Klee sets and Chebyshev centers for the right Bregman distance

Heinz H. Bauschke,[*] Mason S. Macklem,[†] Jason B. Sewell,[‡] and Xianfu Wang[§]

January 20, 2010

Abstract

We systematically investigate the farthest distance function, farthest points, Klee sets, and Chebyshev centers, with respect to Bregman distances induced by Legendre functions. These objects are of considerable interest in Information Geometry and Machine Learning; when the Legendre function is specialized to the energy, one obtains classical notions from Approximation Theory and Convex Analysis.

The contribution of this paper is twofold. First, we provide an affirmative answer to a recently-posed question on whether or not every Klee set with respect to the right Bregman distance is a singleton. Second, we prove uniqueness of the Chebyshev center and we present a characterization that relates to previous works by Garkavi, by Klee, and by Nielsen and Nock.

2000 Mathematics Subject Classification: Primary 41A65, 46B20; Secondary 47H05, 49J52, 49J53, 90C25.

Keywords: Bregman distance, Chebyshev center, convex function, farthest point, Fenchel conjugate, Ioffe-Tikhomirov Theorem, Itakura-Saito distance, Klee set, Kullback-Leibler divergence, Legendre function, Mahalanobis distance, maximal monotone operator, subdifferential operator.

[*]Mathematics, Irving K. Barber School, UBC Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: `heinz.bauschke@ubc.ca`.

[†]Mathematics, Irving K. Barber School, UBC Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: `mason.macklem@ubc.ca`.

[‡]Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail: `j4sewell@uwaterloo.ca`.

[§]Mathematics, Irving K. Barber School, UBC Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: `shawn.wang@ubc.ca`.

1 Introduction

Throughout this paper,

(1) $\mathbb{R}^J$ is the standard Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$.

Suppose that S is a nonempty subset of $\mathbb{R}^J$ such that for every point in $\mathbb{R}^J$, there exists a unique farthest point in S, where "farthest" is understood in the standard Euclidean distance sense. Then S is said to be a *Klee set*, and it is known that S must be a *singleton*; see, e.g., [1, 17, 18, 20, 21] for further information. (The situation in Hilbert space remains unsettled to this day. One of the main obstacles is the possible absence of compactness of bounded closed sets in infinite-dimensional Hilbert spaces.)

In [8], Klee sets were revisited from a new perspective by using measures of fairly different from distances induced by norms. To describe and follow up on this viewpoint, we assume throughout that

(2) $$f\colon \mathbb{R}^J \to \left]-\infty,+\infty\right] \text{ is a convex function of Legendre type.}$$

Recall that for a convex function $g\colon \mathbb{R}^J \to \left]-\infty,+\infty\right]$, the *(essential) domain* is $\operatorname{dom} g = \big\{x \in \mathbb{R}^J \mid g(x) \in \mathbb{R}\big\}$ and $x^* \in \mathbb{R}^J$ is a *subgradient* of g at a point $x \in \operatorname{dom} g$, written $x^* \in \partial g(x)$, if $(\forall h \in \mathbb{R}^J)\ g(x) + \langle h, x^*\rangle \leq g(x+h)$; this induces the corresponding set-valued *subdifferential operator* $\partial g\colon \mathbb{R}^J \rightrightarrows \mathbb{R}^J$. (For basic terminology and results from Convex Analysis not stated explicitly in this paper, we refer the reader to [9, 24, 26, 28].) Then g is said to be *essentially smooth* if g is differentiable on $\operatorname{int}\operatorname{dom} g$ (the interior of its domain), and $\|\nabla g(x)\| \to +\infty$ whenever x approaches a point in the boundary $\operatorname{bdry}\operatorname{dom} g$; g is *essentially strictly convex* if g is strictly convex on every convex subset of $\operatorname{dom}\partial g = \big\{x \in \mathbb{R}^J \mid \partial g(x) \neq \varnothing\big\}$; and g is a *convex function of Legendre type* — often simply called a *Legendre function* — if g is both essentially smooth and essentially strictly convex. See [5, 10, 11, 24] for further information on Legendre functions. It will be convenient to set

(3) $$U := \operatorname{int}\operatorname{dom} f.$$

Many examples of Legendre functions exist; however, in this paper, we focus mainly on the following.

Example 1.1 (Legendre functions) The following are Legendre functions, each evaluated at a point $x \in \mathbb{R}^J$.

(i) *(halved) energy:* $f(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}\sum_j x_j^2$.

(ii) $f(x) = \frac{1}{2}\langle x, Ax\rangle$, where $A \in \mathbb{R}^{J\times J}$ is symmetric and positive definite.

(iii) *negative entropy:* $f(x) = \begin{cases} \sum_j \big(x_j \ln(x_j) - x_j\big), & \text{if } x \in \mathbb{R}^J_+; \\ +\infty, & \text{otherwise.} \end{cases}$

(iv) *negative logarithm:* $f(x) = \begin{cases} -\sum_j \ln(x_j), & \text{if } x \in \mathbb{R}^J_{++}; \\ +\infty, & \text{otherwise.} \end{cases}$

Note that $U = \mathbb{R}^J$ in (i) and (ii), whereas $U = \mathbb{R}^J_{++}$ in (iii) and (iv).

Legendre functions are of considerable interest to us because they give rise to a very nice measure of discrepancy between points, nowadays termed the "Bregman distance"; see, e.g., [12, 13, 14].

Definition 1.2 (Bregman distance) *The* Bregman distance *with respect to f, written D_f or simply D, is the function*

$$D\colon \mathbb{R}^J \times \mathbb{R}^J \to [0,+\infty] : (x,y) \mapsto \begin{cases} f(x) - f(y) - \langle \nabla f(y), x-y\rangle, & \textit{if } y \in U; \\ +\infty, & \textit{otherwise}. \end{cases} \tag{4}$$

Although well established, the term "Bregman distance" is a misnomer because a Bregman distance is in general neither symmetric nor does it satisfy the triangle inequality. However, the Bregman distance is able to distinguish different points in the sense (see [3, Theorem 3.7.(iv)]) that

$$\big(\forall x \in \mathbb{R}^J\big)\big(\forall y \in U\big) \quad D(x,y) = 0 \quad \Leftrightarrow \quad x = y. \tag{5}$$

Example 1.3 (Bregman distances) The Bregman distances corresponding to the Legendre functions of Example 1.1 between two points x and y in $\mathbb{R}^J$ are as follows.

(i) *halved squared Euclidean distance:* $D(x,y) = \frac{1}{2}\|x-y\|^2$.

(ii) *halved squared Mahalanobis distance:* $D(x,y) = \frac{1}{2}\langle x-y, A(x-y)\rangle$.

(iii) *Kullback-Leibler divergence:* $D(x,y) = \begin{cases} \sum_j \big(x_j \ln(x_j/y_j) - x_j + y_j\big), & \text{if } x \in \mathbb{R}^J_+ \text{ and } y \in \mathbb{R}^J_{++}; \\ +\infty, & \text{otherwise.} \end{cases}$

(iv) *Itakura-Saito distance:* $D(x,y) = \begin{cases} \sum_j \big(\ln(y_j/x_j) + x_j/y_j - 1\big), & \text{if } x \in \mathbb{R}^J_{++} \text{ and } y \in \mathbb{R}^J_{++}; \\ +\infty, & \text{otherwise.} \end{cases}$

See e.g. [2] and [3] for further examples of Bregman distances.

From now on, we assume that C is a subset of $\mathbb{R}^J$ such that

$$\varnothing \neq C \subseteq U. \tag{6}$$

Definition 1.4 (right Bregman farthest-distance function and farthest-point map) *The* right Bregman farthest-distance function *is*

$$\overrightarrow{F}_C \colon \mathbb{R}^J \to [0,+\infty] : x \mapsto \sup_{c\in C} D(x,c), \tag{7}$$

and the corresponding right Bregman farthest-point map *is*

$$\overrightarrow{Q}_C \colon \mathbb{R}^J \rightrightarrows \mathbb{R}^J : x \mapsto \begin{cases} \operatorname{argmax}_{c\in C} D(x,c) = \big\{c \in C \;\big|\; D(x,c) = \overrightarrow{F}_C(x)\big\}, & \text{if } x \in \operatorname{dom} f; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{8}$$

Observe that

$$\overrightarrow{F}_C \ \text{ is convex and lower semicontinuous,} \tag{9}$$

that

$$\operatorname{dom} \overrightarrow{Q}_C \subseteq \operatorname{dom} \overrightarrow{F}_C \subseteq \operatorname{dom} f, \tag{10}$$

and that

$$\text{if } C \text{ is compact, then } \operatorname{dom} \overrightarrow{Q}_C = \operatorname{dom} \overrightarrow{F}_C = \operatorname{dom} f. \tag{11}$$

We are now ready to continue the discussion on Klee sets started earlier by introducing a notion central to this paper.

Definition 1.5 ($\overrightarrow{D}$-Klee set) *The set C is said to be $\overrightarrow{D}$-Klee, if for every $x \in U$, $\overrightarrow{Q}_C x$ is a singleton.*

The asymmetry of D gives also rise to the *left* Bregman farthest-distance function and associated farthest-point map and Klee sets. These objects were analyzed in [8] and are not treated here. In fact, under additional assumptions, right and left notions may be related to each other via duality. However, the duality approach was not powerful enough to settle the question, raised in [8, Remark 7.3], *whether or not every $\overrightarrow{D}$-Klee set is a singleton when f does not have full domain* as is the case when D is, e.g., the Kullback-Leibler divergence or the Itakura-Saito distance. *The* *first contribution of this paper is to settle this question entirely*, for manifestations of f that are even more general than those considered in [8]. In fact, in Theorem 3.2 we prove that the answer is affirmative in the present setting.

