Journal of Convex Analysis
Volume 4 (1997), No. 1, 27–67

Legendre Functions and the Method of Random Bregman Projections

Heinz H. Bauschke
*Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
e-mail: bauschke@cecm.sfu.ca*

Jonathan M. Borwein[1]
*Centre for Experimental & Constructive Mathematics,
Simon Fraser University, Burnaby, B.C., Canada V5A 1S6.
jborwein@cecm.sfu.ca*

Received March 13, 1995
Revised manuscript received November 22, 1995

The convex feasibility problem, that is, finding a point in the intersection of finitely many closed convex sets in Euclidean space, arises in various areas of mathematics and physical sciences. It can be solved by the classical method of cyclic orthogonal projections, where, by projecting cyclically onto the sets, a sequence is generated that converges to a point in the intersection. In 1967, Bregman extended this method to non-orthogonal projections based on a new notion of distance, nowadays called "Bregman distance". The Bregman distance is induced by a convex function. If this function is a so-called "zone consistent Bregman function", then Bregman's method works; however, deciding on this can be difficult. In this paper, Bregman's method is studied within the powerful framework of Convex Analysis. New insights are obtained and the rich class of "Bregman/Legendre functions" is introduced. Bregman's method still works, if the underlying function is Bregman/Legendre or more generally if it is Legendre but some constraint qualification holds additionally. The key advantage is the broad *applicability* and *verifiability* of these concepts. The results presented here are complementary to recent work by Censor and Reich on the method of random Bregman projections (where the sets are projected onto infinitely often – not necessarily cyclically). Special attention is given to examples, some of which connect to Pythagorean means and to Convex Analysis on the Hermitian or symmetric matrices.

Keywords: Bregman function, Bregman projection, convex feasibility problem, convex function, convex set, essentially smooth function, essentially strictly convex function, Hermitian matrix, Legendre function, projection.

1991 Mathematics Subject Classification: Primary 49M45; Secondary 47H17, 52A41, 65F10, 65K05, 90C25.

1. Introduction

Numerous problems in mathematics and physical sciences can be recast in terms of the famous *convex feasibility problem*:

$$\text{Given closed convex intersecting sets } C_1, \ldots, C_N, \text{ find a point in } C_1 \cap \cdots \cap C_N .$$

[1] Research supported by NSERC and by the Shrum Endowment.

ISSN 0944-6532 / $ 2.50

Typically, the points in the intersection are the sought-after *solutions* of a given problem and the sets $C_1, \dots, C_N$ correspond to some *constraints*. The convex feasibility problem arises in diverse areas such as best approximation theory, conformal mapping theory, image reconstruction, minimization of convex functions, and statistical estimation. Often, it is possible to calculate the orthogonal projection onto the constraints; thus, denoting the orthogonal projection onto the i^{th} constraint set by P_i, one can solve the convex feasibility problem by the classical *method of cyclic orthogonal projections*:

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

$$y_0 \xmapsto{P_1} y_1 \xmapsto{P_2} y_2 \cdots \xmapsto{P_N} y_N \xmapsto{P_1} y_{N+1} \xmapsto{P_2} \cdots .$$

The sequence (y_n) converges to a solution if the underlying space is some Euclidean space $\mathbb{R}^J$ as we will henceforth assume. (The interested reader is referred to [3] and the references therein for an attempt at review of the ever growing number of publications on projection algorithms.)

Bregman [4] generalized this method in 1967 by allowing (potentially) *non-orthogonal* projections which are constructed as follows:

Given a "sufficiently well-behaved" convex function f, consider the so-called *Bregman "distance"*

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

between two points x and y. (It is worth emphasizing that D_f is *not* a distance function in the sense of metric topology.) Distances between points induce distances between points and sets (as usual, by taking the appropriate infimum) and hence projections onto the sets C_i, denoted by $P_i^{(f)}$ and called *Bregman projections* onto C_i with respect to f.

Not surprisingly, the *method of cyclic Bregman projections* arises by simply replacing P_i by $P_i^{(f)}$ in the method of cyclic orthogonal projections. Now if f is what is called a "zone consistent Bregman function", then the method of cyclic Bregman projections indeed produces a sequence converging to a solution of the convex feasibility problem; see the work by Bregman [4], by Censor and Elfving [6], by Censor and Lent [7], and by De Pierro and Iusem [29].

