Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 159 (2009) 3–25
Journal of Approximation Theory
www.elsevier.com/locate/jat

Bregman distances and Chebyshev sets

Heinz H. Bauschke[a,*], Xianfu Wang[a], Jane Ye[b], Xiaoming Yuan[c]

[a] *Mathematics, Irving K. Barber School, The University of British Columbia Okanagan, Kelowna, B.C. V1V 1V7, Canada*
[b] *Department of Mathematics and Statistics, University of Victoria, Victoria, B.C. V8W 3P4, Canada*
[c] *Department of Mathematics, Hong Kong Baptist University, PR China*

Received 24 December 2007; received in revised form 24 July 2008; accepted 1 August 2008
Available online 13 November 2008

Communicated by C.K. Chui and H.N. Mhaskar

Dedicated to the memory of G.G. Lorentz

Abstract

A closed set of a Euclidean space is said to be Chebyshev if every point in the space has one and only one closest point in the set. Although the situation is not settled in infinite-dimensional Hilbert spaces, in 1932 Bunt showed that in Euclidean spaces a closed set is Chebyshev if and only if the set is convex. In this paper, from the more general perspective of Bregman distances, we show that if every point in the space has a unique nearest point in a closed set, then the set is convex. We provide two approaches: one is by nonsmooth analysis; the other by maximal monotone operator theory. Subdifferentiability properties of Bregman nearest distance functions are also given.

Keywords: Bregman distance; Bregman projection; Chebyshev set with respect to a Bregman distance; Legendre function; Maximal monotone operator; Nearest point; Subdifferential operators

1. Introduction

Throughout the paper, $\mathbb{R}^J$ is the standard Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$, and Γ is the set of proper lower semicontinuous convex functions on $\mathbb{R}^J$. Let C

* Corresponding address: Mathematics, Irving K. Barber School, The University of British Columbia Okanagan, Room 108, Science Building, Kelowna, B.C. V1V 1V7, Canada.
E-mail addresses: heinz.bauschke@ubc.ca (H.H. Bauschke), shawn.wang@ubc.ca (X. Wang), janeye@math.uvic.ca (J. Ye), xmyuan@hotmail.com (X. Yuan).

0021-9045/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jat.2008.08.014

be a nonempty closed subset of $\mathbb{R}^J$. If each $x \in \mathbb{R}^J$ has a unique nearest point in C, the set C is said to be Chebyshev. The famous Chebyshev set problem inquires: "Is a Chebyshev set necessarily convex?". It has been studied by many authors, see [1,6,14,16,7,17–19,26,27,29] and the references therein. Although answered in the affirmative by Bunt in 1932 [8] (see also [22]), we look at the problem from the more general point of view of Bregman distances.

Let

$$f:\mathbb{R}^J \to]-\infty,+\infty] \text{ be convex and differentiable on } U := \operatorname{int}\operatorname{dom} f \neq \varnothing. \tag{1}$$

The *Bregman distance* associated with f is defined by

$$D:\mathbb{R}^J \times \mathbb{R}^J \to [0,+\infty]:(x,y) \mapsto \begin{cases} f(x)-f(y)-\langle \nabla f(y), x-y\rangle, & \text{if } y \in U; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2}$$

Assume that $C \subset U$. It is a natural generalization of the Chebyshev problem to ask the following:

> "If every $x \in U$ has – in terms of the Bregman distance – a unique nearest point in C, i.e., C is Chebyshev for the Bregman distance, must C be convex?"

We give two approaches to our affirmative answer: one uses beautiful properties of maximal monotone operators: Rockafellar's virtual convexity theorem on ranges of maximal monotone operators; the other uses generalized subdifferentials from nonsmooth analysis, which allows us to characterize Chebyshev sets. We also study subdifferentiabilities of Bregman distance functions associated to closed sets. These nonsmooth analysis results are interesting in their own right, since Bregman distances have found tremendous applications in Statistics, Engineering, and Optimization; see the recent books [9,11] and the references therein.

The function D does not define a metric, since it is not symmetric and does not satisfy the triangle inequality. It is thus remarkable that it is not only possible to derive many results on projections and distances similar to the one obtained in finite-dimensional Euclidean spaces, but also to provide a general framework for best approximations.

The paper is organized as follows. In Section 2, we state our assumptions on f and provide some concrete choices. In Section 3, we characterize left Bregman nearest points and geodesics. We show that the Bregman normal is a proximal normal. In Section 4, when f is Legendre and 1-coercive and C is Chebyshev, we show that the composition of the Bregman nearest point map and ∇f^* is maximal monotone. This allows us to apply Rockafellar's theorem on virtual convexity of range of maximal monotone operator to obtain that a Chebyshev set is convex. In Section 5, we study subdifferentiability properties of the left Bregman distance function. Formulas for the Clarke subdifferential, the limiting subdifferential and the Dini subdifferential are given. In Section 6, we give a complete characterizations of Chebyshev sets. Our approach generalizes the results given by Hiriart-Urruty [19,17] from the Euclidean to the Bregman setting. Finally, in Section 7 we show that the convexity of Chebyshev sets for right Bregman projections of f can be studied by using the left Bregman projections of f^*. We give an example showing that even if the right Bregman projection is single-valued, the set C need not be convex.

Notation: In $\mathbb{R}^J$, the closed ball centered at x with radius $\delta > 0$ is denoted by $B_\delta(x)$ and the closed unit ball is $B = B_1(0)$. For a set S, the expressions $\operatorname{int} S$, $\operatorname{cl} S$, $\operatorname{conv} S$ signify the interior, closure, and convex hull of S respectively. For a set-valued mapping $T : \mathbb{R}^J \rightrightarrows \mathbb{R}^J$, we use $\operatorname{ran} T$ and $\operatorname{dom} T$ for its range and domain, and T^{-1} for its set-valued inverse, i.e., $x \in T^{-1}(y) \Leftrightarrow y \in T(x)$. For a function $f : \mathbb{R}^J \to]-\infty,+\infty]$, $\operatorname{dom} f$ is the domain of f, and f^* is its Fenchel conjugate; $\operatorname{conv} f$ ($\operatorname{cl}\operatorname{conv} f$) denotes the convex hull (closed convex hull) of f.

For a differentiable function f, $\nabla f(x)$ and $\nabla^2 f(x)$ denote the gradient vector and the Hessian matrix at x. Our notation is standard and follows, e.g., [23,24].

2. Standing assumptions and examples

From now on, and until the end of Section 6, our standing assumptions on f and C are:

(A1) $f \in \Gamma$ is a convex function of Legendre type, i.e., f is essentially smooth and essentially strictly convex in the sense of [23, Section 26].
(A2) f is 1-coercive, i.e., $\lim_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$. An equivalent requirement is $\operatorname{dom} f^* = \mathbb{R}^J$ (see [24, Theorem 11.8(d)]).
(A3) The set C is a nonempty closed subset of U.

Important instances of functions satisfying the above conditions are:

Example 2.1. Let $x = (x_j)_{1\le j\le J}$ and $y = (y_j)_{1\le j\le J}$ be two points in $\mathbb{R}^J$.

(i) *(Halved) Energy:* If $f\colon x \mapsto \frac{1}{2}\|x\|^2$, then $U = \mathbb{R}^J$,

$$D(x, y) = \frac{1}{2}\|x - y\|^2,$$

and $\nabla^2 f(x) = \mathrm{Id}$ for every $x \in \mathbb{R}^J$. Note that $f^*(x) = \frac{1}{2}\|x\|^2$, $\operatorname{dom} f^* = \mathbb{R}^J$, and $\nabla^2 f^* = \mathrm{Id}$.

(ii) *(Negative) Boltzmann–Shannon Entropy:* If $f\colon x \mapsto \sum_{j=1}^J x_j \ln(x_j) - x_j$ for $x \ge 0$, $+\infty$ otherwise (where $x \ge 0$ means $x_j \ge 0$ for $1 \le j \le J$ and similarly for $x > 0$, and $0 \ln 0 = 0$), then $U = \{x \in \mathbb{R}^J : x > 0\}$, and

$$D(x, y) = \begin{cases} \displaystyle\sum_{j=1}^J x_j \ln(x_j/y_j) - x_j + y_j, & \text{if } x \ge 0 \text{ and } y > 0;\\ +\infty, & \text{otherwise} \end{cases}$$

is the so-called *Kullback–Leibler divergence*. Note that

$$\nabla^2 f(x) = \begin{pmatrix} 1/x_1 & 0 & \cdots & 0 & 0\\ 0 & 1/x_2 & 0 & \cdots & 0\\ \vdots & 0 & \ddots & 0 & 0\\ 0 & \ldots & 0 & 1/x_{J-1} & 0\\ 0 & \ldots & 0 & 0 & 1/x_J \end{pmatrix},$$

that $f^*(x) = \sum_{j=1}^J \mathrm{e}^{x_j}$ with $\operatorname{dom} f^* = \mathbb{R}^J$, and that

$$\nabla^2 f^*(x) = \begin{pmatrix} \mathrm{e}^{x_1} & 0 & \cdots & 0 & 0\\ 0 & \mathrm{e}^{x_2} & 0 & \cdots & 0\\ \vdots & 0 & \ddots & 0 & 0\\ 0 & \ldots & 0 & \mathrm{e}^{x_{J-1}} & 0\\ 0 & \ldots & 0 & 0 & \mathrm{e}^{x_J} \end{pmatrix}.$$

(iii) *(Negative) Fermi–Dirac Entropy*: If $f : x \mapsto \sum_{j=1}^J x_j \ln x_j + (1 - x_j)\ln(1 - x_j)$, then $U = \{x \in \mathbb{R}^J : 0 < x < 1\}$ and

$$D(x,y)=\begin{cases}\displaystyle\sum_{j=1}^{J} x_j\ln(x_j/y_j)+(1-x_j)\ln((1-x_j)/(1-y_j)),\\ \quad\text{if } 1\geq x\geq 0 \text{ and } 1>y>0;\\ +\infty, \quad \text{otherwise.}\end{cases}$$

While

$$\nabla^2 f(x)=\begin{pmatrix}\frac{1}{x_1(1-x_1)} & 0 & \cdots & 0\\ 0 & \frac{1}{x_2(1-x_2)} & 0 & 0\\ \vdots & 0 & \ddots & 0\\ 0 & \cdots & 0 & \frac{1}{x_J(1-x_J)}\end{pmatrix},$$

$\forall 0<x<1, x\in\mathbb{R}^J$,

we have $f^*(x)=\sum_{j=1}^{J}\ln(1+e^{x_j})$ with

$$\nabla^2 f^*(x)=\begin{pmatrix}\frac{e^{x_1}}{(1+e^{x_1})^2} & 0 & \cdots & 0\\ 0 & \frac{e^{x_2}}{(1+e^{x_2})^2} & 0 & 0\\ \vdots & 0 & \ddots & 0\\ 0 & \cdots & 0 & \frac{e^{x_J}}{(1+e^{x_J})^2}\end{pmatrix}, \quad \forall x\in\mathbb{R}^J.$$

(iv) In general, we can let $f: x\mapsto\sum_{i=1}^{J}\phi(x_i)$ where $\phi:\mathbb{R}\to]-\infty,+\infty]$ is a Legendre function. Then $U=(\operatorname{int}\operatorname{dom}\phi)^J$,

$$D(x,y)=\sum_{j=1}^{J}\phi(x_j)-\phi(y_j)-\phi'(y_j)(x_j-y_j), \quad \forall\, x\in\mathbb{R}^J, y\in U.$$

In particular, one can use $\phi(t)=|t|^p/p$ with $p>1$.