Another related line of work concerns *Chebyshev centers*. Again, let us start by reviewing the classical situation in Euclidean spaces. Let S be a nonempty compact subset of $\mathbb{R}^J$. The *Chebyshev center* is the center of the smallest closed ball one can place in $\mathbb{R}^J$ that entirely captures the set S. The Chebyshev center exists and is unique, and a classical result due to Garkavi and Klee (see Corollary 4.5 below) provides a geometric characterization of it. Unlike Klee sets, Chebyshev centers have already been investigated in the context of Bregman distances — see, e.g., the work by Nielsen and Nock [22, 23] (see Corollary 4.6 below) and the references therein — however; it is assumed there that S is *finite*. *The second contribution of this paper is to extend the classical work in Euclidean space by Garkavi and Klee and the recent work by Nielsen and Nock on Chebyshev centers of finite sets with respect to Bregman distances*. In Theorem 4.4, we prove existence and uniqueness for Chebyshev centers of compact sets with respect the Bregman distance, and we present a geometric characterization of it.

The remainder of the paper is organized as follows. In Section 2 we collect and present several results that will make the proofs of the main results more structured and easier to follow. The main result in Section 3 is Theorem 3.2, which states that every compact $\overrightarrow{D}$-Klee set is indeed a singleton. In Section 4, we guarantee existence and uniqueness of the $\overrightarrow{D}$-Chebyshev center, and we characterize it geometrically. In Section 5, we illustrate our results with an example for three Bregman distances.

2 Auxiliary Results

In this section, we collect several results that will make the proofs of the main results easier to follow. We start with two identities that are straightforward consequences of (4).

Lemma 2.1 (See [15, Lemma 3.1].) *Let x be in $\mathbb{R}^J$, and let y and z be in U. Then*

$$D(x,z) - D(y,z) = D(x,y) + \langle x - y, \nabla f(y) - \nabla f(z)\rangle . \tag{12}$$

Lemma 2.2 (See [6, Remark 2.5].) *Let x_1 and x_2 be in* $\operatorname{dom} f$, *and let y_1 and y_2 be in U. Then*

$$\langle x_1 - x_2, \nabla f(y_1) - \nabla f(y_2)\rangle = D(x_2,y_1) + D(x_1,y_2) - D(x_1,y_1) - D(x_2,y_2). \tag{13}$$

Lemma 2.3 *The Bregman distance D is continuous on $U \times U$.*

Proof. This follows from [24, Theorem 10.1 and Corollary 25.5.1]. ■

Fact 2.4 (Rockafellar) (See [24, Theorem 26.5].) *The gradient operator ∇f is a continuous bijection between U and* $\operatorname{int}\operatorname{dom} f^*$, *with continuous inverse $(\nabla f)^{-1} = \nabla f^*$. Furthermore, f^* is also a convex function of Legendre type.*

Recall that a function $g\colon \mathbb{R}^J \to \,]-\infty,+\infty]$ is *coercive* if all its lower level sets are bounded; equivalently, if $\lim_{\|x\|\to+\infty} g(x) = +\infty$. The following is thus clear.

(14) If $g\colon \mathbb{R}^J \to \,]-\infty,+\infty]$ is coercive and lower semicontinuous, then $\operatorname{argmin} g \neq \varnothing$.

Here $\operatorname{argmin} g$ denotes the set of minimizers of g.

Fact 2.5 (See [24, Corollary 14.2.2].) *Let $g\colon \mathbb{R}^J \to \,]-\infty,+\infty]$ be convex, lower semicontinuous, and proper, and let $x^* \in \mathbb{R}^J$. Then $g(\,\cdot\,) - \langle\,\cdot\,, x^*\rangle$ is coercive if and only if $x^* \in \operatorname{int}\operatorname{dom} g^*$.*

Fact 2.6 (Ioffe-Tikhomirov) (See [28, Theorem 2.4.18].) *Let A be a compact Hausdorff space, let $g_a\colon \mathbb{R}^J \to \,]-\infty,+\infty]$ be convex for every $a \in A$, and set $g := \sup_{a\in A} g_a$. Assume that $(\forall x \in \mathbb{R}^J)$ $A \to \,]-\infty,+\infty] : a \mapsto g_a(x)$ is upper semicontinuous and that $x_0 \in \operatorname{dom} g$ is a point such that $(\forall a \in A)$ g_a is continuous at x_0. Then*

$$\partial g(x_0) = \overline{\operatorname{conv}} \bigcup_{\{a\in A \,|\, g(x_0)=g_a(x_0)\}} \partial g_a(x_0). \tag{15}$$

The following result is a Bregman-distance version of the well known characterization of the metric projection onto a nonempty closed convex set.

Lemma 2.7 (See [3, Proposition 3.16].) *Suppose that C is closed and convex, and let $y \in U \smallsetminus C$. Then there exists a unique point $\bar{c} \in C$ such that*

$$\big(\forall c \in C\big) \quad \langle c - \bar{c}, \nabla f(y) - \nabla f(\bar{c})\rangle \leq 0. \tag{16}$$

Lemma 2.8 *Suppose that C is compact, and let $x \in U \smallsetminus ((\nabla f^*)(\overline{\operatorname{conv}}\,(\nabla f(C))))$. Then there exists $y \in (\nabla f^*)(\overline{\operatorname{conv}}\,(\nabla f(C))) \subset U$ such that*

$$\big(\forall c \in C\big) \quad D(x,c) \geq D(x,y) + D(y,c). \tag{17}$$

Proof. Set $S := \nabla f(C)$ and $V := \operatorname{int}\operatorname{dom} f^* = \nabla f(U)$. Since C is compact and ∇f is continuous (Fact 2.4), the set S is compact. Using [24, Theorem 17.2], we deduce that $\operatorname{conv} S = \overline{\operatorname{conv}}\, S$ is a nonempty proper compact subset of V. Using Fact 2.4 again, we see that

$$(\nabla f^*)(\overline{\operatorname{conv}}\, S) \;\text{ is a proper compact subset of } U \tag{18}$$

and that $x^* := \nabla f(x) \in V \smallsetminus (\overline{\operatorname{conv}}\, S)$. Applying Lemma 2.7 (to f^*, $\overline{\operatorname{conv}}\, S$, and x^*), we obtain a point $y^* \in \overline{\operatorname{conv}}\, S$ such that

$$\big(\forall v \in \overline{\operatorname{conv}}\, S\big) \quad \langle v - y^*, \nabla f^*(x^*) - \nabla f^*(y^*)\rangle \leq 0. \tag{19}$$

Now set $y := \nabla f^*(y^*)$. Then (19) yields

$$\big(\forall c \in C\big) \quad \langle \nabla f(y) - \nabla f(c), x - y\rangle \geq 0. \tag{20}$$

Combining this with Lemma 2.1, we estimate

$$(21) \qquad \big(\forall c \in C\big) \quad D(x,c) - D(y,c) = D(x,y) + \langle \nabla f(y) - \nabla f(c), x - y\rangle \geq D(x,y),$$

which completes the proof. ■

Let X and Y be nonempty subsets of $\mathbb{R}^J$ and let $A\colon X \rightrightarrows Y$ be a set-valued operator, i.e., $(\forall x \in X)$ $Ax \subseteq Y$. Denote the *graph* of A by $\operatorname{gr} A := \big\{(x,y) \in X \times Y \;\big|\; y \in Ax\big\}$. We say that A is *monotone from X to Y*, if

$$(22) \qquad \big(\forall (x,x^*) \in \operatorname{gr} A\big)\big(\forall (y,y^*) \in \operatorname{gr} A\big) \quad \langle x - y, x^* - y^*\rangle \geq 0.$$

If A is monotone from X to Y and every proper set-valued extension from X to Y is not monotone, then A is *maximal monotone from X to Y*. If $X = Y = \mathbb{R}^J$, we will simply speak of monotone and maximal monotone operators; this is the usual and well known setting.

We now present a variant of [25, Example 12.7], which is a sufficient condition for maximal monotonicity.

Proposition 2.9 *Let S be a nonempty open subset of $\mathbb{R}^J$, let Y be a subset of $\mathbb{R}^J$, and let $A\colon S \to Y$ be monotone and continuous. Then A is maximal monotone from S to Y.*

Proof. Suppose that $(\bar{x},\bar{y}) \in S \times Y$ satisfies

$$(23) \qquad \big(\forall x \in S\big) \quad \langle \bar{x} - x, \bar{y} - Ax\rangle \geq 0,$$

and denote the closed unit ball in $\mathbb{R}^J$ by B. Then for all sufficiently small $\epsilon > 0$, we have $\bar{x} + \epsilon B \subseteq S$ and hence $(\forall b \in B)$ $\langle \bar{x} - (\bar{x} + \epsilon b), \bar{y} - A(\bar{x} + \epsilon b)\rangle \geq 0$ and so $\langle b, \bar{y} - A(\bar{x} + \epsilon b)\rangle \leq 0$. Letting $\epsilon \to 0^+$ for fixed but arbitrary $b \in B$, and using continuity of A at $\bar{x}$, we deduce that $\langle b, \bar{y} - A\bar{x}\rangle \leq 0$. Taking the supremum over $b \in B$, we obtain $\|\bar{y} - A\bar{x}\| = 0$. Hence $(\bar{x},\bar{y}) = (\bar{x}, A\bar{x}) \in \operatorname{gr} A$, as required. ■

Our first result reveals a monotonicity property of $\overrightarrow{Q}_C$. (See also [27] and [8, Proposition 7.1], and [7], where we discuss Chebyshev sets instead of Klee sets.)