Further progress was made within last year: Censor and Reich (Theorem 3.2 in [12]) established convergence for the method of *random* Bregman projections (that is, every constraint set is picked up infinitely often – no particular order or periodicity is required); Kiwiel [20] reported related results. Alber and Butnariu [1] investigated the method of cyclic Bregman projections in reflexive Banach spaces. It should be noted that Bregman functions appear also in recent papers on Proximal Point Methods; see, for instance, Burachik's [5], Chen and Teboulle's [13], Eckstein's [14], Iusem's [19], and Teboulle's [31].

Unfortunately, *verifying* that a function f is a Bregman function is not an easy task; moreover, we are not aware of any non-trivial condition sufficient for "zone consistency".

The objective in this paper is to analyze the method of random Bregman projections within the framework of Convex Analysis and to provide verifiable *conditions ensuring convergence of the method.*

The paper is organized as follows.

In Section 2, we recall and collect basic facts on well-known concepts of Convex Analysis: essential smoothness, essential strict convexity, Legendre functions, and coercivity.

The importance of these concepts becomes clear immediately in Section 3: Legendre functions are "zone consistent" (Theorem 3.14) – as already indicated, we are not aware of any other (non-trivial) sufficient condition.

Section 4 deals in detail with the class of Bregman functions. The results in the previous sections allow us to *remove* one (redundant) axiom and to simplify another (Remarks 4.2). Theorem 4.7 completely characterizes the important subclasses of "boundary coercive" and "zone coercive" Bregman functions in terms of conjugate function f^*.

In Section 5, we propose the notion of a "Bregman/Legendre function". The class of Bregman/Legendre functions lies strictly between the class of Legendre functions and the class of functions which are both Bregman and Legendre. Not surprisingly, we designed "Bregman/Legendreness" so that the method of random Bregman projections works. Just as for Bregman functions, checking for Bregman/Legendreness can be difficult; however, the situation on the real line is highly satisfactory: a Legendre function is Bregman/Legendre if and only if the domain of its conjugate is open (Theorem 5.8); moreover, this extends to all *separable* multi-dimensional Legendre functions. The class of Bregman/Legendre functions is closed under a variety of operations; hence, it is easy to construct new Bregman/Legendre functions. We are convinced that the notion of a Bregman/Legendre function will become useful in other contexts as well.

Many examples of Bregman/Legendre functions are presented in Section 6. By means of a two-dimensional example, we demonstrate that – unlike for the one-dimensional case – Bregman/Legendreness of a Legendre function really demands more than mere openness of the domain of the conjugate. The important question of computability of Bregman projections is addressed; we offer a new view of Bregman projections onto hyperplanes in Remark 6.13.

Section 7 starts with Pierra's product space formalization [28], which has become a standard tool in the field. Following Censor, Elfving, and Reich [6], [12], we discuss previous results in this light. It is perhaps surprising that some Pythagorean means can be viewed as Bregman projections onto the diagonal in this product space. The second half of this section connects – based on recent work by Lewis [22, 25, 24, 26] – to the increasingly popular area of Convex Analysis on the Hermitian or symmetric matrices. For example, Hadamard's inequality can be viewed in the context of Bregman distances as a "measure of non-diagonality". Moreover, the (negative) "Burg entropy", $\sum_j -\ln x_j$, corresponds to the logarithmic barrier function which lies at the heart of modern Linear Programming algorithms such as Interior Point Methods.

The last Section 8 contains our main result stating that the method of random Bregman functions works if f is Bregman/Legendre or if f is Legendre and the constraints and the interior of the domain of f have a point in common (Theorem 8.1). The important special case when each constraint set is a hyperplane is also investigated. These results are complementary to recent results by Censor and Reich [12] and partially generalize earlier work [4], [6], [7], [29]. Whereas Censor and Reich allow operators more general than Bregman projections, we build on the class of Bregman/Legendre functions: again, the important "Burg entropy" is included in our analysis but excluded from the class of Bregman functions.

Throughout the paper, we assume that

$$E \text{ is a Euclidean space } \mathbb{R}^J \text{ with inner product } \langle\cdot,\cdot\rangle \text{ and induced norm } \|\cdot\|.$$

Almost all the facts we use from Convex Analysis can be found in Rockafellar's fundamental book [30]. The notation is fairly standard: Given convex functions f and g on E, the *domain* of f (*conjugate function* of f, *recession function* of f, *gradient* of f, *subgradient* of f, *infimal convolution* of f and g, respectively) is denoted by $\operatorname{dom} f$ (f^*, $f0^+$, ∇f, ∂f, $f\Box g$, respectively). The indicator function of a set C is denoted ι_C and its *interior* (*boundary*, *closure*, respectively) is abbreviated by $\operatorname{int} C$ ($\operatorname{bd} C$, $\operatorname{cl} C$, respectively). Finally, I stands for the *identity mapping* or *identity matrix* and, for sequences, the symbol "$\longrightarrow$" indicates *convergence*.

