(x)$. That is, $\operatorname{Fix} P_\lambda f = (\partial f)^{-1}(0) \cap \operatorname{int}(\operatorname{dom} g)$. □

Theorem 4.4 implies that each fixed point of $P_\lambda f$ is a critical point of f in the case when $\lambda f + g$ is convex for some $\lambda \in (0, \lambda_f)$, and a minimizer of f in the case when f is convex. Therefore, we may find the critical point via finding a fixed point of $P_\lambda f$. We assume that $\operatorname{dom} f \cap \operatorname{dom} g \subset \operatorname{int}(\operatorname{dom} g)$, and propose the next algorithm, following the idea from [8].

Algorithm 1. *Step* 0. Let $\lambda \in (0, \lambda_f)$ be such that $g + \lambda f$ is convex. Let $x_0 \in \operatorname{int}(\operatorname{dom} g)$ and $\mu > 0$ be given, and set $k = 1$.
Step 1. Set

$$y^k = x^k + \mu \nabla g(x^k), \tag{4.19}$$

$$x^{k+1} = (I + \mu \partial(\lambda f + g))^{-1}(y^k). \tag{4.20}$$

Step 2. If $x^{k+1} = x^k$, then stop; otherwise, set $k = k + 1$ and go back to Step 1.

In view of the convexity of $\lambda f + g$ and $\operatorname{dom} f \cap \operatorname{dom} g \subset \operatorname{int}(\operatorname{dom} g)$, it is easy to see that the sequences $\{x_n\}$ and $\{y_n\}$ are well defined.

Proposition 4.1 (*See [8]*). *Let $\lambda \in (0, \lambda_f)$ be such that $\lambda f + g$ is convex, and let $\mu > 0$. Then, for $x \in \operatorname{int}(\operatorname{dom} g)$, $x \in P_\lambda f(x)$ if and only if $x = (I + \mu \partial(\lambda f + g))^{-1}(x + \mu \nabla g(x))$.*

Proof.

$$\begin{aligned} x = \ & (I + \mu \partial(\lambda f + g))^{-1}(x + \mu \nabla g(x)) \\ & \Leftrightarrow x + \mu \nabla g(x) \in x + \mu \partial(\lambda f + g)(x) \\ & \Leftrightarrow \nabla g(x) \in \partial(\lambda f + g)(x) \\ & \Leftrightarrow x \in P_\lambda f(x). \quad \square \end{aligned}$$

Theorem 4.5 (*See [8]*). *Suppose that $f^* := \inf_{x \in \mathbb{R}^n} f(x)$ is finite.*

(i) *If Algorithm 1 stops after the k steps, i.e., $x^k = x^{k+1}$, then $x^k \in P_\lambda f(x^k)$;*
(ii) *If the sequence $\{x^k\}$ generated by Algorithm 1 is infinite, then $f(x^k) \to f^*$ as $k \to \infty$. Moreover, if $\{x^k\}$ is bounded, then any cluster point of $\{x^k\}$ is a minimizer of f.*

Proof. (i) This follows from Proposition 4.1.
(ii) From the definition of $\{x^k\}$, we have that

$$\frac{1}{\mu}(x^k - x^{k+1}) + \nabla g(x^k) \in \partial(\lambda f + g)(x^{k+1}).$$

It follows that

$$(\lambda f + g)(x) - (\lambda f + g)(x^{k+1}) \geq \left\langle \frac{1}{\mu}(x^k - x^{k+1}) + \nabla g(x^k), x - x^{k+1} \right\rangle, \quad \forall x \in \mathbb{R}^n,$$

or, equivalently,

$$\lambda f(x) - \lambda f(x^{k+1}) \geq g(x^{k+1}) - g(x) + \langle \nabla g(x^k), x - x^{k+1} \rangle + \frac{1}{\mu} \langle x^k - x^{k+1}, x - x^{k+1} \rangle, \quad \forall x \in \mathbb{R}^n. \tag{4.21}$$

Letting $x = x^k$ in (4.21) and noting that $g(x^{k+1}) \geq g(x^k) - \langle \nabla g(x^k), x^{k+1} - x^k \rangle$, we get

$$\lambda[f(x^k) - f(x^{k+1})] \geq \frac{1}{\mu}\|x^k - x^{k+1}\|^2 > 0. \tag{4.22}$$

This implies that $\{f(x^k)\}$ is strictly decreasing and has a lower bound f^*. Hence, $f(x^k) \to f^*$ as $k \to \infty$. The last conclusion is obvious. □

Acknowledgment

The research of the second author was partly supported by the National Natural Sciences Grant (No. 11071052).