Proposition 2.10 *The set-valued operator $-\nabla f \circ \overrightarrow{Q}_C\colon \mathbb{R}^J \rightrightarrows \mathbb{R}^J$ is monotone.*

Proof. Assume that (x,x^*) and (y,y^*) lie in $\operatorname{gr} \overrightarrow{Q}_C$. It follows from (8) and Lemma 2.2 (applied to $x_1 = x$, $x_2 = y$, $y_1 = y^*$, and $y_2 = x^*$) that

$$(24) \qquad \begin{aligned} 0 &\leq \big(D(x,x^*) - D(x,y^*)\big) + \big(D(y,y^*) - D(y,x^*)\big) \\ &= \langle x - y, \nabla f(y^*) - \nabla f(x^*)\rangle \\ &= \langle x - y, (-\nabla f)(x^*) - (-\nabla f)(y^*)\rangle, \end{aligned}$$

as required. ■

Proposition 2.11 *Suppose that C is closed, and that $((x_n, y_n))_{n\in\mathbb{N}}$ is a sequence in $(\operatorname{gr} \overrightarrow{Q}_C) \cap (U \times \mathbb{R}^J)$ such that $(x_n, y_n) \to (x, y) \in U \times \mathbb{R}^J$. Then $(x, y) \in \operatorname{gr} \overrightarrow{Q}_C$.*

Proof. Since $\operatorname{ran} \overrightarrow{Q}_C \subseteq C$, the sequence $(y_n)_{n\in\mathbb{N}}$ lies in C and it satisfies $(\forall n \in \mathbb{N})\ D(x_n, y_n) = \overrightarrow{F}_C(x_n)$. Because C is closed, $y \in C \subseteq U$. By Lemma 2.3, D is continuous on $U \times U$. Altogether, in view of (9), we deduce

$$\overrightarrow{F}_C(x) \leq \varliminf_{n\in\mathbb{N}} \overrightarrow{F}_C(x_n) = \varliminf_{n\in\mathbb{N}} D(x_n, y_n) = D(x, y) \leq \overrightarrow{F}_C(x). \tag{25}$$

Therefore, $\overrightarrow{F}_C(x) = D(x, y)$, i.e., $y \in \overrightarrow{Q}_C(x)$. ■

Proposition 2.12 *Suppose that $\overline{C} \subseteq U$. Then $\operatorname{gr} \overrightarrow{Q}_C \subseteq \operatorname{gr} \overrightarrow{Q}_{\overline{C}}$.*

Proof. Take $(x, y) \in \operatorname{gr} \overrightarrow{Q}_C$. Then $y \in C \subseteq \overline{C}$ and $(\forall c \in C)\ D(x, y) \geq D(x, c)$. Since $\overline{C} \subseteq U$ and $D(x, \cdot)$ is continuous on U (Lemma 2.3), it follows that $(\forall \bar{c} \in \overline{C})\ D(x, y) \geq D(x, \bar{c})$. Thus, $y \in \overrightarrow{Q}_{\overline{C}}(x)$. ■

Remark 2.13 Assume that $\bar{c} \in \overline{C} \cap \operatorname{bdry} U$. In view of (6), there exists a sequence $(c_n)_{n\in\mathbb{N}}$ in $C \subseteq U$ such that $c_n \to \bar{c}$; hence, by [3, Theorem 3.8.(i)], $(\forall x \in U)\ D(x, c_n) \to +\infty$. Therefore, the assumption that $\overline{C}$ be a subset of U is very natural in Proposition 2.12 and elsewhere in this paper.

Proposition 2.14 *Suppose that $(\forall x \in \operatorname{dom} f)\ D(x, \cdot\,)$ is convex on U. Then $\operatorname{gr} \overrightarrow{Q}_C \subseteq \operatorname{gr} \overrightarrow{Q}_{\operatorname{conv} C}$.*

Proof. Take $(x, y) \in \operatorname{gr} \overrightarrow{Q}_C$. Then $x \in \operatorname{dom} f$, $y \in C \subseteq \operatorname{conv} C$, and $(\forall c \in C)\ D(x, c) \leq D(x, y)$. Now let $z \in \operatorname{conv} C$, say $z = \sum_{i=1}^n \lambda_i c_i$, where each $\lambda_i \in [0, 1]$, each $c_i \in C$, and $\sum_{i=1}^n \lambda_i = 1$. Then $D(x, z) \leq \sum_{i=1}^n \lambda_i D(x, c_i) \leq \sum_{i=1}^n \lambda_i D(x, y) = D(x, y)$ and therefore $y \in \overrightarrow{Q}_{\operatorname{conv} C}(x)$. ■

Proposition 2.15 *Suppose that $(\forall x \in \operatorname{dom} f)\ D(x, \cdot\,)$ is strictly convex on U. Then $\overrightarrow{Q}_C = \overrightarrow{Q}_{\operatorname{conv} C}$.*

Proof. In view of Proposition 2.14, we need to show only that $\operatorname{gr} \overrightarrow{Q}_{\operatorname{conv} C} \subseteq \operatorname{gr} \overrightarrow{Q}_C$. To this end, let $(x, y) \in \operatorname{gr} \overrightarrow{Q}_{\operatorname{conv} C}$. Then $x \in \operatorname{dom} f$, $y \in \operatorname{conv} C$, and $(\forall s \in \operatorname{conv} C)\ D(x, s) \leq D(x, y)$. In particular, $(\forall c \in C)\ D(x, c) \leq D(x, y)$. The proof is complete as soon as we have verified that $y \in C$. Assume to the contrary that $y \notin C$. Then $y = \sum_{i=1}^n \lambda_i c_i$, where $n \geq 2$, each $\lambda_i > 0$, each $c_i \in C$, and where the c_i are pairwise distinct and $\sum_{i=1}^n \lambda_i = 1$. But then $D(x, y) < \sum_{i=1}^n \lambda_i D(x, c_i) \leq \sum_{i=1}^n \lambda_i D(x, y) = D(x, y)$, which is absurd. ■

The next result shows that when D is separately convex (see [4] for a systematic discussion of separate and joint convexity of D), then the farthest-point distance is "blind" to the convex hull.

Proposition 2.16 *Suppose that $(\forall x \in \operatorname{dom} f)\ D(x, \cdot\,)$ is convex. Then $\overrightarrow{F}_{\operatorname{conv} C} = \overrightarrow{F}_C$.*

Proof. This follows from [24, Theorem 32.2]. ■

The following result was motivated by a question posed by one of the referees.

Proposition 2.17 *Suppose that C is compact and that $(\forall x \in \operatorname{dom} f)$ $D(x,\cdot)$ is strictly convex on U. Denote the set of extreme points of* $\operatorname{conv} C$ *by E. Then $\overrightarrow{Q}_C = \overrightarrow{Q}_{\operatorname{conv} C} = \overrightarrow{Q}_E$ and* $\operatorname{ran} \overrightarrow{Q}_C \subseteq E$.

Proof. Denote the set of extreme points of $\operatorname{conv} C$ by E. By [24, Corollary 18.3.1], $E \subseteq C$. Furthermore, [24, Theorem 17.2] implies that $\operatorname{conv} C$ is compact. In turn, [24, Corollary 18.5.1] yields $\operatorname{conv} C = \operatorname{conv} E$. Applying Proposition 2.15 twice, we deduce that $\overrightarrow{Q}_C = \overrightarrow{Q}_{\operatorname{conv} C} = \overrightarrow{Q}_{\operatorname{conv} E} = \overrightarrow{Q}_E$. Therefore, $\operatorname{ran} \overrightarrow{Q}_C = \operatorname{ran} \overrightarrow{Q}_E \subseteq E$. ■

3 Klee Sets are Singletons

The following result will be critical in the proof of our first main result (Theorem 3.2).

Theorem 3.1 *Suppose that C is compact. Then* $\operatorname{argmin} \overrightarrow{F}_C$ *is a nonempty subset of U.*

Proof. By (11), $\operatorname{dom} \overrightarrow{F}_C = \operatorname{dom} f$. Since $C \subset U$, it follows from Fact 2.4 that $\nabla f(C) \subset \nabla f(U) = \operatorname{int} \operatorname{dom} f^*$. In view of Fact 2.5, we deduce that

$$(26) \qquad \big(\forall c \in C\big) \quad f(\,\cdot\,) - \langle\,\cdot\,, \nabla f(c)\rangle \quad \text{is coercive.}$$

Since $(\forall c \in C)$ $D(\,\cdot\,, c) = (f(\,\cdot\,) - \langle\,\cdot\,, \nabla f(c)\rangle) + (\langle c, \nabla f(c)\rangle - f(c))$, it follows that

$$(27) \qquad \big(\forall c \in C\big) \quad D(\,\cdot\,, c) \quad \text{is coercive.}$$

In turn, this implies that

$$(28) \qquad \overrightarrow{F}_C(\,\cdot\,) = \sup_{c \in C} D(\,\cdot\,, c) \quad \text{is coercive, convex, lower semicontinuous, and proper.}$$

In view of (28) and (14), $\operatorname{argmin} \overrightarrow{F}_C \neq \varnothing$. Let

$$(29) \qquad x_0 \in \operatorname{argmin} \overrightarrow{F}_C.$$

It suffices to show that

$$(30) \qquad x_0 \in U.$$

Assume to the contrary that $x_0 \notin U$. In view of (10) and (29), $x_0 \in (\operatorname{dom} f \smallsetminus U) \subseteq \operatorname{bdry} \operatorname{dom} f$. Now fix an arbitrary point $x_1 \in U$ and set

$$(31) \qquad \big(\forall \epsilon \in\,]0,1[\,\big) \quad x_\epsilon := (1-\epsilon)x_0 + \epsilon x_1.$$

By [24, Theorem 6.1], $(\forall \epsilon \in \,]0,1])$, $x_\epsilon \in U$. Set $S := \nabla f(C)$. As already observed in the proof of Lemma 2.8, $\operatorname{conv} S = \overline{\operatorname{conv}}\, S$ is a proper compact subset of $\operatorname{int}\operatorname{dom} f^*$. Thus, there exists $\bar{\epsilon} \in \,]0,1]$ such that $(\forall \epsilon \in \,]0,\bar{\epsilon}])$ $x_\epsilon \in U \smallsetminus (\nabla f^*)(\overline{\operatorname{conv}}\, S)$. Lemma 2.8 now yields

$$(32) \qquad \big(\forall \epsilon \in \,]0,\bar{\epsilon}]\,\big)\big(\exists y_\epsilon \in (\nabla f^*)(\overline{\operatorname{conv}}\, S)\big)\big(\forall c \in C\big) \quad D(x_\epsilon, c) \geq D(x_\epsilon, y_\epsilon) + D(y_\epsilon, c).$$

On the one hand, while f is not necessarily continuous at x_0, it is at least *continuous along the line segment* $[x_0, x_1]$ (see [24, Theorem 7.5]); consequently,

$$(33) \qquad \lim_{\epsilon \to 0^+} f(x_\epsilon) = f(x_0).$$

On the other hand, the net $(y_\epsilon)_{\epsilon \in]0,\bar{\epsilon}]}$ lies in $\nabla f^*(\overline{\operatorname{conv}}\, S)$, which is a compact set. After passing to a subnet and relabeling if necessary, we assume that there exists a point $y_0 \in \mathbb{R}^J$ such that

$$(34) \qquad \lim_{\epsilon \to 0^+} y_\epsilon = y_0 \in \nabla f^*(\overline{\operatorname{conv}}\, S) \subset U.$$