2. Tools

The concepts in this section are fundamental to our analysis.

2.1. Essential smoothness

Definition 2.1. (Rockafellar's Section 26 in [30]) Suppose f is a closed convex proper function on E with $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Then f is *essentially smooth*, if f is differentiable on $\operatorname{int}(\operatorname{dom} f)$ and

$$\left.\begin{aligned} (x_n) \text{ in } \operatorname{int}(\operatorname{dom} f), \\ x_n \to x \in \operatorname{bd}(\operatorname{dom} f) \end{aligned}\right\} \Rightarrow \|\nabla f(x_n)\| \to +\infty.$$

Fact 2.2. (Rockafellar's Theorem 26.1 and Lemma 26.2 in [30]) *Suppose f is closed convex proper on E. Then the following are equivalent.*

(i) *f is essentially smooth.*
(ii) *f is differentiable on $\operatorname{int}(\operatorname{dom} f)$ and*

$$\lim_{t\downarrow 0}\langle \nabla f(x+t(y-x)), y-x\rangle = -\infty, \quad \forall x \in \operatorname{bd}(\operatorname{dom} f), \forall y \in \operatorname{int}(\operatorname{dom} f).$$

(iii) *$\partial f(x) = \emptyset$, $\forall x \in \operatorname{bd}(\operatorname{dom} f)$, and $\partial f(y) = \{\nabla f(y)\}$, $\forall y \in \operatorname{int}(\operatorname{dom} f)$.*

2.2. Essential strict convexity

Definition 2.3. (Rockafellar's Section 26 in [30]) Suppose f is closed convex proper on E. Then f is *essentially strictly convex*, if f is strictly convex on every convex subset of $\operatorname{dom} \partial f$.

Fact 2.4. (Rockafellar's Theorem 26.3 in [30]) *Suppose f is closed convex proper on E. Then the following are equivalent.*

(i) *f is essentially strictly convex.*
(ii) *f^* is essentially smooth.*
(iii) *$\partial f(x) \cap \partial f(y) = \emptyset$, $\quad \forall x, y \in \operatorname{dom} f, x \neq y$.*

We add another useful characterization which follows easily from Fact 2.2 and Fact 2.4.

Proposition 2.5. *Suppose f is closed convex proper on E. Then the following are equivalent.*

(i) *f is essentially strictly convex.*
(ii) $\operatorname{range} \partial f = \operatorname{int}(\operatorname{dom} f^*) = \operatorname{dom} \nabla f^*$.
(iii) $\{y\} = \partial f^*(\partial f(y)), \quad \forall y \in \operatorname{dom} \partial f$.

Corollary 2.6. *Suppose f is closed convex proper on E with* $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. *If f is essentially strictly convex and differentiable on* $\operatorname{int}(\operatorname{dom} f)$*, then*

$$\nabla f(y) \in \operatorname{int}(\operatorname{dom} f^*) \quad \textit{and} \quad \nabla f^* \nabla f(y) = y, \qquad \forall y \in \operatorname{int}(\operatorname{dom} f).$$

The following example shows that without essential strict convexity, the $\operatorname{range} \partial f$ need not be equal to $\operatorname{int}(\operatorname{dom} f^*)$:

Example 2.7. ("positive energy")
Let $f(x) = \frac{1}{2}|x|^2$, if $x \geq 0$; $+\infty$, otherwise on $E = \mathbb{R}$. Then $f^*(x^*) = \frac{1}{2}|x^*|^2$, if $x^* \geq 0$; 0, otherwise. Hence

$$\nabla f(\operatorname{int}(\operatorname{dom} f)) = I(]0, +\infty[) =]0, +\infty[\subsetneqq \mathbb{R} = \operatorname{int}(\operatorname{dom} f^*).$$

2.3. Legendre functions

Imposing essential smoothness and essential strict convexity together leads to Legendre functions, an extremely nice class of convex functions.

Definition 2.8. (Rockafellar's Section 26 in [30]) Suppose f is closed convex proper on E. Then f is *Legendre* (or a *Legendre function* or a *convex function of Legendre type*), if f is both essentially smooth and essentially strictly convex, i.e. f satisfies

L0. $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$.
L1. f is differentiable on $\operatorname{int}(\operatorname{dom} f)$.
L2. $\lim_{t \downarrow 0} \langle \nabla f(x + t(y - x)), y - x \rangle = -\infty, \quad \forall x \in \operatorname{bd}(\operatorname{dom} f),\ \forall y \in \operatorname{int}(\operatorname{dom} f)$.
L3. f is strictly convex on $\operatorname{int}(\operatorname{dom} f)$.