The following result (see [23, Theorem 26.5]) plays an important role in what follows.

Fact 2.2 (*Rockafellar*). *A convex function f is of Legendre type if and only if f^* is. In this case, the gradient mapping*

$$\nabla f: U\to\operatorname{int}\operatorname{dom} f^*: x\mapsto\nabla f(x),$$

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

3. Bregman distances and projection operators

We start with:

Definition 3.1. The *left Bregman nearest distance function* to C is defined by

$$\overleftarrow{D}_C: U\to[0,+\infty]: y\mapsto\inf_{x\in C} D(x,y), \tag{3}$$

and the *left Bregman nearest point map* (i.e., the classical Bregman projector) onto C is

$$\overleftarrow{P}_C : U \to U : y \mapsto \underset{x \in C}{\operatorname{argmin}}\ D(x, y) = \{x \in C : D(x, y) = \overleftarrow{D}_C(y)\}.$$

The *right Bregman distance* and *right Bregman projector* onto C are defined analogously and denoted by $\overrightarrow{D}_C$ and $\overrightarrow{P}_C$, respectively. Note that while in [4] the authors consider proximity operators associated with convex set C, here our set C need not be convex and we do not assume that $D(\cdot, \cdot)$ is jointly convex.

We shall often need the following identity

$$D(c, y) - D(x, y) = f(c) - f(x) - \langle \nabla f(y), c - x\rangle, \quad \forall c \in U, x \in U, y \in U \tag{4}$$

which is an immediate consequence of the definition.

Our first result characterizes the left Bregman nearest point.

Proposition 3.2. *Let $x \in C$ and $y \in U$.*

(i) *Then*

$$x \in \overleftarrow{P}_C(y) \quad \Leftrightarrow \quad D(c, x) \geq \langle \nabla f(y) - \nabla f(x), c - x\rangle \quad \forall c \in C. \tag{5}$$

If C is convex, then

$$x \in \overleftarrow{P}_C(y) \quad \Leftrightarrow \quad 0 \geq \langle \nabla f(y) - \nabla f(x), c - x\rangle \quad \forall c \in C. \tag{6}$$

(ii) *Suppose that $x \in \overleftarrow{P}_C(y)$. Then the Bregman projection of*

$$z_\lambda = \nabla f^*(\lambda \nabla f(y) + (1 - \lambda)\nabla f(x)) \quad \text{with } 0 \leq \lambda < 1, \tag{7}$$

on C is singleton with

$$\overleftarrow{P}_C(z_\lambda) = x. \tag{8}$$

If C is convex, (8) holds for every $\lambda \geq 0$.

Proof. (i): By definition, $x \in \overleftarrow{P}_C(y)$ if and only if

$$0 \leq D(c, y) - D(x, y) \quad \forall c \in C;$$

equivalently, $f(c) - f(x) \geq \langle \nabla f(y), c - x\rangle$ by (4). Subtracting $\langle \nabla f(x), c - x\rangle$ from both sides, we obtain

$$D(c, x) \geq \langle \nabla f(y) - \nabla f(x), c - x\rangle.$$

Hence (5) holds.

The convex counterpart (6) is well known and follows, e.g., from [2, Proposition 3.16].

(ii) Assume that $x \in \overleftarrow{P}_C(y)$ and $z_\lambda = \nabla f^*(\lambda \nabla f(y) + (1 - \lambda)\nabla f(x))$ with $0 \leq \lambda < 1$. Then by (5),

$$D(c, x) \geq \langle \nabla f(y) - \nabla f(x), c - x\rangle \quad \forall c \in C. \tag{9}$$

Take $c \in C$. By Fact 2.2, $\nabla f \circ \nabla f^* = \mathrm{Id}$, we have

$$\langle \nabla f(z_\lambda) - \nabla f(x), c - x\rangle \tag{10}$$
$$= \langle \nabla f \circ \nabla f^*(\lambda \nabla f(y) + (1 - \lambda)\nabla f(x)) - \nabla f(x), c - x\rangle \tag{11}$$
$$= \langle (\lambda \nabla f(y) + (1 - \lambda)\nabla f(x)) - \nabla f(x), c - x\rangle \tag{12}$$
$$= \lambda \langle \nabla f(y) - \nabla f(x), c - x\rangle. \tag{13}$$

If $\langle \nabla f(y) - \nabla f(x), c - x \rangle \leq 0$, then

$$\lambda \langle \nabla f(y) - \nabla f(x), c - x \rangle \leq 0 \leq D(c, x);$$

if $\langle \nabla f(y) - \nabla f(x), c - x \rangle \geq 0$, then using $0 \leq \lambda < 1$ and (9),

$$\lambda \langle \nabla f(y) - \nabla f(x), c - x \rangle \leq \langle \nabla f(y) - \nabla f(x), c - x \rangle \leq D(c, x).$$

In either case, by (10) we have

$$\langle \nabla f(z_\lambda) - \nabla f(x), c - x \rangle \leq D(c, x) \quad \forall c \in C.$$

Hence $x \in \overleftarrow{P}_C(z_\lambda)$ by (5). We proceed to show that $\overleftarrow{P}_C(z_\lambda)$ is a singleton. If $\lambda = 0$, then $z_\lambda = x$, $\overleftarrow{P}_C(x) = \{x\}$ by strict convexity of f. It remains to consider the case $0 < \lambda < 1$. Let $\hat{x} \in \overleftarrow{P}_C(z_\lambda)$. Then $D(x, z_\lambda) = D(\hat{x}, z_\lambda)$, which is

$$f(\hat{x}) - f(x) - \langle \nabla f(z_\lambda), \hat{x} - x \rangle = 0,$$

by (4). Using $z_\lambda = \nabla f^*(\lambda \nabla f(y) + (1 - \lambda) \nabla f(x))$, we have

$$f(\hat{x}) - f(x) - \langle \lambda \nabla f(y) + (1 - \lambda) \nabla f(x), \hat{x} - x \rangle = 0,$$

so that

$$\lambda[f(\hat{x}) - f(x) - \langle \nabla f(y), \hat{x} - x \rangle] + (1 - \lambda)[f(\hat{x}) - f(x) - \langle \nabla f(x), \hat{x} - x \rangle] = 0,$$

and

$$\lambda[f(x) - f(\hat{x}) - \langle \nabla f(y), x - \hat{x} \rangle] = (1 - \lambda)[f(\hat{x}) - f(x) - \langle \nabla f(x), \hat{x} - x \rangle].$$

This gives, by (4), $\lambda[D(x, y) - D(\hat{x}, y)] = (1 - \lambda) D(\hat{x}, x)$ and hence

$$D(x, y) - D(\hat{x}, y) = \frac{1 - \lambda}{\lambda} D(\hat{x}, x),$$

since $1 > \lambda > 0$. If $\hat{x} \neq x$, then $D(\hat{x}, x) > 0$ by the strict convexity of f so that $D(x, y) > D(\hat{x}, y)$, and this contradicts that $x \in \overleftarrow{P}_C(y)$. Therefore, $\overleftarrow{P}_C(z_\lambda) = \{x\}$.

When C is convex, by (6), $x \in \overleftarrow{P}_C(y)$ if and only if

$$\langle \nabla f(y) - \nabla f(x), c - x \rangle \leq 0, \quad \forall c \in C. \tag{14}$$

If $z_\lambda = \nabla f^*(\lambda \nabla f(y) + (1 - \lambda) \nabla f(x))$ with $\lambda \geq 0$, then

$$\begin{aligned} &\langle \nabla f(z_\lambda) - \nabla f(x), c - x \rangle \\ &\quad = \langle \nabla f \circ \nabla f^*(\lambda \nabla f(y) + (1 - \lambda) \nabla f(x)) - \nabla f(x), c - x \rangle \qquad (15) \\ &\quad = \lambda \langle \nabla f(y) - \nabla f(x), c - x \rangle \leq 0. \qquad (16) \end{aligned}$$

By (6), $x \in \overleftarrow{P}_C(z_\lambda)$. Applying (8), we see that $x = \overleftarrow{P}_C(z_\lambda)$. Indeed, select $\lambda_1 > \lambda$. Since

$$z_\lambda = \nabla f^*(\lambda \nabla f(y) + (1 - \lambda) \nabla f(x)) \Rightarrow \nabla f(z_\lambda) = \nabla f(x) + \lambda(\nabla f(y) - \nabla f(x)), \tag{17}$$

$$\begin{aligned} z_{\lambda_1} &= \nabla f^*(\lambda_1 \nabla f(y) + (1 - \lambda_1) \nabla f(x)) \Rightarrow \nabla f(z_{\lambda_1}) \\ &= \nabla f(x) + \lambda_1(\nabla f(y) - \nabla f(x)). \qquad (18) \end{aligned}$$

Solve (18) for $\nabla f(y) - \nabla f(x)$ and put into (17) to get

$$\nabla f(z_\lambda) = \left(1 - \frac{\lambda}{\lambda_1}\right) \nabla f(x) + \frac{\lambda}{\lambda_1} \nabla f(z_{\lambda_1}).$$

This gives

$$z_\lambda = \nabla f^*((1-\lambda/\lambda_1)\nabla f(x) + \lambda/\lambda_1 \nabla f(z_{\lambda_1})).$$

As $x \in \overleftarrow{P}_C(z_{\lambda_1})$, (8) applies. ■

It is interesting to point out a connection to the *proximal normal cone* $N_C^P(x)$ of C at $x \in C$ Recall that

$$N_C^P(x) := \{t(y-x) : \ t \geq 0, x \in \ P_C(y), y \in \mathbb{R}^J\},$$

in which P_C denotes the usual projection on C in terms of the Euclidean norm, and each vector $t(y-x)$ is called a proximal normal to C at x; see, e.g., [13, Section 1.1] for further information.