Combining (33) and (34), invoking Lemma 2.3, and taking the limit in (32), we obtain altogether that

$$(35) \qquad \big(\forall c \in C\big) \quad D(x_0, c) \geq D(x_0, y_0) + D(y_0, c).$$

Since $x_0 \in \operatorname{bdry}\operatorname{dom} f$ and $y_0 \in \operatorname{int}\operatorname{dom} f = U$, (5) results in $D(x_0, y_0) > 0$. Taking the supremum in (35) over $c \in C$, we deduce that

$$(36) \qquad \overrightarrow{F}_C(x_0) \geq D(x_0, y_0) + \overrightarrow{F}_C(y_0) > \overrightarrow{F}_C(y_0),$$

which contradicts (29). Therefore, we have verified (30), and the proof is complete. ■

Theorem 3.2 (every $\overrightarrow{D}$-Klee set is a singleton) *Suppose that C is compact and that C is $\overrightarrow{D}$-Klee. Then C is a singleton.*

Proof. Recall that

$$(37) \qquad \overrightarrow{F}_C(\,\cdot\,) = \sup_{c \in C} D(\,\cdot\,, c) = \sup_{c \in C} \Big(\big(f(\,\cdot\,) - \langle\,\cdot\,, \nabla f(c)\rangle\,\big) + \big(\,\langle c, \nabla f(c)\rangle - f(c)\big)\Big).$$

Because C is $\overrightarrow{D}$-Klee, if $x \in U$, then $\overrightarrow{Q}_C x$ is the unique point in C such that $\overrightarrow{F}_C(x) = D(x, \overrightarrow{Q}_C x)$ and $(\forall c \in C \smallsetminus \{\overrightarrow{Q}_C x\})$ $\overrightarrow{F}_C(x) > D(x, c)$. In view of Theorem 3.1, we take $x_0 \in \operatorname{argmin} \overrightarrow{F}_C \subset U$. Using the Fact 2.6, we obtain

$$(38) \qquad 0 \in \partial \overrightarrow{F}_C(x_0) = \nabla f(x_0) - \nabla f(\overrightarrow{Q}_C x_0).$$

Hence $\nabla f(x_0) = \nabla f(\overrightarrow{Q}_C(x_0))$ and thus $x_0 = \overrightarrow{Q}_C(x_0)$. Therefore, $C = \{x_0\}$. ■

Corollary 3.3 (Klee) *Suppose that C is a compact Klee set with respect to the Euclidean distance. Then C is a singleton.*

Proof. (See also [20].) This follows from Theorem 3.2 when $f = \frac{1}{2}\|\cdot\|^2$. ■

We conclude this section with two results concerning $\overrightarrow{D}$-Klee sets that are not assumed to be compact. When considering classical Klee sets, a standard assumption is *closedness*. The next result illustrates this assumption in the present Bregman distance setting.

Proposition 3.4 *Suppose that $\overline{C}$ is a compact subset of U, and that $U \subseteq \operatorname{dom} \overrightarrow{Q}_C$. Then $\overline{C}$ is $\overrightarrow{D}$-Klee if and only if C is $\overrightarrow{D}$-Klee and $\overrightarrow{Q}_C$ is continuous on U.*

Proof. "$\Rightarrow$": Since $\overline{C}$ is compact, Theorem 3.2 implies that $\overline{C}$ is a singleton, say $\overline{C} = \{y\}$. But then $C = \{y\} = \overline{C}$ is also $\overrightarrow{D}$-Klee, and $\overrightarrow{Q}_C|_U \equiv \{y\}$ is clearly continuous on U.

"$\Leftarrow$": Proposition 2.10 implies that both $-\nabla f \circ \overrightarrow{Q}_C\big|_U$ and $-\nabla f \circ \overrightarrow{Q}_{\overline{C}}\big|_U$ are monotone from U to $\mathbb{R}^J$. Furthermore, since $\overrightarrow{Q}_C$ is continuous on U, so is $-\nabla f \circ \overrightarrow{Q}_C\big|_U$. Thus, by Proposition 2.9, $-\nabla f \circ \overrightarrow{Q}_C\big|_U$ is maximal monotone from U to $\mathbb{R}^J$. On the other hand, Proposition 2.12 implies that

$$\operatorname{gr}\big(-\nabla f \circ \overrightarrow{Q}_C\big|_U\big) \subseteq \operatorname{gr}\big(-\nabla f \circ \overrightarrow{Q}_{\overline{C}}\big|_U\big). \tag{39}$$

Altogether, $-\nabla f \circ \overrightarrow{Q}_C\big|_U = -\nabla f \circ \overrightarrow{Q}_{\overline{C}}\big|_U$, which yields $\overrightarrow{Q}_C\big|_U = \overrightarrow{Q}_{\overline{C}}\big|_U$. Since $\overrightarrow{Q}_C\big|_U$ is single-valued, so is $\overrightarrow{Q}_{\overline{C}}\big|_U$. Therefore, $\overline{C}$ is $\overrightarrow{D}$-Klee. ■

If the underlying Bregman distance D is strictly convex in the second variable, then we obtain the following result.

Proposition 3.5 *Suppose that $(\forall x \in U)\ D(x,\cdot)$ is strictly convex on U. Then $\operatorname{conv} C$ is $\overrightarrow{D}$-Klee if and only if C is $\overrightarrow{D}$-Klee.*

Proof. This is an immediate consequence of Proposition 2.15. ■

4 Characterization of Chebyshev Centers

The proof of our second main result (Theorem 4.4) relies upon the next two results.

Proposition 4.1 *Suppose that C is compact. Then $\overrightarrow{F}_C$ is proper, lower semicontinuous, and convex, with $\operatorname{dom} \overrightarrow{F}_C = \operatorname{dom} f = \operatorname{dom} \overrightarrow{Q}_C$. Furthermore, $\overrightarrow{F}_C$ is strictly convex on $\operatorname{dom} \partial \overrightarrow{F}_C = \operatorname{int} \operatorname{dom} f = U$.*

Proof. We observed already (see (9) and (11)) that $\overrightarrow{F}_C$ is convex and lower semicontinuous, and that $\operatorname{dom} \overrightarrow{F}_C = \operatorname{dom} f = \operatorname{dom} \overrightarrow{Q}_C$. Hence $\overrightarrow{F}_C$ is proper. Now set

$$g\colon \mathbb{R}^J \to \left]-\infty,+\infty\right] : x \mapsto \max_{c\in C}\big(\langle c - x, \nabla f(c)\rangle - f(c)\big), \tag{40}$$

and note that g is convex with $\operatorname{dom} g = \mathbb{R}^J = \operatorname{int}\operatorname{dom}\partial g$ (see [24, Theorem 23.4]). Furthermore,

$$\overrightarrow{F}_C = f + g. \tag{41}$$

By the subdifferential sum rule (see [24, Theorem 23.8]), we have $\partial \overrightarrow{F}_C = \partial f + \partial g$ and hence $\operatorname{dom}\partial\overrightarrow{F}_C = \operatorname{dom}(\partial f) \cap \operatorname{dom}(\partial g) = \operatorname{dom}(\partial f) \cap \mathbb{R}^J = \operatorname{dom}\partial f$. On the other hand, since f is a Legendre function, it follows from [24, Theorem 26.1] that $\operatorname{dom}\partial f = \operatorname{int}\operatorname{dom} f$. Altogether, $\operatorname{dom}\partial\overrightarrow{F}_C = \operatorname{int}\operatorname{dom} f = U$. Using once more the assumption that f is a Legendre function, we have that f is strictly convex on $\operatorname{int}\operatorname{dom} f = U$, and therefore so is $\overrightarrow{F}_C = f + g$. ■

Recall that for a proper convex function $g : \mathbb{R}^J \to \left]-\infty, +\infty\right]$, the *directional derivative* of g at $x \in \operatorname{dom} g$ in the direction $h \in \mathbb{R}^J$ is defined by

$$g'(x;h) = \lim_{t\to 0^+} \frac{g(x+th) - g(x)}{t}. \tag{42}$$

Theorem 4.2 (directional derivative) *Suppose that C is compact, let $x \in \operatorname{dom} f$, and let $h \in \mathbb{R}^J$. Then*

$$\overrightarrow{F}_C'(x;h) = \sup\left\{ f'(x;h) - \langle h, \nabla f(y)\rangle \;\middle|\; y \in \overrightarrow{Q}_C(x) \right\}. \tag{43}$$

If $x \notin U$ and $x + h \in U$, then $\overrightarrow{F}_C'(x;h) = -\infty$.