Fact 2.9. (Rockafellar's Theorem 26.5 in [30]) *A convex function f is of Legendre type, if and only if its conjugate f^* is. In this case, the gradient mapping*

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

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

In view of Example 2.7 and Fact 2.9, the positive energy is not Legendre. Many examples of Legendre functions will be provided later on.

2.4. Coercivities

Definition 2.10. A function f on E is *coercive* (or 0-*coercive*; see Definition X.1.3.7 in [17]) if f has bounded lower level sets: $\{x \in E : f(x) \leq r\}$ is bounded, $\forall r \in \mathbb{R}$; equivalently,

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

For instance, affine functions are not coercive, whereas the norm $\|\cdot\|$ is.

Fact 2.11. (Rockafellar's Corollary 14.2.2 in [30]) *Suppose f is closed convex proper on E and $x^* \in E$. Then $x^* \in \operatorname{int}(\operatorname{dom} f^*)$ if and only if the function $f(\cdot) - \langle x^*, \cdot\rangle$ is coercive.*

Remark 2.12. Fact 2.11 is a very powerful tool of convex analysis: existence of minimizers is often guaranteed by coercivity and closedness of the function; Fact 2.11 relates coercivity of this function to continuity of its conjugate function which is often much easier to check.

Before we discuss an even stronger version of coercivity, we review facts on the recession function (see Theorem 8.5 in [30] for the definition).

Fact 2.13. (Rockafellar's Corollary 13.3.4.(c) in [30]) *Suppose f is closed convex proper on E. Then*

$$\operatorname{int}(\operatorname{dom} f^*) = \{x^* \in E : \langle x^*, e\rangle < (f0^+)(e),\ \forall e \in E \setminus \{0\}\}.$$

Proposition 2.14. *Suppose f is closed convex proper on E. Suppose further (x_n) is a sequence in E with $\lim_n \|x_n\| = +\infty$ and $(x_n/\|x_n\|)$ convergent. Then $\underline{\lim}_n f(x_n)/\|x_n\| \geq (f0^+)(\lim_n x_n/\|x_n\|)$.*

Proof. Let $q := \lim_n x_n/\|x_n\| \neq 0$ and $L := \lim_n f(x_n)/\|x_n\| \in [-\infty, +\infty]$ (we assume without loss of generality that $(f(x_n)/\|x_n\|)$ converges — after passing to a subsequence if necessary). Fix any $\hat{x} \in \operatorname{dom} f$ and let $d_n := x_n - \hat{x}$, for all n. It is easy to check that $\lim_n \|x_n\|/\|d_n\| = 1$ and $\lim_n d_n/\|d_n\| = q$. Fix an arbitrary positive ρ. Then $\|d_n\| \geq \rho$ eventually; hence

$$L = \lim_n \frac{f(\hat{x} + \|d_n\| \frac{d_n}{\|d_n\|}) - f(\hat{x})}{\|d_n\|} \geq \lim_n \frac{f(\hat{x} + \rho \frac{d_n}{\|d_n\|}) - f(\hat{x})}{\rho} \geq \frac{f(\hat{x} + \rho q) - f(\hat{x})}{\rho}.$$

Now let tend ρ to $+\infty$ to conclude $L \geq (f0^+)(q)$. □

Definition 2.15. A function f on E is called *super-coercive* (or 1-*coercive*; see Definition X.1.3.7 in [17]), if

$$\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Evidently, any super-coercive function — for instance, $\frac{1}{2}\|\cdot\|^2$ — is coercive. The converse is not true: the norm $\|\cdot\|$ is only coercive. Super-coercivity has various characterizations.

Proposition 2.16. *Suppose f is closed convex proper on E. Then the following are equivalent:*

(i) *f is super-coercive.*

(ii) *f is* co-finite *(see Section 13 in [30]), i.e. $(f0^+)(e) = +\infty$, $\forall e \neq 0$; equivalently,*

$$\lim_{\lambda \to +\infty} \frac{f(\lambda e)}{\lambda} = +\infty, \quad \forall e \neq 0.$$

(iii) $\operatorname{dom} f^* = E$.
(iv) $f(\cdot) - \langle x^*, \cdot\rangle$ *is coercive*, $\forall x^* \in E$.