Proposition 3.3. *Suppose that f is twice continuously differentiable on U, let $y \in U$, and suppose that $x \in \overleftarrow{P}_C(y)$. Then $\nabla f(y) - \nabla f(x) \in N_C^P(x)$.*

Proof. By Proposition 3.2(i),

$$D(c,x) \geq \langle \nabla f(y) - \nabla f(x), c-x\rangle \quad \forall c \in C. \tag{19}$$

Since the Hessian of f is continuous, using Taylor's formula, we obtain

$$D(c,x) = f(c) - f(x) - \langle \nabla f(x), c-x\rangle = \frac{1}{2}\langle c-x, \nabla^2 f(\xi)(c-x)\rangle$$
$$\text{where } \xi \in [c,x]. \tag{20}$$

Fix $\delta > 0$. Since $\nabla^2 f$ is continuous on the compact set $C \cap B_\delta(x)$, there exists $\sigma = \sigma(x,\delta) > 0$ such that $\|\nabla^2 f(\xi)\| \leq 2\sigma$ for every $\xi \in C \cap B_\delta(x)$. Then (20) gives $D(c,x) \leq \sigma\|c-x\|^2$. By (19),

$$\sigma\|c-x\|^2 \geq \langle \nabla f(y) - \nabla f(x), c-x\rangle \quad \forall c \in C \cap B_\delta(x).$$

By [13, Proposition 1.1.5.(b) on page 25], $\nabla f(y) - \nabla f(x) \in N_C^P(x)$. ■

The following example illustrates the geodesics $\{z_\lambda : \ 0 \leq \lambda \leq 1\}$ given by (7).

Example 3.4. Let $x = (x_j)_{1\leq j\leq J}$ and $y = (y_j)_{1\leq j\leq J}$ be two points in $\mathbb{R}^J$.

(i) If $f: x \mapsto \frac{1}{2}\|x\|^2$, then $\nabla f = \nabla f^* = \mathrm{Id}$. We have

$$z_\lambda = \lambda y + (1-\lambda)x,$$

for $\lambda \in [0,1]$. Hence z_λ is a componentwise *arithmetic mean* of x and y.

(ii) If $f: x \mapsto \sum_{j=1}^J x_j \ln(x_j) - x_j$, then

$$\nabla f(x) = (\ln x_1, \ldots, \ln x_n),$$
$$f^*: x^* \mapsto \sum_{j=1}^J \exp(x_j^*),$$

so that

$$\nabla f^*(x^*) = (\exp x_1^*, \ldots, \exp x_J^*).$$

We have

$$\begin{aligned} z_\lambda &= \nabla f^*(\lambda \nabla f(y) + (1-\lambda)\nabla f(x)) && (21)\\ &= \nabla f^*(\lambda \ln y_1 + (1-\lambda)\ln x_1, \ldots, \lambda \ln y_J + (1-\lambda)\ln x_J) && (22)\\ &= (\exp(\lambda \ln y_1 + (1-\lambda)\ln x_1), \ldots, \exp(\lambda \ln y_J + (1-\lambda)\ln x_J)) && (23)\\ &= (y_1^\lambda x_1^{1-\lambda}, \ldots, y_J^\lambda x_J^{1-\lambda}). && (24) \end{aligned}$$

Hence z_λ is a componentwise *geometric mean* of x and y.

(iii) If $f: x \mapsto \sum_{j=1}^J \exp(x_j)$, then $f^*: x^* \mapsto \sum_{j=1}^J x_j^* \ln(x_j^*) - x_j^*$ so that

$$\nabla f(x) = (\exp(x_1), \ldots, \exp(x_J)), \quad \nabla f^*(x^*) = (\ln x_1^*, \ldots, \ln x_J^*).$$

Hence

$$z_\lambda = (\ln(\lambda \exp(y_1) + (1-\lambda)\exp(x_1)), \ldots, \ln(\lambda \exp(y_J) + (1-\lambda)\exp(x_J))).$$

Define the *symmetrization* of D for $x, y \in U$ by

$$S(x, y) := D(x, y) + D(y, x) = \langle \nabla f(x) - \nabla f(y), x - y\rangle.$$

Proposition 3.5. *Given $x, y \in U$ and $0 < \lambda < 1$, set*

$$z_\lambda := \nabla f^*(\lambda \nabla f(y) + (1-\lambda)\nabla f(x)).$$

Then we have:

(i) $D(x, y) = D(x, z_\lambda) + D(z_\lambda, y) + \frac{1-\lambda}{\lambda} S(x, z_\lambda)$.
(ii) $S(x, y) = \frac{1}{1-\lambda} S(y, z_\lambda) + \frac{1}{\lambda} S(z_\lambda, x)$.

Proof. Since $z_\lambda = \nabla f^*(\lambda \nabla f(y) + (1-\lambda)\nabla f(x))$, and

$$D(x, z_\lambda) = f(x) - f(z_\lambda) - \langle \nabla f(z_\lambda), x - z_\lambda\rangle,$$

we have

$$\begin{aligned} D(x, z_\lambda) &= f(x) - f(z_\lambda) - \langle \lambda \nabla f(y) + (1-\lambda)\nabla f(x), x - z_\lambda\rangle && (25)\\ &= \lambda[f(x) - f(z_\lambda) - \langle \nabla f(y), x - z_\lambda\rangle] \\ &\quad + (1-\lambda)[f(x) - f(z_\lambda) - \langle \nabla f(x), x - z_\lambda\rangle] && (26)\\ &= \lambda[D(x, y) - D(z_\lambda, y)] - (1-\lambda)[f(z_\lambda) - f(x) - \langle \nabla f(x), z_\lambda - x\rangle] && (27)\\ &= \lambda[D(x, y) - D(z_\lambda, y)] - (1-\lambda)D(z_\lambda, x). && (28) \end{aligned}$$

Hence $(1-\lambda)[D(x, z_\lambda) + D(z_\lambda, x)] + \lambda D(x, z_\lambda) + \lambda D(z_\lambda, y) = \lambda D(x, y)$. Dividing both sides by λ yields

$$D(x, y) = D(x, z_\lambda) + D(z_\lambda, y) + \frac{1-\lambda}{\lambda} S(x, z_\lambda), \tag{29}$$

which is (i).

To see (ii), we rewrite

$$z_\lambda = \nabla f^*((1-\lambda)\nabla f(x) + \lambda \nabla f(y)).$$

Applying (i), we get

$$D(y, x) = D(y, z_\lambda) + D(z_\lambda, x) + \frac{\lambda}{1-\lambda} S(z_\lambda, y). \tag{30}$$

Adding (29) and (30), we obtain

$$S(x, y) = D(x, y) + D(y, x) \tag{31}$$

$$= [D(z_\lambda, y) + D(y, z_\lambda)] + [D(z_\lambda, x) + D(x, z_\lambda)] + \frac{1-\lambda}{\lambda} S(x, z_\lambda) + \frac{\lambda}{1-\lambda} S(z_\lambda, y) \tag{32}$$

$$= \left(1 + \frac{\lambda}{1-\lambda}\right) S(z_\lambda, y) + \left(1 + \frac{1-\lambda}{\lambda}\right) S(x, z_\lambda) \tag{33}$$

$$= \frac{1}{1-\lambda} S(y, z_\lambda) + \frac{1}{\lambda} S(z_\lambda, x), \tag{34}$$

which is (ii). ■

4. Bregman nearest points and maximal monotone operators

We shall need the following pointwise version of a concept due to Rockafellar and Wets [24, Definition 1.16].

Definition 4.1. Let $g : \mathbb{R}^J \times \mathbb{R}^J \to]-\infty, +\infty]$ and let $\bar{y} \in \mathbb{R}^J$. We say that g is level bounded in the first variable locally uniformly at $\bar{y}$, if for every $\alpha \in \mathbb{R}$, there exists $\delta > 0$ such that

$$\bigcup_{y \in B_\delta(\bar{y})} \{x \in \mathbb{R}^J : g(x, y) \le \alpha\} \quad \text{is bounded.}$$

Proposition 4.2. *The Bregman distance D is level bounded in the first variable locally uniformly at every point in U.*

Proof. Suppose the opposite. Then there exist $\bar{y} \in U$ and $\bar{\alpha} \in \mathbb{R}$ such that for every $n \in \{1, 2, \ldots\}$, there exist $y_n \in U$ and $x_n \in \operatorname{dom} f$ such that

$$\|y_n - \bar{y}\| < \frac{1}{n}, \qquad D(x_n, y_n) \le \bar{\alpha}, \quad \text{and} \quad \|x_n\| > n.$$

We then have $y_n \to \bar{y}$, $\|x_n\| \to \infty$ and

$$D(x_n, y_n) \le \bar{\alpha}. \tag{35}$$

Now

$$D(x_n, y_n) = f(x_n) - f(y_n) - \langle \nabla f(y_n), x_n - y_n \rangle \tag{36}$$

$$= f(x_n) - \langle \nabla f(y_n), x_n \rangle + [-f(y_n) + \langle \nabla f(y_n), y_n \rangle]. \tag{37}$$

Since f is Legendre, ∇f is continuous on U. When $n \to \infty$, we have

$$-f(y_n) + \langle \nabla f(y_n), y_n \rangle \to -f(\bar{y}) + \langle \nabla f(\bar{y}), \bar{y} \rangle, \tag{38}$$

and

$$f(x_n) - \langle \nabla f(y_n), x_n \rangle = \|x_n\| \left(\frac{f(x_n)}{\|x_n\|} - \left\langle \nabla f(y_n), \frac{x_n}{\|x_n\|} \right\rangle \right) \tag{39}$$

$$\ge \|x_n\| \left(\frac{f(x_n)}{\|x_n\|} - \|\nabla f(y_n)\| \right) \to \infty, \tag{40}$$

since $\|\nabla f(y_n)\| \to \|\nabla f(\bar{y})\|$ and $\lim f(x_n)/\|x_n\| = +\infty$. (38)-(40) and (37) altogether show that $D(x_n, y_n) \to \infty$, but this contradicts (35). ■

The following result will be very useful later.