Proof. Recall that $\operatorname{dom}\overrightarrow{F}_C = \operatorname{dom} f = \operatorname{dom}\overrightarrow{Q}_C$ by Proposition 4.1, so let $y \in \overrightarrow{Q}_C(x)$. Then

$$(\forall t > 0) \quad \overrightarrow{F}_C(x+th) \geq D(x+th, y) = f(x+th) - f(y) - \langle x + th - y, \nabla f(y)\rangle \tag{44}$$

and

$$\overrightarrow{F}_C(x) = D(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle. \tag{45}$$

Hence, $(\forall t > 0)$ $\overrightarrow{F}_C(x+th) - \overrightarrow{F}_C(x) \geq f(x+th) - f(x) - \langle th, \nabla f(y)\rangle$. Dividing by t and taking the infimum over $t > 0$ yields

$$\overrightarrow{F}_C'(x;h) \geq f'(x;h) - \langle h, \nabla f(y)\rangle. \tag{46}$$

Taking the supremum over $y \in \overrightarrow{Q}_C(x)$ yields

$$\overrightarrow{F}_C'(x;h) \geq \sup\left\{ f'(x;h) - \langle h, \nabla f(y)\rangle \;\middle|\; y \in \overrightarrow{Q}_C(x) \right\}. \tag{47}$$

If $[x, x+h] \cap \operatorname{dom} f = \{x\}$, then $f'(x;h) = +\infty$; hence, (43) follows from (47). Thus, we assume that $[x, x+h] \cap \operatorname{dom} f$ contains a nontrivial line segment. Let $(t_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 1\right[$ such that $t_n \to 0^+$ and $(x + t_n h)_{n\in\mathbb{N}}$ lies in $\operatorname{dom} f$. Furthermore, for every $n \in \mathbb{N}$, let $c_n \in \overrightarrow{Q}_C(x + t_n h)$. After passing to a subsequence and relabeling if necessary, we also assume that $c_n \to \bar{c} \in C$. Then, for every $n \in \mathbb{N}$,

$$\overrightarrow{F}_C(x + t_n h) = D(x + t_n h, c_n) = f(x + t_n h) - f(c_n) - \langle x + t_n h - c_n, \nabla f(c_n)\rangle \tag{48}$$

and $\overrightarrow{F}_C(x) \geq D(x, c_n) = f(x) - f(c_n) - \langle x - c_n, \nabla f(c_n)\rangle$; consequently,

$$\frac{\overrightarrow{F}_C(x + t_n h) - \overrightarrow{F}_C(x)}{t_n} \leq \frac{f(x + t_n h) - f(x)}{t_n} - \langle h, \nabla f(c_n)\rangle \,. \tag{49}$$

Letting $n \to +\infty$ in (49), we deduce that

$$\overrightarrow{F}'_C(x; h) \leq f'(x; h) - \langle h, \nabla f(\bar{c})\rangle \,. \tag{50}$$

On the other hand, using line segment continuity of f and $\overrightarrow{F}_C$ at x (see [24, Corollary 7.5.1]), and continuity of both f and ∇f on U, we see that letting $n \to +\infty$ in (48) yields $\overrightarrow{F}_C(x) = D(x, \bar{c})$. Hence $\bar{c} \in \overrightarrow{Q}_C(x)$. It thus follows from (50) that $\overrightarrow{F}'_C(x; h) \leq \sup \big\{ f'(x; h) - \langle h, \nabla f(y)\rangle \;\big|\; y \in \overrightarrow{Q}_C(x) \big\}$. Combining this with (47), we deduce that (43) holds. The "If" statement follows from (43) and [24, Theorem 23.3]. ■

Theorem 4.3 (subdifferential) *Suppose that C is compact, and let $x \in U$. Then*

$$\partial \overrightarrow{F}_C(x) = \nabla f(x) - \operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(x)\big). \tag{51}$$

Proof. By Theorem 4.2 and [24, Theorem 23.4], $\overrightarrow{F}'_C(x; \cdot)$ is the support function of both $\nabla f(x) - \nabla f\big(\overrightarrow{Q}_C(x)\big)$ and $\partial \overrightarrow{F}_C(x)$. Therefore, the latter set (which is closed and convex already) is the closed convex hull of the former set. Since $\overrightarrow{Q}_C(x)$ is a compact subset of U by Proposition 2.11, it follows from the continuity of ∇f on U and from [24, Theorem 17.2] that $\overline{\operatorname{conv}}\, \nabla f\big(\overrightarrow{Q}_C(x)\big) = \operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(x)\big)$. This completes the proof. ■

Theorem 4.4 (uniqueness and characterization of the $\overrightarrow{D}$-Chebyshev center) *Suppose that C is compact. Then $\overrightarrow{F}_C$ has a unique minimizer $x \in \operatorname{dom} f$, called the* $\overrightarrow{D}$-Chebyshev center *of C, and characterized by*

$$x \in \nabla f^*\Big(\operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(x)\big)\Big). \tag{52}$$

Proof. Theorem 3.1 states that $\operatorname{argmin} \overrightarrow{F}_C$ is a nonempty subset of U. In view of the strict convexity of $\overrightarrow{F}_C$ on U (Proposition 4.1), $\operatorname{argmin} \overrightarrow{F}_C$ is a singleton, say $\{x\}$. By Theorem 4.3, $0 \in \partial \overrightarrow{F}_C(x) = \nabla f(x) - \operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(x)\big)$ and thus $\nabla f(x) \in \operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(x)\big)$. Now apply Fact 2.4. ■

Corollary 4.5 (Garkavi-Klee) (See [16] and also [19].) *Suppose that C is compact and that $x \in \mathbb{R}^J$. Then x is the Chebyshev center of C with respect to the Euclidean distance if and only if*

$$x \in \operatorname{conv} \overrightarrow{Q}_C(x). \tag{53}$$

Corollary 4.6 (Nock-Nielsen) (See [23] and also [22].) *Suppose that C is finite. Then the $\overrightarrow{D}$-Chebyshev center of C is the unique point $x \in U$ characterized by*

$$x \in \nabla f^*\Big(\operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(x)\big)\Big). \tag{54}$$

Corollary 4.7 *Suppose that C is compact and that it contains at least 2 points, and let $x \in U$ be the $\overrightarrow{D}$-Chebyshev center of C. Then $\overrightarrow{Q}_C(x)$ must contain at least 2 points.*

Proof. Suppose to the contrary that $\overrightarrow{Q}_C(x)$ is a singleton. Then (52) implies that $\overrightarrow{Q}_C(x) = \{x\}$, i.e., that x is its own farthest point in C. In view of (5) and the assumption that C contains a point different from x, this is absurd. ■

5 Constructing and Visualizing Chebyshev Centers

We work in the Euclidean plane, i.e., we assume that $J = 2$, and we let D be the halved squared Euclidean distance, the Kullback-Leibler divergence, or the Itakura-Saito distance (see Example 1.3). Set

(55) $$\mathbf{c}_0 = (1, a) \quad \text{and} \quad \mathbf{c}_1 = (a, 1), \quad \text{where} \quad a \in \left]1, +\infty\right[,$$

and

(56) $$(\forall \lambda \in \mathbb{R}) \quad \mathbf{c}_\lambda = (1-\lambda)\mathbf{c}_0 + \lambda \mathbf{c}_1.$$

Furthermore, we assume that

(57) $$\begin{aligned} C = \operatorname{conv}\{\mathbf{c}_0, \mathbf{c}_1\} &= \big\{\mathbf{c}_\lambda \;\big|\; \lambda \in [0,1]\big\} = \big\{\big(1-\lambda+\lambda a, (1-\lambda)a + \lambda\big) \;\big|\; \lambda \in [0,1]\big\} \\ &= \big\{((a-1)\lambda + 1, (1-a)\lambda + a) \;\big|\; \lambda \in [0,1]\big\}. \end{aligned}$$

Note that $C \subset \mathbb{R}^2_{++} \subseteq U$, and that C is compact and convex. In view of Theorem 4.4, the $\overrightarrow{D}$-Chebyshev center $\mathbf{z}$ of C is characterized by

(58) $$\mathbf{z} \in \nabla f^*\Big(\operatorname{conv} \nabla f\big(\overrightarrow{Q}_C(\mathbf{z})\big)\Big).$$

Our aim in this section is to determine $\mathbf{z}$ and related objects, and to visualize them. It will be convenient to set

(59) $$\Delta = \big\{(x, x) \;\big|\; x \in \mathbb{R}\big\}.$$

Relying heavily on symmetries of the Bregman distance D and the set C (both with respect to switching coordinates), we now start the process of pinpointing $\mathbf{z}$.

Proposition 5.1 $\mathbf{z} \in \Delta$.

Proof. For $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, set $\mathbf{x}^\intercal = (x_2, x_1)$. Observe that for the choices of D considered in this section, $(\forall \mathbf{x} \in \mathbb{R}^2)(\forall \mathbf{y} \in \mathbb{R}^2)\ D(\mathbf{x}, \mathbf{y}) = D(\mathbf{x}^\intercal, \mathbf{y}^\intercal)$ and that $C^\intercal = \big\{\mathbf{c}^\intercal \;\big|\; \mathbf{c} \in C\big\} = C$. Thus, $(\forall \mathbf{x} \in \mathbb{R}^2)\ \overrightarrow{F}_C(\mathbf{x}) = \overrightarrow{F}_C(\mathbf{x}^\intercal)$. Since $\mathbf{z}$ is the *unique* minimizer of $\overrightarrow{F}_C$, we must have that $\mathbf{z} = \mathbf{z}^\intercal$, i.e., that $\mathbf{z} \in \Delta$. ■

Example 5.2 (halved squared Euclidean distance) Suppose D is as in Example 1.3(i), and let $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$. Then

$$\overrightarrow{Q}_C(\mathbf{x}) = \begin{cases} \{\mathbf{c}_0\}, & \text{if } x_2 < x_1; \\ \{\mathbf{c}_1\}, & \text{if } x_1 < x_2; \\ \{\mathbf{c}_0, \mathbf{c}_1\}, & \text{if } x_1 = x_2, \end{cases} \tag{60}$$

and $\mathbf{z} = \mathbf{c}_{1/2} = \big(\frac{1}{2}(1+a), \frac{1}{2}(1+a)\big)$.

Proof. Set

$$d_{\mathbf{x}} \colon \mathbb{R} \to [0, +\infty[\,:\, \lambda \mapsto D(\mathbf{x}, c_\lambda). \tag{61}$$

Then for every $\lambda \in \mathbb{R}$, we have

$$d_{\mathbf{x}}(\lambda) = (a-1)^2\lambda^2 + (1-a)(x_1 - x_2 + a - 1)\lambda + \frac{(x_1-1)^2 + (x_2-a)^2}{2}, \tag{62}$$

$$d'_{\mathbf{x}}(\lambda) = (x_1 - x_2 - 1 + a)(1-a) + 2\lambda(1-a)^2 \text{ and } d''_{\mathbf{x}}(\lambda) = 2(a-1)^2. \tag{63}$$

Hence $\overrightarrow{Q}_C(\mathbf{x}) \subseteq \{\mathbf{c}_0, \mathbf{c}_1\}$. Since $d_{\mathbf{x}}(0) - d_{\mathbf{x}}(1) = (1-a)(x_2 - x_1)$, we obtain (60). Furthermore, since C is convex and $\mathbf{c}_{1/2} \in \Delta$, we have $\mathbf{c}_{1/2} \in C = \operatorname{conv}\{\mathbf{c}_0, \mathbf{c}_1\} = \operatorname{conv} \overrightarrow{Q}_C(\mathbf{c}_{1/2})$. Therefore, the characterization (58) of $\mathbf{z}$ yields $\mathbf{z} = \mathbf{c}_{1/2}$. (Alternatively, one may verify that $\mathbf{c}_{1/2}$ is the unique minimizer of the function $\Delta \to [0, +\infty[\,:\, (x,x) \mapsto d_{(x,x)}(0) = \overrightarrow{F}_C(x,x)$.) ■

Example 5.3 (Kullback-Leibler divergence) Suppose D is as in Example 1.3(iii), and let $\mathbf{x} = (x_1, x_2) \in U$. Then

$$\overrightarrow{Q}_C(\mathbf{x}) = \begin{cases} \{\mathbf{c}_0\}, & \text{if } x_2 < x_1; \\ \{\mathbf{c}_1\}, & \text{if } x_1 < x_2; \\ \{\mathbf{c}_0, \mathbf{c}_1\}, & \text{if } x_1 = x_2, \end{cases} \tag{64}$$

and $\mathbf{z} = \big(\sqrt{a}, \sqrt{a}\,\big)$.