Proof. "(i)⇒(ii)": is trivial.
"(ii)⇐(i)": if not, then there is a sequence (x_n) such that $\|x_n\| \to +\infty$ but $(f(x_n)/\|x_n\|)$ is not tending to $+\infty$. Without loss (subsequence!) $+\infty > M \geq f(x_n)/\|x_n\|$ and $x_n/\|x_n\| \to e \neq 0$. Proposition 2.14 implies $M \geq (f0^+)(e)$, which contradicts the co-finiteness of f.
"(ii)⇔(iii)": is Rockafellar's Corollary 13.3.1 in [30].
"(iii)⇔(iv)": clear from Fact 2.11. □

3. Bregman distances and Bregman projections

3.1. Bregman distances

Definition 3.1. (see Bregman's Equation 1.4 in [4]) Suppose f is closed convex proper on E with $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. If f is differentiable on $\operatorname{int}(\operatorname{dom} f)$, then the corresponding *Bregman "distance"* D_f is defined by

$$D_f : E \times \operatorname{int}(\operatorname{dom} f) \to [0, +\infty] : (x, y) \mapsto f(x) - f(y) - \langle \nabla f(y), x - y\rangle.$$

Although commonly used, the term "distance" is misleading: neither is D_f symmetric (unless f is quadratic, as Iusem [18] proved) nor does D_f satisfy the triangle inequality (check $\|\cdot\|^2$). Nonetheless, D_f has some "good" distance-like features provided that f is "nice enough".

Proposition 3.2. *Suppose f is closed convex proper on E with* $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. *If f is differentiable on* $\operatorname{int}(\operatorname{dom} f)$, *then:*
(i) $D_f(x, y) = f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x\rangle$, $\forall x \in E$, $\forall y \in \operatorname{int}(\operatorname{dom} f)$.
(ii)

$$\left.\begin{aligned} (y_n) \text{ in } \operatorname{int}(\operatorname{dom} f), \\ y_n \to y \in \operatorname{int}(\operatorname{dom} f) \end{aligned}\right\} \Rightarrow\ D_f(y, y_n) \to 0$$

Proof. (i): By Theorem 23.5 in [30], $f(y) + f^*(\nabla f(y)) = \langle \nabla f(y), y\rangle$; now substitute.
(ii): This is clear from (i), Corollary 25.5.1 and Theorem 23.5 of [30]. □

A sharper form of Proposition 3.2.(ii) holds on the real line for essentially smooth functions.

Proposition 3.3. *Suppose f is essentially smooth on* $\mathbb{R}$. *Then*

$$\left.\begin{aligned} (y_n) \text{ in } \operatorname{int}(\operatorname{dom} f), \\ y_n \to y \in \operatorname{dom} f \end{aligned}\right\} \Rightarrow\ D_f(y, y_n) \to 0.$$

Proof. By Proposition 3.2.(ii), without loss $y \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$. We can assume there exists some $\epsilon > 0$ such that $[y, y + 2\epsilon] \subseteq \operatorname{dom} f$ so that $y + \epsilon \in \operatorname{int}(\operatorname{dom} f)$ (the case when y is a right endpoint is treated similarly). By essential smoothness,

$$\langle \nabla f(y_n), \epsilon\rangle = \langle \nabla f(y + (\tfrac{y_n - y}{\epsilon})[(y + \epsilon) - y]), [(y + \epsilon) - y]\rangle \to -\infty.$$

In particular, $-\nabla f(y_n)(y - y_n) < 0$ and hence

$$0 \leq D_f(y, y_n) < f(y) - f(y_n), \qquad \text{for all large } n.$$

Now apply Corollary 7.5.1 of [30]. □

Remark 3.4. Of course, Proposition 3.3 extends to all *separable* essentially smooth functions. On the other hand, the proposition does not necessarily hold for non-separable essentially smooth functions; see Example 7.32.

The next proposition is good for building examples. We omit its simple proof.