Theorem 4.3. *The following hold.*

(i) *For each $y \in U$, the set $\overleftarrow{P}_C(y)$ is nonempty and compact. Moreover, $\overleftarrow{D}_C$ is continuous on U.*
(ii) *If $x_n \in \overleftarrow{P}_C(y_n)$ and $y_n \to y \in U$, then the sequence $(x_n)_{n=1}^\infty$ is bounded, and all its cluster points lie in $\overleftarrow{P}_C(y)$.*
(iii) *Let $y \in U$ and $\overleftarrow{P}_C(y) = \{x\}$. If $x_n \in \overleftarrow{P}_C(y_n)$ and $y_n \to y$, then $x_n \to x$; consequently, $\overleftarrow{P}_C$ is continuous at y.*

Proof. Fix $\bar{y} \in U$ and $\delta > 0$ such that $B_\delta(\bar{y}) \subset U$. Consider the proper lower semicontinuous function $g : \mathbb{R}^J \times \mathbb{R}^J \to]-\infty, +\infty]$ defined by

$$(x, y) \mapsto D(x, y) + \iota_C(x) + \iota_{B_\delta(\bar{y})}(y).$$

Observe that $\operatorname{dom} g = C \times B_\delta(\bar{y})$. For every $y \in \mathbb{R}^J$ and $\alpha \in \mathbb{R}$, we have

$$\{x \in \mathbb{R}^J : g(x, y) \le \alpha\} = \begin{cases} C \cap \{x \in \mathbb{R}^J : D(x, y) \le \alpha\}, & \text{if } y \in B_\delta(\bar{y}); \\ \varnothing, & \text{otherwise.} \end{cases} \tag{41}$$

We now show that

$$g \text{ is level bounded in the first variable locally uniformly at every point in } \mathbb{R}^J. \tag{42}$$

To this end, fix $\bar{z} \in \mathbb{R}^J$ and $\alpha \in \mathbb{R}$.

Case 1: $\bar{z} \notin B_\delta(\bar{y})$.

Let $\epsilon > 0$ be so small that $B_\delta(\bar{y}) \cap B_\epsilon(\bar{z}) = \varnothing$. Then (41) yields

$$\bigcup_{z \in B_\epsilon(\bar{z})} \{x \in \mathbb{R}^J : g(x, z) \le \alpha\} = \varnothing,$$

which is certainly bounded.

Case 2: $\bar{z} \in B_\delta(\bar{y})$.

Since $B_\delta(\bar{y}) \subset U$, we have $\bar{z} \in U$. Proposition 4.2 guarantees the existence of $\epsilon > 0$ such that

$$\bigcup_{z \in B_\epsilon(\bar{z})} \{x \in \mathbb{R}^J : D(x, z) \le \alpha\} \quad \text{is bounded.}$$

In view of (41), the set

$$\bigcup_{z \in B_\epsilon(\bar{z}) \cap B_\delta(\bar{y})} C \cap \{x \in \mathbb{R}^J : D(x, z) \le \alpha\} = \bigcup_{z \in B_\epsilon(\bar{z})} \{x \in \mathbb{R}^J : g(x, z) \le \alpha\}$$

is bounded as well.

Altogether, we have verified (42).

Define a function m at $y \in \mathbb{R}^J$ by

$$m(y) := \inf_{x \in \mathbb{R}^J} g(x, y) = \begin{cases} \inf_{x \in C} D(x, y) = \overleftarrow{D}_C(y), & \text{if } y \in B_\delta(\bar{y}); \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $m = \overleftarrow{D}_C + \iota_{B_\delta(\bar{y})}$ and

$$\operatorname*{argmin}_{x \in \mathbb{R}^J} g(x, y) = \begin{cases} \operatorname*{argmin}_{x \in C} D(x, y) = \overleftarrow{P}_C(y), & \text{if } y \in B_\delta(\bar{y}); \\ \varnothing, & \text{otherwise.} \end{cases}$$

Now (42) and [24, Theorem 1.17(a)] imply that if $y \in B_\delta(\bar{y})$, then $\overleftarrow{P}_C(y)$ is nonempty and compact. In particular, $\overleftarrow{P}_C(\bar{y}) \neq \varnothing$ and compact. Take $\bar{x} \in \overleftarrow{P}_C(\bar{y})$. As

$$g(\bar{x}, \cdot) = D(\bar{x}, \cdot) + \iota_{B_\delta(\bar{y})}$$

is continuous at $\bar{y}$, by [24, Theorem 1.17(c)], the function m is continuous at $\bar{y}$. Hence $\overleftarrow{D}_C$ is continuous at $\bar{y}$. Since $\bar{y} \in U$ is arbitrary, this proves (i). Next, [24, Theorem 1.17(b)] gives (ii) since $\overleftarrow{D}_C$ is continuous on U.

Finally, (iii) is an immediate consequence of (ii). ■

Our next result states that $\overleftarrow{P}_C \circ \nabla f^*$ is a monotone operator. This is also related to [3, Proposition 3.32.(ii)(c)], which establishes a stronger property when C is convex.

Proposition 4.4. *For every x, y in U,*

$$\langle \overleftarrow{P}_C(y) - \overleftarrow{P}_C(x), \nabla f(y) - \nabla f(x) \rangle \geq 0; \tag{43}$$

consequently, $\overleftarrow{P}_C \circ \nabla f^*$ is monotone.

Proof. Since

$$D(\overleftarrow{P}_C(x), y) \geq D(\overleftarrow{P}_C(y), y), \qquad D(\overleftarrow{P}_C(y), x) \geq D(\overleftarrow{P}_C(x), x),$$

we use (4) to get

$$f(\overleftarrow{P}_C(x)) - f(\overleftarrow{P}_C(y)) - \langle \nabla f(y), \overleftarrow{P}_C(x) - \overleftarrow{P}_C(y) \rangle \geq 0,$$
$$f(\overleftarrow{P}_C(y)) - f(\overleftarrow{P}_C(x)) - \langle \nabla f(x), \overleftarrow{P}_C(y) - \overleftarrow{P}_C(x) \rangle \geq 0.$$

Adding these inequalities yields

$$\langle \nabla f(y), \overleftarrow{P}_C(y) - \overleftarrow{P}_C(x) \rangle - \langle \nabla f(x), \overleftarrow{P}_C(y) - \overleftarrow{P}_C(x) \rangle \geq 0,$$

i.e., (43). The monotonicity now follows from Fact 2.2 and our assumption that $\operatorname{dom} f^* = \mathbb{R}^J$. ■

Definition 4.5. The set C is *Chebyshev with respect to the left Bregman distance*, or simply $\overleftarrow{D}$-Chebyshev, if for every $x \in U$, $\overleftarrow{P}_C(x)$ is a singleton.

For some instances of f, it is known that if C is convex, then it is $\overleftarrow{D}$-Chebyshev (see, e.g., [2, Theorem 3.14]) and $\overleftarrow{P}_C$ is continuous (see, e.g., [4, Proposition 3.10(i)]). The next result is a refinement.

Proposition 4.6. *Suppose that C is $\overleftarrow{D}$-Chebyshev. Then $\overleftarrow{P}_C : U \to C$ is continuous. Hence $\overleftarrow{P}_C \circ \nabla f^*$ is continuous and maximal monotone.*

Proof. While the continuity of $\overleftarrow{P}_C$ follows from Theorem 4.3(iii), Proposition 4.4 shows that $\overleftarrow{P}_C \circ \nabla f^*$ is monotone. Since $\overleftarrow{P}_C$ is continuous on U and $\nabla f^* : \mathbb{R}^J \to U$ is continuous, we conclude that $\overleftarrow{P}_C \circ \nabla f^*$ is continuous on $\mathbb{R}^J$. Altogether, since $\overleftarrow{P}_C \circ \nabla f^*$ is single-valued, it is maximal monotone on $\mathbb{R}^J$ by [24, Example 12.7]. ■

Rockafellar's well-known result on the virtual convexity of the range of a maximal monotone operator allows us to show that $\overleftarrow{D}$-Chebyshev sets are convex. Our proof extends a Hilbert space technique due to Berens and Westphal [5].

Theorem 4.7 ($\overleftarrow{D}$-*Chebyshev Sets are Convex*). *Suppose that C is $\overleftarrow{D}$-Chebyshev. Then C is convex.*

Proof. By Proposition 4.6, $\overleftarrow{P}_C \circ \nabla f^*$ is a maximal monotone operator on $\mathbb{R}^J$. Using [24, Theorem 12.41] (or [25, Theorem 19.2]), $\mathrm{cl}[\mathrm{ran}\, \overleftarrow{P}_C \circ \nabla f^*]$ is convex. Since $\mathrm{ran}\, \nabla f^* = U$ and $C \subset U$, it follows that

$$C \supset \mathrm{ran}\left(\overleftarrow{P}_C \circ \nabla f^*\right) = \overleftarrow{P}_C(\nabla f^*(\mathbb{R}^J)) = \overleftarrow{P}_C(U) \supset \overleftarrow{P}_C(C) = C,$$

from which $\mathrm{cl}[\mathrm{ran}\, \overleftarrow{P}_C \circ \nabla f^*] = \mathrm{cl}\, C = C$. Hence C is convex. ■

Corollary 4.8. *The set C is $\overleftarrow{D}$-Chebyshev if and only if it is convex.*

5. Subdifferentiabilities of Bregman distances

Let us show that $\overleftarrow{D}_C$ is locally Lipschitz on U.

Proposition 5.1. *Suppose f is twice continuously differentiable on U. Then the left Bregman distance function satisfies*

$$\overleftarrow{D}_C = f^* \circ \nabla f - (f + \iota_C)^* \circ \nabla f = [f^* - (f + \iota_C)^*] \circ \nabla f, \tag{44}$$

and it is locally Lipschitz on U.