Proof. Set

$$d_{\mathbf{x}} \colon \mathbb{R} \to [0, +\infty] : \lambda \mapsto D(\mathbf{x}, \mathbf{c}_\lambda). \tag{65}$$

Then $\operatorname{dom} d_{\mathbf{x}} = \big\{\lambda \in \mathbb{R} \mid \mathbf{c}_\lambda \in U\big\} = \,]-1/(a-1), a/(a-1)[\, \supset [0,1]$. For every $\lambda \in \operatorname{dom} d_{\mathbf{x}}$, we have

$$d_{\mathbf{x}}(\lambda) = -x_1 \ln\left(\frac{(a-1)\lambda + 1}{x_1}\right) - x_1 + 1 - x_2 \ln\left(\frac{(1-a)\lambda + a}{x_2}\right) - x_2 + a, \tag{66}$$

$$d'_{\mathbf{x}}(\lambda) = -\frac{x_1(a-1)}{(a-1)\lambda+1} - \frac{x_2(1-a)}{(1-a)\lambda+a}, \tag{67}$$

and

$$d''_{\mathbf{x}}(\lambda) = \frac{x_1(a-1)^2}{\big((a-1)\lambda+1\big)^2} + \frac{x_2(1-a)^2}{\big((1-a)\lambda+a\big)^2} > 0. \tag{68}$$

Thus, $d_{\mathbf{x}}$ has no local maximizers in $\operatorname{dom} d_{\mathbf{x}}$ and therefore $\overrightarrow{Q}_C(\mathbf{x}) \subseteq \{\mathbf{c}_0, \mathbf{c}_1\}$. Because of

$$D(\mathbf{x}, \mathbf{c}_0) - D(\mathbf{x}, \mathbf{c}_1) = d_{\mathbf{x}}(0) - d_{\mathbf{x}}(1) = (x_1 - x_2)\ln(a), \tag{69}$$

we see that (64) must hold. Finally, (64) implies that

$$\begin{aligned}
\big(\sqrt{a}, \sqrt{a}\,\big) &= \Big(\exp\big(\tfrac{1}{2}(0+\ln(a))\big), \exp\big(\tfrac{1}{2}(\ln(a)+0)\big)\Big) \\
&= (\exp\times\exp)\Big(\tfrac{1}{2}\big(\ln(1),\ln(a)\big) + \tfrac{1}{2}\big(\ln(a),\ln(1)\big)\Big) \\
&= \nabla f^*\big(\tfrac{1}{2}\nabla f(\mathbf{c}_0) + \tfrac{1}{2}\nabla f(\mathbf{c}_1)\big) \\
&\in \nabla f^*\Big(\operatorname{conv}\nabla f\Big(\overrightarrow{Q}_C\big(\sqrt{a}, \sqrt{a}\,\big)\Big)\Big).
\end{aligned} \tag{70}$$

In view of the characterization (58) of $\mathbf{z}$, we deduce that $\mathbf{z} = \big(\sqrt{a}, \sqrt{a}\,\big)$. ■

Remark 5.4

(i) The fact that the extreme points $\{\mathbf{c}_0, \mathbf{c}_1\}$ play a role in Example 5.2 and Example 5.3 is not surprising since in these cases $D(\mathbf{x}, \cdot)$ is convex for every $\mathbf{x} \in U$ (see, e.g., [4]) so that [24, Corollary 32.3.2] applies.

(ii) Note that $\mathbf{z}$ is the (coordinate-wise) arithmetic mean of $\mathbf{c}_0$ and $\mathbf{c}_1$ when D is the halved squared Euclidean distance (Example 5.2), and that $\mathbf{z}$ is the (coordinate-wise) geometric mean of $\mathbf{c}_0$ and $\mathbf{c}_1$ when D is the Kullback-Leibler divergence (Example 5.3). This might nurture the conjecture that $\mathbf{z}$ is the (coordinate-wise) harmonic mean of $\mathbf{c}_0$ and $\mathbf{c}_1$ for the Itakura-Saito distance — depending on the location of a, this is sometimes but not always the case (see Example 5.5 and Lemma 5.6).

Example 5.5 (Itakura-Saito distance) Suppose that D is as in Example 1.3(iv). Set

$$g = g(a) = \frac{a(a+1)}{(a-1)^2}\ln\left(\frac{(a+1)^2}{4a}\right) \quad \text{and} \quad h = h(a) = \frac{2a}{a+1}. \tag{71}$$

Then

$$\mathbf{z} = \begin{cases} (h,h), & \text{if } g < h; \\ (g,g), & \text{if } g \geq h; \end{cases} \quad \text{and} \quad \overrightarrow{Q}_C(\mathbf{z}) = \begin{cases} \{\mathbf{c}_0, \mathbf{c}_1\}, & \text{if } g < h; \\ \{\mathbf{c}_0, \mathbf{c}_{1/2}, \mathbf{c}_1\}, & \text{if } g \geq h. \end{cases} \tag{72}$$

Proof. Set

$$\mathbf{g} = (g, g) \quad \text{and} \quad \mathbf{h} = (h, h), \tag{73}$$

and note that a straightforward computation yields

$$\Delta \cap \nabla f^*\Big(\operatorname{conv} \nabla f\big(\{\mathbf{c}_0, \mathbf{c}_1\}\big)\Big) = \{\mathbf{h}\}. \tag{74}$$

Let $\mathbf{x} = (x, x) \in U \cap \Delta$ and set

$$d_{\mathbf{x}} \colon \mathbb{R} \to [0, +\infty] \colon \lambda \mapsto D(\mathbf{x}, c_\lambda). \tag{75}$$

Then $\operatorname{dom} d_{\mathbf{x}} = \big\{\lambda \in \mathbb{R} \;\big|\; \mathbf{c}_\lambda \in U\big\} = \left]-1/(a-1), a/(a-1)\right[\supset [0,1]$. For every $\lambda \in \operatorname{dom} d_{\mathbf{x}}$, we have

$$d_{\mathbf{x}}(\lambda) = \ln\left(\frac{(a-1)\lambda+1}{x}\right) + \frac{x}{(a-1)\lambda+1} + \ln\left(\frac{(1-a)\lambda+a}{x}\right) + \frac{x}{(1-a)\lambda+a} - 2. \tag{76}$$

and hence

$$d'_{\mathbf{x}}(\lambda) = \frac{a-1}{(a-1)\lambda+1} - \frac{x(a-1)}{\big((a-1)\lambda+1\big)^2} + \frac{1-a}{(1-a)\lambda+a} - \frac{x(1-a)}{\big((1-a)\lambda+a\big)^2}. \tag{77}$$

We note in passing that an elementary calculation results in

$$d_{\mathbf{x}}(0) - d_{\mathbf{x}}(\tfrac{1}{2}) = \ln\left(\frac{4a}{(a+1)^2}\right) + \left(\frac{(a-1)^2}{a(a+1)}\right)x. \tag{78}$$

Now observe that $d'_{\mathbf{x}}(\frac{1}{2}) = 0$ and that $d'_{\mathbf{x}}(\lambda)$ in (77) is also a quotient of two polynomials (in λ), where the numerator is a polynomial of degree 3 or less. Thus, $d'_{\mathbf{x}}$ has at most two further roots different from $\frac{1}{2}$, which would have to be centered symmetrically around $\frac{1}{2}$ because of the symmetry of $d_{\mathbf{x}}$ about $\frac{1}{2}$. Furthermore, $d_{\mathbf{x}}(\lambda) \to +\infty$ as λ approaches either boundary point of $\operatorname{dom} d_{\mathbf{x}}$. Hence, critical points of $d_{\mathbf{x}}$ that are different from $\frac{1}{2}$ cannot be local maximizers. Therefore, $\overrightarrow{Q}_C(\mathbf{x}) \subseteq \{\mathbf{c}_0, \mathbf{c}_{1/2}, \mathbf{c}_1\}$. The symmetry of D and C yields that exactly one of the following holds: $\overrightarrow{Q}_C(\mathbf{x}) = \{\mathbf{c}_{1/2}\}$, $\overrightarrow{Q}_C(\mathbf{x}) = \{\mathbf{c}_0, \mathbf{c}_1\}$, or $\overrightarrow{Q}_C(\mathbf{x}) = \{\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_1\}$. Combining this with (78), we obtain the equivalences

$$\overrightarrow{Q}_C(\mathbf{x}) = \{\mathbf{c}_0, \mathbf{c}_1\} \quad \Leftrightarrow \quad d_{\mathbf{x}}(0) - d_{\mathbf{x}}(\tfrac{1}{2}) > 0 \quad \Leftrightarrow \quad x > g. \tag{79}$$

Let us now turn to the $\overrightarrow{D}$-Chebyshev center $\mathbf{z}$ of C. Since $\mathbf{z} \in \Delta$ (Proposition 5.1) and $\overrightarrow{Q}_C(\mathbf{z})$ must contain at least 2 points (Corollary 4.7), we write $\mathbf{z} = (z, z)$ and we deduce that either $\overrightarrow{Q}_C(\mathbf{z}) = \{\mathbf{c}_0, \mathbf{c}_1\}$ or $\overrightarrow{Q}_C(\mathbf{z}) = \{\mathbf{c}_0, \mathbf{c}_{1/2}, \mathbf{c}_1\}$. In turn, this means that exactly one of the following two cases holds.