Proposition 3.5. *Suppose $f_1, \ldots, f_J$ are closed convex proper on some Euclidean spaces $E_1, \ldots, E_J$ with $\operatorname{int}(\operatorname{dom} f_1), \ldots, \operatorname{int}(\operatorname{dom} f_J) \neq \emptyset$, respectively. If every f_j is differentiable on $\operatorname{int}(\operatorname{dom} f_j)$ and if $\lambda_1, \ldots, \lambda_J$ are strictly positive real numbers, then*

$$\mathbf{f} \;:\; \mathbf{E} := \textstyle\prod_j E_j \to \,]-\infty, +\infty] \;:\; \mathbf{x} := (x_1, \ldots, x_J) \mapsto \displaystyle\sum_j \lambda_j f_j(x_j)$$

is a closed convex function on $\mathbf{E}$ that is differentiable on $\operatorname{int}(\operatorname{dom} \mathbf{f}) = \prod_j \operatorname{int}(\operatorname{dom} f_j)$. The corresponding Bregman distance of $\mathbf{f}$ is

$$D_{\mathbf{f}}(\mathbf{x}, \mathbf{y}) = \sum_j \lambda_j D_{f_j}(x_j, y_j), \qquad \forall \mathbf{x} \in \mathbf{E},\ \forall \mathbf{y} \in \operatorname{int}(\operatorname{dom} \mathbf{f}).$$

The following proposition will be useful later.

Proposition 3.6. *Suppose f is Legendre on $\mathbb{R}$ and $\operatorname{dom} f = [a, b[$, where $-\infty < a < b \leq +\infty$. If (y_n) is a sequence in $\operatorname{int}(\operatorname{dom} f)$ with $y_n \to b$ and $\nabla f(y_n) \to +\infty$, then $D_f(a, y_n) \to +\infty$.*

Proof. *Case 1: $b < +\infty$.*
Fix any $y \in \,]a, b[$. Then eventually $y_n > y$ and thus $\nabla f(y_n) \geq (f(y_n) - f(y))/(y_n - y)$. Hence

$$\begin{aligned} D_f(a, y_n) &= f(a) - f(y_n) + \nabla f(y_n)(y_n - a) \\ &\geq f(a) - f(y_n) + \frac{f(y_n) - f(y)}{y_n - y}(y_n - a) \\ &= \left\{ f(a) - f(y)\frac{y_n - a}{y_n - y} \right\} + f(y_n) \left\{ \frac{y - a}{y_n - y} \right\}. \end{aligned}$$

The bracketed terms are bounded; moreover, the second bracketed term converges to $(y-a)/(b-y) > 0$. On the other hand, $f(y_n) \to +\infty$, because f is closed and $b \notin \operatorname{dom} f$. The result follows.

Case 2: $b = +\infty$.
Suppose to the contrary $(D_f(a, y_n))$ is not tending to $+\infty$. After passing to a subsequence, we assume $(D_f(a, y_n))$ is bounded. Dividing by $y_n - a$ yields

$$\frac{f(y_n)}{y_n - a} - \nabla f(y_n) \to 0; \quad \text{hence} \quad \frac{f(y_n)}{y_n - a} \to +\infty.$$

Fix any $y > a$. Then eventually $y_n > y$ and as before

$$\begin{aligned} D_f(a, y_n) &= f(a) - f(y_n) + \nabla f(y_n)(y_n - a) \\ &\geq f(a) - f(y_n) + \frac{f(y_n) - f(y)}{y_n - y}(y_n - a) \\ &= \left\{ f(a) - f(y)\frac{y_n - a}{y_n - y} \right\} + \frac{f(y_n)}{y_n - a} \left\{ \frac{y_n - a}{y_n - y}(y - a) \right\}. \end{aligned}$$

Again the bracketed terms are bounded and the second bracketed term converges to $y - a > 0$, which implies $D_f(a, y_n) \to +\infty$. □

Theorem 3.7. ("essential strict convexity helps") *Suppose f is closed convex proper on E, differentiable on $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$, and essentially strictly convex. Suppose further $y \in \operatorname{int}(\operatorname{dom} f)$. Then:*

(i) *$D_f(\cdot, y)$ is closed convex proper on E, differentiable on $\operatorname{int}(\operatorname{dom} f)$, and essentially strictly convex.*
(ii) *$\nabla f(y) \in \operatorname{int}(\operatorname{dom} f^*)$.*
(iii) *$D_f(\cdot, y)$ is coercive.*
(iv) *$D_f(x, y) = 0 \iff x = y$, $\forall x \in E$.*
(v) *$D_f(x, y) = D_{f^*}(\nabla f(y), \nabla f(x))$, $\forall x \in \operatorname{int}(\operatorname{dom} f)$.*
(vi) *If $\operatorname{dom} f^*$ is open, then $D_f(x, \cdot)$ is coercive, $\forall x \in \operatorname{int}(\operatorname{dom} f)$; equivalently,*

$$\left. \begin{array}{r} x \in \operatorname{int}(\operatorname{dom} f), (y_n) \textit{ in } \operatorname{int}(\operatorname{dom} f), \\ (D_f(x, y_n)) \textit{ bounded} \end{array} \right\} \Rightarrow \ (y_n) \textit{ bounded.}$$