Proof. The Mean Value Theorem and the continuity of $\nabla^2 f$ on U imply that ∇f is locally Lipschitz on U. For $y \in U$,

$$\overleftarrow{D}_C(y) = \inf_{c \in C} [f(c) - f(y) - \langle \nabla f(y), c - y \rangle] \tag{45}$$

$$= \inf_c [(f + \iota_C)(c) - \langle \nabla f(y), c \rangle + f^*(\nabla f(y))] \tag{46}$$

$$= f^*(\nabla f(y)) - \sup_c [\langle \nabla f(y), c \rangle - (f + \iota_C)(c)] \tag{47}$$

$$= f^*(\nabla f(y)) - (f + \iota_C)^*(\nabla f(y)). \tag{48}$$

Note that $f + \iota_C \geq f$, $(f + \iota_C)^* \leq f^*$, so $\mathrm{dom}\, f^* \subset \mathrm{dom}(f + \iota_C)^*$. Being convex functions, both $(f + \iota_C)^*$ and f^* are locally Lipschitz on the interior of their respective domains, in particular on $\mathrm{int}\,\mathrm{dom}\, f^* = \mathbb{R}^J$. Since $\nabla f : U \to \mathbb{R}^J$ is locally Lipschitz, we conclude that $\overleftarrow{D}_C$ is locally Lipschitz on U. ■

For a function g that is finite and locally Lipschitz at a point y, we define the *Dini subderivative* and *Clarke subderivative* of g at y in the direction w, denoted respectively by $\mathrm{d}\, g(y)(w)$ and $\hat{\mathrm{d}} g(y)(w)$, via

$$\mathrm{d}\, g(y)(w) := \liminf_{t \downarrow 0} \frac{g(y + tw) - g(y)}{t},$$

$$\hat{\mathrm{d}} g(y)(w) := \limsup_{\substack{x \to y \\ t \downarrow 0}} \frac{g(x + tw) - g(x)}{t},$$

and the corresponding *Dini subdifferential* and *Clarke subdifferential* via

$$\hat\partial g(y) := \{y^* \in \mathbb{R}^J : \langle y^*, w\rangle \le \mathrm{d}\, g(y)(w),\ \forall w \in \mathbb{R}^J\},$$
$$\overline{\partial} g(y) := \{y^* \in \mathbb{R}^J : \langle y^*, w\rangle \le \hat{\mathrm{d}} g(y)(w),\ \forall w \in \mathbb{R}^J\}.$$

Furthermore, the *limiting subdifferential* is defined by

$$\partial_L g(y) := \limsup_{x\to y} \hat\partial g(x),$$

see [24, Definition 8.3]. We say that g is *Clarke regular* at y if $\mathrm{d}\, g(y)(w) = \hat{\mathrm{d}} g(y)(w)$ for every $w \in \mathbb{R}^J$, or equivalently $\hat\partial g(y) = \overline{\partial} g(y)$. For further properties of these subdifferentials and subderivatives, see [15,21,24].

We now study the subdifferentiability of $\overleftarrow{D}_C$ in terms of $\overleftarrow{P}_C$.

Proposition 5.2. *Suppose f is twice continuously differentiable on U. Then the function $-\overleftarrow{D}_C$ is Dini subdifferentiable on U; more precisely, if $y \in U$, then*

$$\nabla^2 f(y)[\overleftarrow{P}_C(y) - y] \subset \hat\partial(-\overleftarrow{D}_C)(y),$$

and thus

$$\nabla^2 f(y)[\operatorname{conv} \overleftarrow{P}_C(y) - y] \subset \hat\partial(-\overleftarrow{D}_C)(y). \tag{49}$$

Proof. Fix $y \in U$. By Theorem 4.3(i), $\overleftarrow{P}_C(y) \neq \varnothing$. Let $x \in \overleftarrow{P}_C(y)$. As $\hat\partial$ is convex-valued, it suffices to show that

$$\nabla^2 f(y)(x - y) \in \hat\partial(-\overleftarrow{D}_C)(y). \tag{50}$$

To this end, let $t > 0$ and $w \in \mathbb{R}^J$. Since for sufficiently small t, $y + tw \in U$,

$$-\overleftarrow{D}_C(y + tw) = \sup_{c\in C} (-f(c) + f(y + tw) + \langle\nabla f(y + tw), c - (y + tw)\rangle) \tag{51}$$
$$\ge -f(x) + f(y + tw) + \langle\nabla f(y + tw), x - (y + tw)\rangle \tag{52}$$

and

$$\overleftarrow{D}_C(y) = f(x) - f(y) - \langle\nabla f(y), x - y\rangle, \tag{53}$$

we have

$$-\overleftarrow{D}_C(y + tw) + \overleftarrow{D}_C(y) \ge f(y + tw) - f(y) + \langle\nabla f(y + tw) - \nabla f(y), x - y\rangle + \langle\nabla f(y + tw), -tw\rangle.$$

Dividing both sides by t and taking the limit inferior as $t \downarrow 0$, we have

$$\mathrm{d}(-\overleftarrow{D}_C)(y)(w) \ge \langle\nabla f(y), w\rangle + \langle\nabla^2 f(y)w, x - y\rangle - \langle\nabla f(y), w\rangle \tag{54}$$
$$= \langle\nabla^2 f(y)(x - y), w\rangle, \tag{55}$$

which gives (50). ■

Lemma 5.3. *Suppose that f is twice continuously differentiable on U, let $y \in U$, and suppose that $\overleftarrow{P}_C(y)$ is a singleton. Then $\overleftarrow{D}_C$ is Dini subdifferentiable at y and*

$$\nabla^2 f(y)(y - \overleftarrow{P}_C(y)) \in \hat\partial\, \overleftarrow{D}_C(y). \tag{56}$$

Proof. Suppose that $\overleftarrow{P}_C(y) = \{x\}$, and fix $w \in \mathbb{R}^J$. Let (t_n) be a positive sequence such that $(y + t_n w)$ lies in U, $t_n \downarrow 0$, and

$$\mathrm{d}\,\overleftarrow{D}_C(y)(w) = \lim_{n\to\infty} \frac{\overleftarrow{D}_C(y + t_n w) - \overleftarrow{D}_C(y)}{t_n}.$$

Select $x_n \in \overleftarrow{P}_C(y + t_n w)$, which is possible by Theorem 4.3(i). We have

$$\begin{aligned}
\overleftarrow{D}_C(y + t_n w) - \overleftarrow{D}_C(y) &= D(x_n, y + t_n w) - D(x_n, y) + D(x_n, y) - D(x, y)\\
&\geq D(x_n, y + t_n w) - D(x_n, y)\\
&= f(x_n) - f(y + t_n w) - \langle \nabla f(y + t_n w), x_n - (y + t_n w)\rangle\\
&\quad - [f(x_n) - f(y) - \langle \nabla f(y), x_n - y\rangle]\\
&= -(f(y + t_n w) - f(y)) - \langle \nabla f(y + t_n w)\\
&\quad - \nabla f(y), x_n - y\rangle + t_n\langle \nabla f(y + t_n w), w\rangle.
\end{aligned}$$

Dividing by t_n, we get

$$\begin{aligned}
&\frac{\overleftarrow{D}_C(y + t_n w) - \overleftarrow{D}_C(y)}{t_n}\\
&\quad \geq -\frac{f(y + t_n w) - f(y)}{t_n} - \frac{\langle \nabla f(y + t_n w) - \nabla f(y), x_n - y\rangle}{t_n} + \langle \nabla f(y + t_n w), w\rangle.
\end{aligned} \tag{57}$$

By Theorem 4.3(iii), $x_n \to x$. Taking limits in (57) yields

$$\mathrm{d}\,\overleftarrow{D}_C(y)(w) \geq -\langle \nabla^2 f(y) w, x - y\rangle = \langle \nabla^2 f(y)(y - x), w\rangle.$$

Since this holds for every $w \in \mathbb{R}^J$, we conclude that $\nabla^2 f(y)(y - x) \in \hat{\partial}\,\overleftarrow{D}_C(y)$. ■

Lemma 5.3 allows us to generalize [24, Example 8.53] from the Euclidean distance to the left Bregman distance. It delineates the differences between the Dini subdifferential, limiting subdifferential and Clarke subdifferential.

Theorem 5.4. *Suppose that f is twice continuously differentiable on U and that for every $u \in U$, $\nabla^2 f(u)$ is positive definite. Set $g = \overleftarrow{D}_C$, and let $y \in U$ and $w \in \mathbb{R}^J$. Then the following hold.*

(i) *The Dini subderivative is*

$$\mathrm{d}\,g(y)(w) = \min_{x \in \overleftarrow{P}_C(y)} \langle \nabla^2 f(y)(y - x), w\rangle, \tag{58}$$

so that the Dini subdifferential of g is

$$\hat{\partial} g(y) = \begin{cases} \{\nabla^2 f(y)[y - \overleftarrow{P}_C(y)]\} & \text{if } \overleftarrow{P}_C(y) \text{ is a singleton;}\\ \varnothing, & \text{otherwise.}\end{cases} \tag{59}$$

The limiting subdifferential is

$$\partial_L g(y) = \nabla^2 f(y)[y - \overleftarrow{P}_C(y)]. \tag{60}$$

The Clarke subderivative is

$$\hat{\mathrm{d}} g(y)(w) = \max_{x \in \overleftarrow{P}_C(y)} \langle \nabla^2 f(y)(y - x), w\rangle, \tag{61}$$

from which we get the Clarke subdifferential

$$\overline{\partial} g(y) = \nabla^2 f(y)[y - \operatorname{conv} \overleftarrow{P}_C(y)]. \tag{62}$$

Hence $-\overleftarrow{D}_C$ *is Clarke regular on* U.