(Case 1) $$\overrightarrow{Q}_C(\mathbf{z}) = \{\mathbf{c}_0, \mathbf{c}_1\},$$

or

$$\text{(Case 2)} \qquad \overrightarrow{Q}_C(\mathbf{z}) = \{\mathbf{c}_0, \mathbf{c}_{1/2}, \mathbf{c}_1\}.$$

If (Case 1) holds, then (58), Proposition 5.1, (74), and (79) yield that $\mathbf{z} = \mathbf{h}$ and that $z > g$. Thus,

$$\text{(Case 1)} \;\Rightarrow\; z = h > g. \tag{80}$$

Using (78), we obtain the implication

$$\text{(Case 2)} \;\Rightarrow\; z = g. \tag{81}$$

We now assume momentarily that $g < h$. Then, by (79), $\overrightarrow{Q}_C(\mathbf{h}) = \{\mathbf{c}_0, \mathbf{c}_1\}$ and hence $\mathbf{h} \in \nabla f^*(\operatorname{conv} \nabla \overrightarrow{Q}_C(\mathbf{h}))$ by (74). In view of the characterization (58) of $\mathbf{z}$, we obtain $\mathbf{z} = \mathbf{h}$ and hence $z = h$. We thus have verified the first case of (72).

Finally, we assume that $g \geq h$. In view of (80), (Case 1) cannot hold. Thus, (Case 2) must hold and (81) yields that $z = g$, i.e., that $\mathbf{z} = \mathbf{g}$. ■

The formula for $\mathbf{z}$ given in Example 5.5 immediately raises the question on how g and h relate to each other, viewed as functions of a. In the following result, we provide an alternative description of the inequality $g < h$.

Lemma 5.6 *Let the functions g and h be defined on the interval $I =]1, +\infty[$ by*

$$g(x) = \frac{x(x+1)}{(x-1)^2} \ln\left(\frac{(x+1)^2}{4x}\right) \quad \textit{and} \quad h(x) = \frac{2x}{x+1}. \tag{82}$$

Then there exists a real number $\widetilde{a} \in I$ such that

$$\big(\forall x \in I\big) \quad \begin{cases} g(x) < h(x), & \textit{if } x < \widetilde{a}; \\ g(x) = h(x), & \textit{if } x = \widetilde{a}; \\ g(x) > h(x), & \textit{if } x > \widetilde{a}. \end{cases} \tag{83}$$

In fact, $\widetilde{a} \approx 17.63$.

Proof. Observe that

$$\begin{aligned} h(x) > g(x) &\Leftrightarrow \frac{2(x-1)^2}{(x+1)^2} > \ln\left(\frac{(x+1)^2}{4x}\right) = 2\ln(x+1) - \ln(4x) \\ &\Leftrightarrow k(x) := \frac{2(x-1)^2}{(x+1)^2} - 2\ln(x+1) + \ln(4x) > 0. \end{aligned} \tag{84}$$

Since

$$\begin{aligned} k'(x) &= \frac{8(x-1)}{(x+1)^3} - \frac{2}{x+1} + \frac{1}{x} = \frac{-(x-1)(x^2-6x+1)}{x(x+1)^3} \\ &= \frac{-(x-1)\big(x-(3-2\sqrt{2})\big)\big(x-(3+2\sqrt{2})\big)}{x(x+1)^3}, \end{aligned} \tag{85}$$

we set $\xi = 3 + 2\sqrt{2} \approx 5.83$, and we deduce that k is strictly increasing on $]1, \xi]$ and that k is strictly decreasing on $[\xi, +\infty[$. On the other hand, $k(1) = 0$ and $\lim_{x\to+\infty} k(x) = -\infty$. Altogether, there must exist some number $\widetilde{a} > \xi$ such that $k > 0$ on $]1, \widetilde{a}[$, $k(\widetilde{a}) = 0$, and $k < 0$ on $]\widetilde{a}, +\infty[$. In view of (84), we obtain (83). Finally, the proclaimed approximation $\widetilde{a} \approx 17.63$ follows from Maple, Mathematica, or by simple bisection. ■

Remark 5.7 Consider again Example 5.5 and its notation. Define numbers $\mu_0, \mu_{1/2}, \mu_1$ according to the following two alternatives:

$$g < h \quad \Rightarrow \quad \begin{cases} \mu_0 = \mu_1 = \frac{1}{2}; \\ \\ \mu_{1/2} = 0, \end{cases} \tag{86}$$

or

$$g \geq h \quad \Rightarrow \quad \begin{cases} \mu_0 = \mu_1 = \dfrac{(a-1)^2 - 2a\ln\left(\dfrac{(a+1)^2}{4a}\right)}{(a-1)^2\ln\left(\dfrac{(a+1)^2}{4a}\right)}; \\ \\ \mu_{1/2} = \dfrac{-2(a-1)^2 + (a+1)^2\ln\left(\dfrac{(a+1)^2}{4a}\right)}{(a-1)^2\ln\left(\dfrac{(a+1)^2}{4a}\right)}. \end{cases} \tag{87}$$

One may verify that $\{\mu_0, \mu_{1/2}, \mu_1\} \subset [0, 1]$, that $\mu_0 + \mu_{1/2} + \mu_1 = 1$, and that

$$\mathbf{z} = \nabla f^*\Big(\mu_0 \nabla f(\mathbf{c}_0) + \mu_{1/2}\nabla f(\mathbf{c}_{1/2}) + \mu_1 \nabla f(\mathbf{c}_1)\Big). \tag{88}$$

Note that the existence of such convex coefficients is guaranteed by (58).

Remark 5.8 Figure 1 shows the set C, the Chebyshev center $\mathbf{z}$ of C, and the corresponding sphere of radius $\overrightarrow{F}_C(\mathbf{z})$ centered at $\mathbf{z}$, for a variety of values of a (fixed within each row) and for each of the three distances analyzed (fixed within each column). Specifically, shown are $a = 4$ and $a = 8$ over the region $R = [0, 10] \times [0, 10]$ (top two rows), and $a = 16$ and $a = 32$ over the region $R = [0, 50] \times [0, 50]$ (bottom two rows). Each are shown over color-maps indicating $\overrightarrow{F}_C(\mathbf{x})$ for each $\mathbf{x} \in R$, with the interpretation of the colors indicated in the accompanying color-legend. Note that the colors indicate distances from each point in the specified region to the farthest point in C, but are only relative comparisons within each graph; the same color in separate images does not indicate the same numerical magnitude, neither for a fixed distance D nor for a fixed value of a. In addition, the color-maps for the halved squared Euclidean distance and the Kullback-Leibler divergence were calculated using $\overrightarrow{Q}_C(\mathbf{x})$ in Examples 5.2 and 5.3, respectively. However, the color-map for the Itakura-Saito distance was calculated numerically by a discretization of C due to the absence of a corresponding formula for $\overrightarrow{Q}_C(\mathbf{x})$ in Example 5.5. We make the following observations directly from Figure 1:

(i) As predicted by our analysis, for the halved squared Euclidean distance, $\mathbf{z}$ falls on the point $\mathbf{c}_{1/2}$ for all values of a (left-column). The color-map corresponds to $\max\{D(\mathbf{x}, \mathbf{c}_0), D(\mathbf{x}, \mathbf{c}_1)\}$, with $D(\mathbf{x}, \mathbf{c}_0) = D(\mathbf{x}, \mathbf{c}_1)$ along Δ as per (60).

(ii) For the Itakura-Saito distance and for small a (see $a = 4$ and $a = 8$), the endpoints $\mathbf{c}_0$ and $\mathbf{c}_1$ are the farthest points from the Chebyshev center (h, h). When $a \geq \widetilde{a}$ (see Lemma 5.6), then the farthest points from (g, g) are $\{\mathbf{c}_0, \mathbf{c}_{1/2}, \mathbf{c}_1\}$, and $D((g, g), \mathbf{c}_{1/2}) < \overrightarrow{F}_C(h, h)$, visually confirming that (g, g) is now the Chebyshev center (see Figure 1 for $a = 32$).

Remark 5.9 Finally, let us fix $\mathbf{x} = (1, 1)$ and assume that $a = 6$. For the Itakura-Saito distance, we have that the farthest point $\overrightarrow{Q}_C(\mathbf{x})$ is $\mathbf{c}_{1/2}$, which is actually the *nearest* point of C to $\mathbf{x}$ for both of the other distances. Indeed, Figure 2 shows the spheres for the Itakura-Saito distance for a variety of radii. The thickness of the line segments is plotted proportional to the distance from $\mathbf{x}$. (In addition, note that the Itakura-Saito ball is convex for small a, a fact not apparent in Figure 1.)

Acknowledgments

The authors thank the anonymous referees 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.

References

[1] E. Asplund: "Sets with unique farthest points", *Israel Journal of Mathematics*, vol. 5, pp. 201–209, 1967.

[2] A. Banerjee, S. Merugu, I.S. Dhillon, and J. Ghosh: "Clustering with Bregman divergences", *Journal of Machine Learning Research*, vol. 6, pp. 1705–1749, 2005.

[3] H.H. Bauschke and J.M. Borwein: "Legendre functions and the method of random Bregman projections", *Journal of Convex Analysis*, vol. 4, pp 27–67, 1997.

[4] H.H. Bauschke and J.M. Borwein: "Joint and separate convexity of the Bregman distance" in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (Haifa 2000)*, D. Butnariu, Y. Censor, and S. Reich (editors), pp. 23–36, Elsevier, 2001.

[5] H.H. Bauschke, J.M. Borwein and P.L. Combettes: "Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces", *Communications in Contemporary Mathematics*, vol. 3, pp. 615–647, 2001.

[6] H.H. Bauschke and A.S. Lewis: "Dykstra's algorithm with Bregman projections: a convergence proof", *Optimization*, vol. 48, pp. 409–427, 2000.