Proof. (i): is clear by Proposition 3.2.(i).
(ii),(iii): f essentially strictly convex $\Rightarrow \nabla f(y) \in \operatorname{int}(\operatorname{dom} f^*)$ (Proposition 2.5) $\Leftrightarrow f^*$ is continuous at $\nabla f(y) \Leftrightarrow f(\cdot) - \langle \nabla f(y), \cdot \rangle$ is coercive (Fact 2.11) $\Leftrightarrow D_f(\cdot, y)$ is coercive (Proposition 3.2.(i)).
(iv): Fix $x \in E$. Then $D_f(x, y) = 0 \Leftrightarrow f(x) + f^*(\nabla f(y)) = \langle \nabla f(y), x \rangle \Leftrightarrow x = \nabla f^*(\nabla f(y))$ $\Leftrightarrow x = y$ (Corollary 2.6).
(v): $D_{f^*}(\nabla f(y), \nabla f(x)) = f^*(\nabla f(y)) + f^{**}(\nabla f^*(\nabla f(x))) - \langle \nabla f^*(\nabla f(x)), \nabla f(y) \rangle = f^*(\nabla f(y)) + f(x) - \langle x, \nabla f(y) \rangle = D_f(x, y)$.
(vi): Suppose $x \in \operatorname{int}(\operatorname{dom} f)$, (y_n) in $\operatorname{int}(\operatorname{dom} f)$ with $(D_f(x, y_n)) = (f(x) + f^*(\nabla f(y_n)) - \langle x, \nabla f(y_n) \rangle)$ bounded. Since $x \in \operatorname{int}(\operatorname{dom} f)$, the function $f^*(\cdot) - \langle x, \cdot \rangle$ is coercive. Hence $(\nabla f(y_n))$ is bounded. f^* is closed, thus all cluster points of $(\nabla f(y_n))$ lie in $\operatorname{dom} f^* = \operatorname{int}(\operatorname{dom} f^*)$. It follows that $(y_n) = (\nabla f^*(\nabla f(y_n)))$ is bounded, too (Corollary 2.6 and Corollary 25.5.1 of [30]. □

Theorem 3.8. ("essential smoothness helps") *Suppose f is closed convex proper on E, differentiable on $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$, and essentially smooth. Then:*

(i) $$\left. \begin{array}{r} x \in \operatorname{int}(\operatorname{dom} f), (y_n) \textit{ in } \operatorname{int}(\operatorname{dom} f), \\ y_n \to y \in \operatorname{bd}(\operatorname{dom} f) \end{array} \right\} \Rightarrow \ D_f(x, y_n) \to +\infty.$$

$$\text{(ii)}\quad \left.\begin{array}{r} x \in \operatorname{int}(\operatorname{dom} f), (y_n) \textit{ in } \operatorname{int}(\operatorname{dom} f), \\ y_n \to y \in \operatorname{cl}(\operatorname{dom} f), D_f(x, y_n) \textit{ bounded} \end{array}\right\} \Rightarrow \begin{array}{r} y \in \operatorname{int}(\operatorname{dom} f) \\ \textit{(and } D_f(y, y_n) \to 0\textit{).} \end{array}$$

$$\text{(iii)}\quad \left.\begin{array}{r} (x_n) \textit{ in } \operatorname{dom} f,\ x_n \to x \in \operatorname{dom} f, \\ (y_n) \textit{ in } \operatorname{int}(\operatorname{dom} f), y_n \to y \in \operatorname{dom} f, \\ \{x, y\} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset, D_f(x_n, y_n) \to 0 \end{array}\right\} \Rightarrow \begin{array}{c} D_f(x, y) = 0 \\ \textit{(and } y \in \operatorname{int}(\operatorname{dom} f)\textit{).} \end{array}$$