References

[1] J.-J. Moreau, Proximit et dualit dans un espace Hilbertien, Bull. Soc. Math. France 93 (1965) 273–299.
[2] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976) 877–898.
[3] H. Attouch, Variational Convergence for Functions and Operators, Pitman, Boston, 1984.
[4] J.-B. Hiriart-Urruty, C. Lemarechal, Convex analysis and Minimization Algorithm I, Springer, New York, 1996.
[5] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer-Verlag, New York, 1998.
[6] F. Bernard, L. Thibault, Prox-regularity of functions and sets in Banach spaces, Set-valued Anal. 12 (2004) 25–47.
[7] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Model. Simul. 4 (2005) 1168–C1200.
[8] A. Hamdi, A Moreau-Yosida regularization of a difference of two convex functions, Appl. Math. E-Notes 5 (2005) 164–170.
[9] W.L. Hare, C. Sagastizbal, Computing proximal points of nonconvex functions, Math. Program. 116 (2009) 221–258.
[10] X.F. Wang, On Chebyshev functions and Klee functions, J. Math. Anal. Appl. 1 (2010) 293–310.
[11] M. Bačák, J.M. Borwein, A. Eberhard, B.S. Mordukhovich, Infimal convolutions and Lipschitian properties of subdifferentials for prox-regular functions in Hilbert spaces, J. Convex Analysis 17 (2010) 732–763.
[12] V. Klee, Convexity of Chebyshev Sets, Math. Ann. 142 (1961) 292–304.
[13] N.W. Efimow, S.B. Stechkin, Approximative compactness and Chebyshev sets, Sov. Math. 2 (1961) 1226–1228.
[14] V.S. Balaganski, L.P. Vlasov, The problem of the convexity of Chebyshev sets, Russian Math. Surveys 51 (1996) 1127–1190.
[15] F. Deutsch, Best Approximations in inner product spaces, in: CMS Books Math./Ouvrages Math., in: SMC, vol. 7, Springer-Verlag, New York, 2001.
[16] J.M. Borwein, Proximality and Chebyshev sets, Optim. Lett. 1 (2007) 21–32.
[17] J.-B. Hiriart-Urruty, Potpourri of conjectures and open questions in nonlinear analysis and optimization, SIAM Rev. 49 (2007) 255–273.
[18] H. Busemann, Note on a theorem on convex sets, Matem. Tidsskr B (1947) 32–34.
[19] 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 (3) (1967) 200–217.
[20] D. Butnariu, Y. Censor, S. Reich, Iterative averaging of entropic projections for solving stochastic convex feasibility problems, Comput. Optim. Appl. 8 (1997) 21–39.
[21] H.H. Bauschke, J.M. Borwein, Legendre functions and the method of random Bregman projections, J. Convex Anal. 4 (1997) 27–67.
[22] H.H. Bauschke, J.M. Borwein, P.L. Combettes, Bregman monotone optimization algorithms, SIAM J. Control Optim. 42 (2003) 596–636.
[23] E. Resmerita, On total convexity, Bregman projections and stability in Banach spaces, J. Convex Anal. 11 (2004) 1–16.
[24] H.H. Bauschke, P.L. Combettes, D. Noll, Joint minimization with alternating Bregman proximity operators, Pac. J. Optim. 2 (2006) 401–424.
[25] H.H. Bauschke, Xianfu Wang, J. Ye, Xiaoming Yuan, Bregman distances and Chebyshev sets, J. Approx. Theory 159 (2009) 3–25.
[26] C. Li, W. Song, J-C. Yao, The Bregman distance, approximate compactness and convexity of Chebyshev sets in Banach spaces, J. Approx. Theory 162 (2010) 1128–1149.
[27] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.
[28] E. Asplund, R.T. Rockafella, Gradients of convex functions, Trans. Amer. Math. Soc. 139 (1969) 443–467.
[29] M. Bougeard, J.-P. Penot, A. Pommellet, Towards minimal assumptions for the infimal convolution regularization, J. Math. Anal. Appl. 64 (1991) 245–270.
[30] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation, I. Basic theory, Springer-Verlag, Berlin, 2006.
[31] R.R. Phelps, Convex Functions Monotone Operators and Differentiability, 2nd edition, in: Lecture Notes in Mathematics, vol. 1364, Springer-Verlag, 1993.