(ii) *If* $y \in C$, *then* g *is strictly differentiable with derivative 0.*

Proof. By Theorem 4.3(i), $\overleftarrow{P}_C(y) \neq \varnothing$. Fix $x \in \overleftarrow{P}_C(y)$ and $t > 0$ sufficiently small so that $y + tw \in U$. In view of $\overleftarrow{D}_C(y + tw) \leq D(x, y + tw)$ and $\overleftarrow{D}_C(y) = D(x, y)$, we have

$$\begin{aligned}
\mathrm{d}\, g(y)(w) &= \liminf_{t\downarrow 0} \frac{\overleftarrow{D}_C(y + tw) - \overleftarrow{D}_C(y)}{t} \leq \liminf_{t\downarrow 0} \frac{D(x, y + tw) - D(x, y)}{t} \\
&= \liminf_{t\downarrow 0} \frac{f(x) - f(y + tw) - \langle \nabla f(y + tw), x - (y + tw)\rangle - [f(x) - f(y) - \langle \nabla f(y), x - y\rangle]}{t} \\
&= \liminf_{t\downarrow 0} \frac{-[f(y + tw) - f(y)] - \langle \nabla f(y + tw) - \nabla f(y), x - y\rangle + t\langle \nabla f(y + tw), w\rangle}{t} \\
&= \liminf_{t\downarrow 0} -\frac{f(y + tw) - f(y)}{t} - \frac{\langle \nabla f(y + tw) - \nabla f(y), x - y\rangle}{t} + \langle \nabla f(y + tw), w\rangle \\
&= -\langle \nabla f(y), w\rangle - \langle \nabla^2 f(y)w, x - y\rangle + \langle \nabla f(y), w\rangle \\
&= \langle \nabla^2 f(y)(y - x), w\rangle.
\end{aligned}$$

Since this holds for every $x \in \overleftarrow{P}_C(y)$, it follows from Theorem 4.3(i) that

$$\mathrm{d}\, g(y)(w) \leq \min_{x \in \overleftarrow{P}_C(y)} \langle \nabla^2 f(y)(y - x), w\rangle.$$

To get the opposite inequality, we consider a positive sequence (t_n) such that $t_n \downarrow 0$, $(y + t_n w)$ lies in U, and

$$\mathrm{d}\, g(y)(w) = \lim_{n\to\infty} \frac{\overleftarrow{D}_C(y + t_n w) - \overleftarrow{D}_C(y)}{t_n}.$$

Select $x_n \in \overleftarrow{P}_C(y + t_n w)$, which is possible by Theorem 4.3(i). Then

$$\begin{aligned}
\overleftarrow{D}_C(y + t_n w) &= D(x_n, y + t_n w) \\
&= f(x_n) - f(y + t_n w) - \langle \nabla f(y + t_n w), x_n - (y + t_n w)\rangle
\end{aligned} \tag{63}$$

and

$$\overleftarrow{D}_C(y) \leq D(x_n, y) \leq f(x_n) - f(y) - \langle \nabla f(y), x_n - y\rangle. \tag{64}$$

By Theorem 4.3(ii), and after taking a subsequence if necessary, we assume that $x_n \to x \in \overleftarrow{P}_C(y)$. We estimate

$$\begin{aligned}
&\frac{\overleftarrow{D}_C(y + t_n w) - \overleftarrow{D}_C(y)}{t_n} \\
&\quad \geq \frac{-[f(y + t_n w) - f(y)] - \langle \nabla f(y + t_n w) - \nabla f(y), x_n - y\rangle + \langle \nabla f(y + t_n w), t_n w\rangle}{t_n} \\
&\quad = \frac{-[f(y + t_n w) - f(y)]}{t_n} - \frac{\langle \nabla f(y + t_n w) - \nabla f(y), x_n - y\rangle}{t_n} + \langle \nabla f(y + t_n w), w\rangle.
\end{aligned} \tag{65}$$

Taking limits, we obtain

$$\mathrm{d}\, g(y)(w) \geq -\langle \nabla^2 f(y)w, x-y\rangle = \langle \nabla^2 f(y)(y-x), w\rangle \geq \min_{x\in \overleftarrow{P}_C(y)} \langle \nabla^2 f(y)(y-x), w\rangle.$$

Therefore, (58) is correct.

For $y^* \in \mathbb{R}^J$, $y^* \in \hat{\partial} g(y)$ if and only if

$$\langle y^*, w\rangle \leq \langle \nabla^2 f(y)(y-x), w\rangle \quad \forall x \in \overleftarrow{P}_C(y), w \in \mathbb{R}^J.$$

This holds if and only if $y^* = \nabla^2 f(y)(y-x)$, $\forall x \in \overleftarrow{P}_C(y)$; since $\nabla^2 f(y)$ is invertible, we deduce that $x = y - (\nabla^2 f(y))^{-1} y^*$, so that $\overleftarrow{P}_C(y)$ is unique. Therefore, if $\overleftarrow{P}_C(y)$ is not unique, then $\hat{\partial} g(y)$ has to be empty. Hence (59) holds.

For every $z \in \mathbb{R}^J$, we have

$$\hat{\partial} g(z) \subset \nabla^2 f(z)(z - \overleftarrow{P}_C(z)).$$

The upper semicontinuity of $\overleftarrow{P}_C$ (see Theorem 4.3(ii)) implies through $\partial_L g(y) = \limsup_{z\to y} \hat{\partial} g(z)$ that

$$\partial_L g(y) \subset \nabla^2 f(y)(y - \overleftarrow{P}_C(y)). \tag{66}$$

Equality actually has to hold. Indeed, for $x \in \overleftarrow{P}_C(y)$ and $0 \leq \lambda < 1$, the point

$$z_\lambda := \nabla f^*(\lambda \nabla f(y) + (1-\lambda)\nabla f(x)),$$

has $\overleftarrow{P}_C(z_\lambda) = \{x\}$ by Proposition 3.2(ii). Lemma 5.3 shows that

$$\nabla^2 f(z_\lambda)(z_\lambda - x) \in \hat{\partial} g(z_\lambda),$$

where $\nabla^2 f(z_\lambda)(z_\lambda - x) \to \nabla^2 f(y)(y-x)$ as $\lambda \to 1$, since $\nabla^2 f$ is continuous. Thus $\nabla^2 f(y)(y-x) \in \partial_L g(y)$ and therefore

$$\nabla^2 f(y)(y - \overleftarrow{P}_C(y)) \subset \partial_L g(y). \tag{67}$$

Hence (66) and (67) together give (60).

Since g is locally Lipschitz around $y \in U$ by Proposition 5.1, the singular subdifferential of g at y is 0, so that its polar cone is $\mathbb{R}^J$. Then for every $w \in \mathbb{R}^J$, using [24, Exercise 8.23] we have

$$\hat{\mathrm{d}} g(y)(w) = \sup\{\langle y^*, w\rangle : \ y^* \in \partial_L g(y)\};$$

thus, (61) follows from (60). Now (61) is the same as

$$\hat{\mathrm{d}} g(y)(w) = \max \langle \nabla^2 f(y)(y - \operatorname{conv} \overleftarrow{P}_C(y)), w\rangle.$$

As $\operatorname{conv} \overleftarrow{P}_C(y)$ is compact, we obtain (62) or directly apply [24, Theorem 8.49] and (60).

The Clarke regularity of $-\overleftarrow{D}_C$ follows from combining (49) and (62). Indeed,

$$\nabla^2 f(y)[\operatorname{conv} \overleftarrow{P}_C(y) - y] \subset \hat{\partial}(-\overleftarrow{D}_C)(y) \subset \bar{\partial}(-\overleftarrow{D}_C)(y) = \nabla^2 f(y)[\operatorname{conv} \overleftarrow{P}_C(y) - y],$$

so that $\hat{\partial}(-\overleftarrow{D}_C)(y) = \bar{\partial}(-\overleftarrow{D}_C)(y)$.

(ii): When $y \in C$, $\overleftarrow{P}_C(y) = \{y\}$. By (60), $\partial_L g(y) = \{0\}$, and this implies that g is strictly differentiable at y by [24, Theorem 9.18(b)]. ■

Corollary 5.5. *Suppose that f is twice continuously differentiable and that $\nabla^2 f(y)$ is positive definite, for every $y \in U$. Then for $y \in U$, the following are equivalent:*

(i) $\overleftarrow{D}_C$ *is Dini subdifferentiable at* y;
(ii) $\overleftarrow{D}_C$ *is differentiable at* y;
(iii) $\overleftarrow{D}_C$ *is strictly differentiable at* y;
(iv) $\overleftarrow{D}_C$ *is Clarke regular at* y;
(v) $\overleftarrow{P}_C(y)$ *is a singleton.*

If these hold, we have $\nabla \overleftarrow{D}_C(y) = \nabla^2 f(y)[y - \overleftarrow{P}_C(y)]$.

Proof. (i) ⇒ (ii): By Proposition 5.2, both $-\overleftarrow{D}_C$ and $\overleftarrow{D}_C$ are Dini subdifferentiable. Thus $\overleftarrow{D}_C$ is differentiable at y (see [13, Exercise 3.4.14 on page 143]), and

$$\hat{\partial} \overleftarrow{D}_C(y) = -\hat{\partial}(-\overleftarrow{D}_C)(y) = \{\nabla \overleftarrow{D}_C(y)\}.$$

(ii) ⇒ (i) is clear. (ii) ⇔ (iii) ⇔ (iv): This is a consequence of [30, Theorem 3.4]. (ii) ⇔ (v): If $\overleftarrow{D}_C$ is differentiable at y, then (59) implies that $\overleftarrow{P}_C(y)$ is a singleton. Conversely, if $\overleftarrow{P}_C(y)$ is a singleton, then (62), Proposition 5.1, and [12, Proposition 2.2.4] show that $\overleftarrow{D}_C$ is strictly differentiable and hence differentiable at y. Finally, the gradient formula $\nabla \overleftarrow{D}_C(y) = \nabla^2 f(y)[y - \overleftarrow{P}_C(y)]$ is a consequence of Proposition 5.2 or Lemma 5.3. ■

Corollary 5.6. *Suppose that* f *is twice continuously differentiable on* U *and that* $\nabla^2 f(y)$ *is positive definite for every* $y \in U$. *Then* $\overleftarrow{P}_C$ *is almost everywhere and generically single-valued on* U.

Proof. By Proposition 5.1, $\overleftarrow{D}_C$ is locally Lipschitz on U. Apply Rademacher's Theorem [7, Theorem 9.1.2] or [13, Corollary 3.4.19] to obtain that $\overleftarrow{D}_C$ is differentiable almost everywhere on U. Moreover, since $-\overleftarrow{D}_C$ is Clarke regular on U by Theorem 5.4, we use [20, Theorem 10] to conclude that $-\overleftarrow{D}_C$ is differentiable generically on U, and so is $\overleftarrow{D}_C$. Hence the result follows from Corollary 5.5. ■

6. Characterizations of Chebyshev sets

Definition 6.1. For $g : \mathbb{R}^J \to]-\infty, +\infty]$ (not necessarily convex), we let

$$\partial g(x) := \{s \in \mathbb{R}^J : \ g(y) \geq g(x) + \langle s, y - x\rangle \ \forall y \in \mathbb{R}^J\} \quad \text{if } x \in \operatorname{dom} g;$$

and $\partial g(x) = \varnothing$ otherwise; and the Fenchel conjugate of g is defined by

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

According to [17, Proposition 1.4.3],

$$s \in \partial g(x) \quad \Rightarrow \quad x \in \partial g^*(s), \tag{68}$$

which becomes "⇔" if $g \in \Gamma$. In order to study $\overleftarrow{D}$-Chebyshev sets, we need two preparatory results concerning subdifferentiabilities of $f + \iota_C$ and $(f + \iota_C)^*$. Lemmas 6.2, 6.3 and Theorem 6.6 generalize respectively, and are inspired by [17, Propositions 3.2.1, 3.2.2 and Theorem 3.2.3].