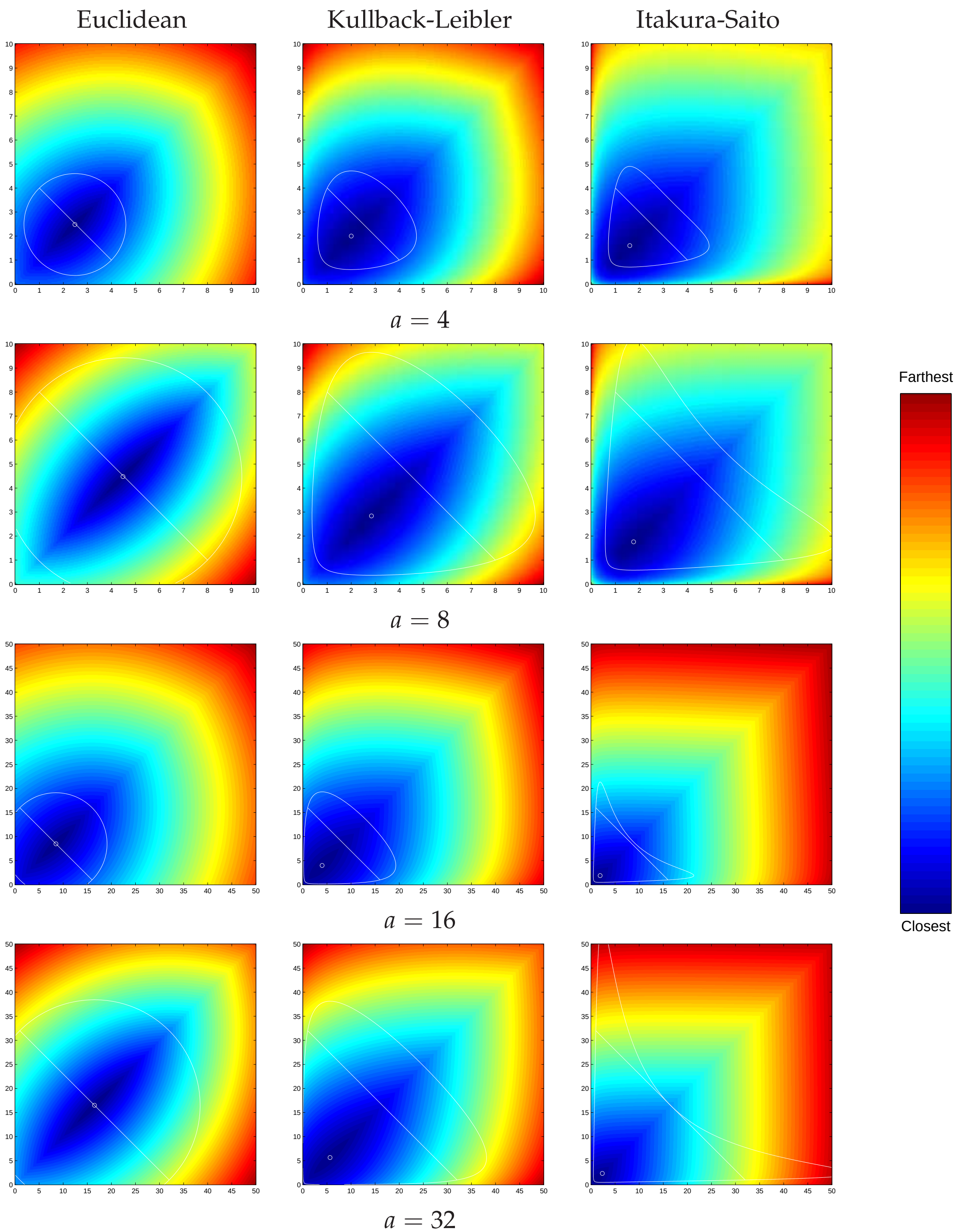

Figure 1: The set C, the Chebyshev center $\mathbf{z}$ of C, and the sphere of radius $\overrightarrow{F}_C(\mathbf{z})$ centered at $\mathbf{z}$, for C the line segment connecting $(1, a)$ and $(a, 1)$ for $a = 4$ and $a = 8$ over the region $[0, 10] \times [0, 10]$; and $a = 16$ and $a = 32$, over the region $[0, 50] \times [0, 50]$. Each are shown over color-maps for the three distances analyzed in Section 5, with the interpretations of the colors indicated in the color-legend.

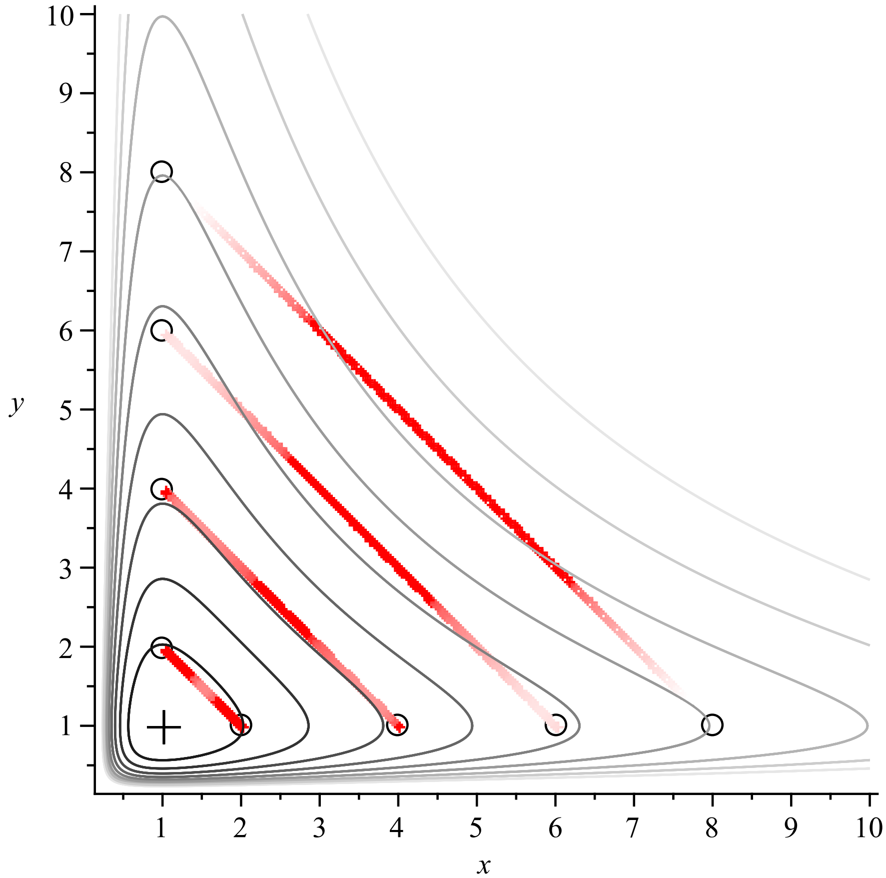

Figure 2: Spheres for the Ikaturo-Saito distance centered at $\mathbf{x} = (1, 1)$, for a variety of radii. Also shown are the line-segments C from $(1, a)$ to $(a, 1)$ for $a = 2, 4, 6, 8$, with plot intensity proportional to the distance from $\mathbf{x}$.

[7] H.H. Bauschke, X. Wang, J. Ye and X. Yuan: "Bregman distances and Chebyshev sets", *Journal of Approximation Theory*, vol. 159, pp. 3–25, 2009.

[8] H.H. Bauschke, X. Wang, J. Ye and X. Yuan: "Bregman distances and Klee sets", *Journal of Approximation Theory*, vol. 158, pp. 170–183, 2009.

[9] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, second edition, Springer-Verlag, 2006.

[10] J. Borwein and J. Vanderwerff: "Convex functions of Legendre type on Banach spaces", *Journal of Convex Analysis*, vol. 8, pp. 569–582, 2001.

[11] J. Borwein and J. Vanderwerff, *Convex Functions: Constructions, Characterizations & Counterexamples*, Cambridge University Press, to appear.

[12] 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. Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.

[13] D. Butnariu and A.N. Iusem, *Totally Convex Functions for Fixed Point Computation in Infinite Dimensional Optimization*, Kluwer, 2000.

[14] Y. Censor and S.A. Zenios, *Parallel Optimization*, Oxford University Press, 1997.

[15] G. Chen and M. Teboulle: "Convergence analysis of a proximal-like minimization algorithm using Bregman functions", *SIAM Journal on Optimization*, vol. 3, pp. 538–543, 1993.

[16] A.L. Garkavi: "On the Čebyšev center and convex hull of a set" (Russian), *Uspehi Matematičeskih Nauk*, vol. 19, pp. 139–145, 1964.

[17] J.-B. Hiriart-Urruty: "La conjecture des points les plus éloignés revisitée", *Annales des Sciences Mathématiques du Québec*, vol. 29, pp. 197–214, 2005.

[18] J.-B. Hiriart-Urruty: "Potpourri of conjectures and open questions in nonlinear analysis and optimization", *SIAM Review*, vol. 49, pp. 255–273, 2007.

[19] V. Klee: "Circumspheres and inner products", *Mathematica Scandinavica*, vol. 8, pp. 363–370, 1960.

[20] V. Klee: "Convexity of Chebyshev sets", *Mathematische Annalen*, vol. 142, pp. 292–304, 1960/61.

[21] T.S. Motzkin, E.G. Straus, and F.A. Valentine: "The number of farthest points", *Pacific Journal of Mathematics*, vol. 3, pp. 221–232, 1953.

[22] F. Nielsen and R. Nock: "On the smallest enclosing information disk", *Information Processing Letters*, vol. 105, pp. 93–97, 2008.

[23] R. Nock and F. Nielsen: "Fitting the smallest enclosing Bregman ball", in *Machine Learning: 16th European Conference on Machine Learning (Porto 2005)*, J. Gama, R. Camacho, P. Brazdil, A. Jorge and L. Torgo (editors), pp. 649–656, Springer Lecture Notes in Computer Science vol. 3720, 2005.

[24] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[25] R.T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer-Verlag, 1998.

[26] S. Simons, *From Hahn-Banach to Monotonicity*, Springer-Verlag, 2008.

[27] U. Westphal and T. Schwartz: "Farthest points and monotone operators", *Bulletin of the Australian Mathematical Society*, vol. 58, pp. 75–92, 1998.

[28] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, 2002.