Lemma 6.2. *Let* $x \in \mathbb{R}^J$. *Then*

$$\partial(f + \iota_C)(x) = \{s \in \mathbb{R}^J : x \in \overleftarrow{P}_C(\nabla f^*(s))\} = (\overleftarrow{P}_C \circ \nabla f^*)^{-1}(x),$$

and consequently $\partial(f + \iota_C) = \left(\overleftarrow{P}_C \circ \nabla f^*\right)^{-1}$.

Proof. The statement is clear if $x \notin C$, so assume $x \in C$. By [17, Theorem 1.4.1],

$$s \in \partial(f + \iota_C)(x) \quad \Leftrightarrow \quad (f + \iota_C)^*(s) + (f + \iota_C)(x) = \langle s, x \rangle. \tag{69}$$

Proposition 5.1 shows that

$$(f + \iota_C)^* = f^* - \overleftarrow{D}_C \circ \nabla f^* \quad \text{on } \mathbb{R}^J.$$

Combining with (69) and since $x \in C$, we get

$$s \in \partial(f + \iota_C)(x) \quad \Leftrightarrow \quad f^*(s) - (\overleftarrow{D}_C \circ \nabla f^*)(s) + f(x) = \langle s, x \rangle;$$

equivalently,

$$\begin{aligned} \overleftarrow{D}_C(\nabla f^*(s)) &= f(x) + f^*(s) - \langle s, x \rangle & (70)\\ &= f(x) + f^*((\nabla f \circ \nabla f^*)(s)) - \langle \nabla f \circ \nabla f^*(s), x \rangle & (71)\\ &= f(x) - f(\nabla f^*(s)) - \langle \nabla f(\nabla f^*(s)), x - \nabla f^*(s) \rangle & (72)\\ &= D(x, \nabla f^*(s)), & (73) \end{aligned}$$

i.e., $x \in \overleftarrow{P}_C(\nabla f^*(s))$. ■

The following result, which establishes the link between $\partial(f + \iota_C)^*$ and $\overleftarrow{P}_C \circ \nabla f^*$, is the cornerstone for the convexity characterization of $\overleftarrow{D}$-Chebyshev sets.

Lemma 6.3. *Let $s \in \mathbb{R}^J$. Then*

$$\partial(f + \iota_C)^*(s) = \operatorname{conv}[\overleftarrow{P}_C(\nabla f^*(s))] = \operatorname{conv}[\overleftarrow{P}_C \circ \nabla f^*(s)]$$

Consequently, $\overleftarrow{P}_C \circ \nabla f^*$ is monotone on $\mathbb{R}^J$.

Proof. Since f is 1-coercive and C is closed, the function $f + \iota_C$ is 1-coercive and lower semicontinuous. We have that $\operatorname{conv}(f + \iota_C)$ is lower semicontinuous by [17, Proposition 1.5.4], and $\operatorname{dom}(f + \iota_C)^* = \mathbb{R}^J$ by [17, Proposition 1.3.8]. Now

$$x \in \partial(f + \iota_C)^*(s) \Leftrightarrow x \in \partial[\operatorname{conv}(f + \iota_C)]^*(s) \Leftrightarrow s \in \partial[\operatorname{conv}(f + \iota_C)](x),$$

in which the first equivalences follows from [17, Corollary 1.3.6] and the second equivalence uses the lower semicontinuity of $\operatorname{conv}(f + \iota_C)$. Using [17, Theorem 1.5.6], $s \in \partial[\operatorname{conv}(f + \iota_C)](x)$ if and only if there exist $x_1, \ldots, x_k \in \mathbb{R}^J$, $\alpha_1, \ldots, \alpha_k > 0$ such that

$$\sum_{j=1}^{k} \alpha_j = 1, \qquad x = \sum_{j=1}^{k} \alpha_j x_j, \quad \text{and} \quad s \in \bigcap_{j=1}^{k} \partial(f + \iota_C)(x_j). \tag{74}$$

But $s \in \partial(f + \iota_C)(x_j)$ is equivalent to

$$x_j \in \overleftarrow{P}_C(\nabla f^*(s)),$$

by Lemma 6.2. Hence (74) gives $\partial(f + \iota_C)^*(s) = \operatorname{conv} \overleftarrow{P}_C(\nabla f^*(s))$. Finally, as a selection of $\partial(f + \iota_C)^*$, which is maximal monotone, the operator $\overleftarrow{P}_C \circ \nabla f^*$ is monotone. ■

Remark 6.4. Let $y \in \mathbb{R}^J = \operatorname{dom} f^*$. Then $(f + \iota_C)^*(y) = f^*(y) - \inf_{x \in C}[f(x) + f^*(y) - \langle y, x \rangle]$. Since

$$\begin{aligned} f(x) + f^*(y) - \langle y, x \rangle &= f(x) + f^*(\nabla f(\nabla f^*(y))) - \langle \nabla f(\nabla f^*(y)), x \rangle \\ &= D(x, \nabla f^*(y)), \end{aligned}$$

we have $(f+\iota_C)^*(y) = f^*(y) - \overleftarrow{D}_C(\nabla f^*(y))$. Hence

$$(f+\iota_C)^* = f^* - \overleftarrow{D}_C \circ \nabla f^*;$$

see also Proposition 5.1. If $f = \frac{1}{2}\|\cdot\|^2$, then

$$(f+\iota_C)^* = \frac{1}{2}\|\cdot\|^2 - \frac{1}{2}d_C^2,$$

where $d_C(y) := \inf\{\|y-x\| : x \in C\}, \forall y \in \mathbb{R}^J$. In this case, Lemma 6.3 is classical; see [18, pages 262–264] or [19].

We also need the following result from [28].

Proposition 6.5 (*Soloviov*). *Let $g : \mathbb{R}^J \to]-\infty, +\infty]$ be lower semicontinuous, and assume that g^* is essentially smooth. Then g is convex.*

Now we are ready for the main result of this section.

Theorem 6.6 (*Characterizations of $\overleftarrow{D}$-Chebyshev Sets*). *The following are equivalent:*

(i) *C is convex;*
(ii) *C is $\overleftarrow{D}$-Chebyshev, i.e., $\overleftarrow{P}_C$ is single-valued on U;*
(iii) *$\overleftarrow{P}_C$ is continuous on U;*
(iv) *$\overleftarrow{D}_C \circ \nabla f^*$ is differentiable on $\mathbb{R}^J$;*
(v) *$f + \iota_C$ is convex.*

When these equivalent conditions hold, we have

$$\nabla(\overleftarrow{D}_C \circ \nabla f^*) = \nabla f^* - \overleftarrow{P}_C \circ \nabla f^* \quad \text{on } \mathbb{R}^J; \tag{75}$$

consequently, $\overleftarrow{D}_C \circ \nabla f^*$ is continuously differentiable.

If, in addition, f is twice continuously differentiable on U and $\nabla^2 f(y)$ is positive definite $\forall y \in U$, then (i)–(iv) *are equivalent to*

(vi) *$\overleftarrow{D}_C$ is differentiable on U.*

In this case, we have

$$\nabla \overleftarrow{D}_C(y) = \nabla^2 f(y)[y - \overleftarrow{P}_C(y)] \quad \forall y \in U; \tag{76}$$

consequently, $\overleftarrow{D}_C$ is continuously differentiable.

Proof. (i) $\Rightarrow$ (ii) is well known; see, e.g., [2, Theorem 3.12]. (ii) $\Rightarrow$ (iii) follows from Theorem 4.3(iii). To see (iii) $\Rightarrow$ (iv), we use Remark 6.4:

$$\overleftarrow{D}_C \circ \nabla f^* = f^* - (f+\iota_C)^*.$$

Since $\overleftarrow{P}_C$ is continuous on U and $\nabla f^* : \mathbb{R}^J \to U$, $\partial(f+\iota_C)^*$ is single-valued on $\mathbb{R}^J$ by Lemma 6.3. Thus, $(f+\iota_C)^*$ is differentiable on $\mathbb{R}^J$. Altogether, $\overleftarrow{D}_C \circ \nabla f^* = f^* - (f+\iota_C)^*$ is differentiable on $\mathbb{R}^J$.

When $f + \iota_C$ is convex, since $C \subset U$ we have that $\operatorname{dom}(f+\iota_C) = C$ is convex, and this shows (v) $\Rightarrow$ (i). We now prove (iv) $\Rightarrow$ (v) and assume (iv). Remark 6.4 shows

$$(f+\iota_C)^* = f^* - \overleftarrow{D}_C \circ \nabla f^*, \tag{77}$$

which implies that

$$(f + \iota_C)^* \text{ is differentiable on } \mathbb{R}^J. \tag{78}$$

Since $f + \iota_C$ is lower semicontinuous, it follows from Proposition 6.5 that $f + \iota_C$ is convex.

When equivalent conditions (i)–(v) hold, (75) follows from Lemma 6.3 and (77). Since ∇f^* is continuous and $\overleftarrow{P}_C$ is continuous by (iii), we obtain that $\overleftarrow{D}_C \circ \nabla f^*$ is continuously differentiable. When $\nabla^2 f(y)$ is positive definite, $\forall y \in U$, (ii) $\Leftrightarrow$ (vi) by Corollary 5.5. Finally, (76) follows from Theorem 5.4, i.e., (59). This finishes the proof. ■

Remark 6.7. When C is convex, formula (76) follows also from [4, Proposition 3.12(i)]. See also [10, Lemma 3.2] for the first result of this kind, in the (negative) Boltzmann–Shannon entropy setting.

7. Right Bregman projections

In this section, it will be convenient to write D_f for the Bregman distance associated with f (see (2)). Correspondingly, we write $\overleftarrow{P}_C^f$, $\overrightarrow{P}_C^f$ for the corresponding left and right projection operators. While D_f is convex in its first argument, it is not necessarily so in its second argument. The properties of $\overrightarrow{P}_C^f$ can be studied by using $\overleftarrow{P}_{\nabla f(C)}^{f^*}$.

Proposition 7.1. *Let $f \in \Gamma$ be Legendre and $C \subset \operatorname{int}\operatorname{dom} f$. Then for the right Bregman nearest point projection, we have*

$$\overrightarrow{P}_C^f = \nabla f^* \circ \overleftarrow{P}_{\nabla f(C)}^{f^*} \circ \nabla f; \tag{79}$$

or equivalently,

$$\overleftarrow{P}_{\nabla f(C)}^{f^*} = \nabla f \circ \overrightarrow{P}_C^f \circ \nabla f^*. \tag{80}$$

Proof. By [2, Theorem 3.7(v)] (applied to f^* rather than f),

$$D_{f^*}(x^*, y^*) = D_f(\nabla f^*(y^*), \nabla f^*(x^*)) \quad \forall x^*, y^* \in \operatorname{int}\operatorname{dom} f^*.$$

For every $y^* \in \operatorname{int}\operatorname{dom} f^*$, we thus have

$$\begin{aligned}
\overleftarrow{P}_{\nabla f(C)}^{f^*}(y^*) &= \operatorname*{argmin}_{x^* \in \nabla f(C)} D_{f^*}(x^*, y^*) \\
&= \operatorname*{argmin}_{x^* \in \nabla f(C)} D_f(\nabla f^*(y^*), \nabla f^*(x^*)) && (81) \\
&= \nabla f(\overrightarrow{P}_{\nabla f^*(\nabla f(C))}^f(\nabla f^*(y^*))) = \nabla f(\overrightarrow{P}_C^f(\nabla f^*(y^*))) && (82) \\
&= (\nabla f \circ \overrightarrow{P}_C^f \circ \nabla f^*)(y^*), && (83)
\end{aligned}$$

which gives (80). Finally, we see that (79) is equivalent to (80) by using Fact 2.2. ■

Lemma 7.2. *Let $f \in \Gamma$ be Legendre, let $C \subset \mathbb{R}^J$ be such that $\operatorname{cl} C \subset \operatorname{int}\operatorname{dom} f$, and assume that for every $y \in \operatorname{int}\operatorname{dom} f$, $\overleftarrow{P}_C^f(y) \neq \varnothing$. Then C is closed.*

Proof. Assume that $(c_n)_{n=1}^{\infty}$ is a sequence in C, and $c_n \to y$. We need to show that $y \in C$. By assumption $y \in \operatorname{cl} C$, and $y \in U$. If $y \notin C$, then

$$D_f(c, y) = f(c) - f(y) - \langle \nabla f(y), c - y \rangle > 0, \quad \forall c \in C, \tag{84}$$

by, e.g., [2, Theorem 3.7.(iv)]. On the other hand, as f is continuous on U,

$$0 \le \overleftarrow{D}_C^f(y) \le D_f(c_n, y) = f(c_n) - f(y) - \langle \nabla f(y), c_n - y \rangle \to 0.$$

Thus, $\overleftarrow{D}_C^f(y) = 0$. Using (84), we see that this contradicts the assumption that $\overleftarrow{P}_C^f(y) \neq \varnothing$. ■

Theorem 7.3. *Let $f \in \Gamma$ be Legendre, with full domain $\mathbb{R}^J$, and let $C \subset \mathbb{R}^J$ be closed with $\operatorname{cl}(\nabla f(C)) \subset \operatorname{int} \operatorname{dom} f^*$. Assume that $\overrightarrow{P}_C^f(y)$ is a singleton for every $y \in \mathbb{R}^J$. Then $\nabla f(C)$ is convex.*

Proof. We have the facts that f^* is Legendre and f^* is 1-coercive. By (80), $\overleftarrow{P}_{\nabla f(C)}^{f^*}(y)$ is single-valued for every $y \in \operatorname{int} \operatorname{dom} f^*$. As $\operatorname{cl}(\nabla f(C)) \subset \operatorname{int} \operatorname{dom} f^*$, Lemma 7.2 says that the set $\nabla f(C)$ is closed. Hence we apply Theorem 6.6 to f^* and $\nabla f(C)$, and we obtain that $\nabla f(C)$ is convex. ■

Corollary 7.4. *Let f and C satisfy* A1–A3*, assume that f has full domain, and that $\overrightarrow{P}_C^f(y)$ is a singleton for every $y \in \mathbb{R}^J$. Then $\nabla f(C)$ is convex.*

The following example shows that even if $\overrightarrow{P}_C^f(y)$ is a singleton for every $y \in \operatorname{int} \operatorname{dom} f$, the set C may fail to be convex. Thus, Theorem 6.6 fails for the right Bregman projection $\overrightarrow{P}_C^f$. Note that Theorem 7.3 allows us to conclude that $\nabla f(C)$ is convex rather than C.

Example 7.5. Consider the Legendre function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) := e^x + e^y \quad \forall (x, y) \in \mathbb{R}^2,$$

and its Fenchel conjugate

$$f^*: \mathbb{R}^2 \to]-\infty, +\infty]: (x, y) \mapsto \begin{cases} x \ln x - x + y \ln y - y, & \text{if } x \ge 0, y \ge 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Define a compact convex set

$$C := [(0, 0), (1, 2)] = \{(\lambda, 2\lambda) : 0 \le \lambda \le 1\}.$$

As $\nabla f(x, y) = (e^x, e^y)$ for every $(x, y) \in \mathbb{R}^2$, we see that

$$\nabla f(C) = \{(e^\lambda, e^{2\lambda}) : \ 0 \le \lambda \le 1\}$$

is compact but clearly *not convex.*

(i) In view of Theorem 7.3 and the lack of convexity of $\nabla f(C)$, there must exist $(x, y) \in \mathbb{R}^2$ such that $\overrightarrow{P}_C^f(x, y)$ is *multi-valued.*

(ii) Since $\overleftarrow{P}_C^f(x, y)$ is a singleton for every $(x, y) \in \mathbb{R}^2$, and since $\overrightarrow{P}_{\nabla f(C)}^{f^*} = \nabla f \circ \overleftarrow{P}_C^f \circ \nabla f^*$ by Proposition 7.1 (applied to f^* and $\nabla f(C)$), we deduce that $\overrightarrow{P}_{\nabla f(C)}^{f^*}$ is single-valued on $\operatorname{int} \operatorname{dom} f^* = \{(x, y) : \ x > 0, y > 0\}$. Therefore, we obtain the striking conclusion that the analogue of Theorem 6.6 for the right Bregman projection fails even though f^* is Legendre and 1-coercive.

Acknowledgments

We thank two referees and Dan Butnariu for their pertinent comments. Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada. Jane Ye was partially supported by the Natural Sciences and Engineering Research Council of Canada. Xiaoming Yuan was partially supported by the Pacific Institute for the Mathematical Sciences, the University of Victoria, the University of British Columbia Okanagan, and the National Science Foundation of China Grant 10701055.

References

[1] E. Asplund, Chebyshev sets in Hilbert space, Trans. Amer. Math. Soc. 144 (1969) 235–240.
[2] H.H. Bauschke, J.M. Borwein, Legendre functions and the method of random Bregman projections, J. Convex Anal. 4 (1997) 27–67.
[3] H.H. Bauschke, J.M. Borwein, P.L. Combettes, Bregman monotone optimization algorithms, SIAM J. Control Optim. 42 (2003) 596–636.
[4] H.H. Bauschke, P.L. Combettes, D. Noll, Joint minimization with alternating Bregman proximity operators, Pac. J. Optim. 2 (2006) 401–424.
[5] H. Berens, U. Westphal, Kodissipative metrische Projektionen in normierten linearen Räumen, in: P.L. Butzer, B. Sz.-Nagy (Eds.), Linear Spaces and Approximation, in: ISNM, vol. 40, Birkhäuser, 1980, pp. 119–130.
[6] J.M. Borwein, Proximity and Chebyshev sets, Optim. Lett. 1 (2007) 21–32.
[7] J.M. Borwein, A.S. Lewis, Convex Analysis and Nonlinear Optimization, second ed., Springer, New York, 2006.
[8] L.N.H. Bunt, Bijdrage tot de theorie de convexe puntverzamelingen, Thesis, Univ. of Groningen, Amsterdam 1934.
[9] D. Butnariu, A.N. Iusem, Totally Convex Functions for Fixed Point Computation in Infinite Dimensional Optimization, Kluwer, Dordrecht, 2000.
[10] D. Butnariu, Y. Censor, S. Reich, Iterative averaging of entropic projections for solving stochastic convex feasibility problems, Comput. Optim. Appl. 8 (1997) 21–39.
[11] Y. Censor, S.A. Zenios, Parallel Optimization, Oxford University Press, 1997.
[12] F.H. Clarke, Optimization and Nonsmooth Analysis, SIAM, Philadelphia, 1990.
[13] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, P.R. Wolenski, Nonsmooth Analysis and Control Theory, Springer-Verlag, New York, 1998.
[14] F.H. Clarke, R.J. Stern, P.R. Wolenski, Proximal smoothness and the lower-C^2 property, J. Convex Anal. 2 (1995) 117–144.
[15] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley Interscience, New York, 1983.
[16] F. Deutsch, Best Approximation in Inner Product Spaces, Springer-Verlag, New York, 2001.
[17] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms II, Springer, New York, 1996.
[18] J.-B. Hiriart-Urruty, Potpourri of conjectures and open questions in nonlinear analysis and optimization, SIAM Rev. 49 (2007) 255–273.
[19] J.-B. Hiriart-Urruty, Ensembles de Tchebychev vs. ensembles convexes: l'etat de la situation vu via l'analyse convexe non lisse, Ann. Sci. Math. Québec 22 (1998) 47–62.
[20] P.D. Loewen, X. Wang, On the multiplicity of Dini subgradients in separable spaces, Nonlinear Anal. 58 (2004) 1–10.
[21] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation I, Springer-Verlag, Berlin, 2006.
[22] T. Motzkin, Sur quelques propriétés caractéristiques des ensembles convexes, Atti Accad. Naz. Lincei, Rend., VI. Ser. 21 (1935) 562–567.
[23] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.
[24] R.T. Rockafellar, R. J-B Wets, Variational Analysis, Springer-Verlag, New York, 1998.
[25] S. Simons, Minimax and Monotonicity, in: Lecture Notes in Mathematics, 1693, Springer-Verlag, 1998.
[26] I. Singer, Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces, in: Die Grundlehren der mathematischen Wissenschaften, Band 171, Springer-Verlag, 1970.
[27] I. Singer, The Theory of Best Approximation and Functional Analysis, in: Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, vol. 13, Society for Industrial and Applied Mathematics, 1974.

[28] V. Soloviov, Duality for nonconvex optimization and its applications, Anal. Math. 19 (1993) 297–315.
[29] L.P. Vlasov, Approximate properties of sets in normed linear spaces, Russian Math. Surveys 28 (1973) 1–66.
[30] Z. Wu, J.J. Ye, Equivalence among various derivatives and subdifferentials of the distance function, J. Math. Anal. Appl. 282 (2003) 629–647.