Proof. (i): Assume to the contrary that $\underline{\lim}_n D_f(x, y_n) < +\infty$. Without loss (subsequences!) , we assume $\|\nabla f(y_n)\| \to +\infty$ (by essential smoothness), $\nabla f(y_n)/\|\nabla f(y_n)\| \to q \neq 0$, and $(D_f(x, y_n)) = (f(x) + f^*(\nabla f(y_n)) - \langle \nabla f(y_n), x\rangle)$ is bounded. Dividing the last sequence by $\|\nabla f(y_n)\|$ yields $f^*(\nabla f(y_n))/\|\nabla f(y_n)\| \to \langle q, x\rangle$. By Proposition 2.14, $\langle q, x\rangle \geq (f^* 0^+)(q)$; thus (Fact 2.13) $x \notin \operatorname{int}(\operatorname{dom} f)$ which is absurd.
(ii): is equivalent to (i). The "(and $D_f(y, y_n) \to 0$)" part follows from Proposition 3.2.(ii).
(iii): *Claim:* $y \in \operatorname{int}(\operatorname{dom} f)$.
Otherwise, $y \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$. Then $x \in \operatorname{int}(\operatorname{dom} f)$ and $\|\nabla f(y_n)\| \to +\infty$. We assume (subsequence!) that $(\nabla f(y_n)/\|\nabla f(y_n)\|)$ is convergent, say to $q \neq 0$. Now $0 \leftarrow D_f(x_n, y_n) = f(x_n) + f^*(\nabla f(y_n)) - \langle \nabla f(y_n), x_n\rangle$; thus division by $\|\nabla f(y_n)\|$ yields $\lim_n \frac{f^*(\nabla f(y_n))}{\|\nabla f(y_n)\|} = \langle q, x\rangle$. Proposition 2.14 implies $\langle q, x\rangle \geq (f^* 0^+)(q)$. Hence, by Fact 2.13, $x \notin \operatorname{int}(\operatorname{dom} f^{**}) = \operatorname{int}(\operatorname{dom} f)$. This is the desired contradiction and the claim thus holds.
Because f is closed, we get $0 = \lim_n D_f(x_n, y_n) = \lim_n f(x_n) + f^*(\nabla f(y)) - \langle \nabla f(y), x\rangle \geq f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x\rangle = D_f(x, y) \geq 0$. $\square$

3.2. Legendreness

Theorem 3.9. ("Legendreness helps a lot") *Suppose f is Legendre on E. Then:*
(i) $D_{f^*}(x^*, y^*) = D_f(\nabla f^*(y^*), \nabla f^*(x^*))$, $\forall x^*, \forall y^* \in \operatorname{int}(\operatorname{dom} f^*)$.
(ii) *If* $\operatorname{dom} f^*$ *is not open, then* $D_f(x, \cdot)$ *is not coercive,* $\forall x \in \operatorname{dom} f$.

$$\text{(iii)}\quad \left.\begin{array}{r} (x_n) \textit{ in } \operatorname{dom} f, x_n \to x \in \operatorname{dom} f, \\ (y_n) \textit{ in } \operatorname{int}(\operatorname{dom} f), y_n \to y \in \operatorname{dom} f, \\ \{x, y\} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset, D_f(x_n, y_n) \to 0 \end{array}\right\} \Rightarrow x = y.$$

Proof. (i): f^* is essentially smooth and essentially strictly convex, since f is Legendre. Hence Theorem 3.7.(v) applies.
(ii): Fix $y^* \in \operatorname{dom} f^* \setminus \operatorname{int}(\operatorname{dom} f^*)$, $y_0^* \in \operatorname{int}(\operatorname{dom} f^*)$, and let $y_n^* := (1 - 1/n)y^* + (1/n)y_0^*$, for all $n \geq 2$. Then the sequence (y_n^*) lies in $\operatorname{int}(\operatorname{dom} f^*)$ and converges to y^* along the segment between y_0^* and y. Hence $f^*(y_n^*) \to f^*(y^*)$ (Corollary 7.5.1 in [30]). Let $y_n := \nabla f^*(y_n^*)$, for all n; then $\|y_n\| \to +\infty$. Now fix an arbitrary $x \in \operatorname{dom} f$. The sequences $(\langle y_n^*, x\rangle)$, $(f^*(y_n^*))$ are (convergent hence) bounded and so is $(D_f(x, y_n)) = (f(x) + f^*(y_n^*) - \langle y_n^*, x\rangle)$. Therefore, $D_f(x, \cdot)$ is not coercive.
(iii): combine Theorem 3.8.(iii) with Theorem 3.7.(iv). $\square$

Example 3.10. ("Boltzmann/Shannon") Suppose $f(x) = x \ln x - x$ on $\operatorname{dom} f = [0, +\infty[$. Then f is Legendre and $f^* = \exp$ has open domain. Fix any $x > 0$ and

$y_n \to +\infty$. Suppose further that $x_n \downarrow x$. A direct check (see also Iusem's Proposition 9.1 in [19] or Chen and Teboulle's Lemma 3.1 in [13]) gives

$$D_f(x_n, y_n) = D_f(x, y_n) - D_f(x, x_n) - \nabla f(x_n)(x - x_n) + \nabla f(y_n)(x - x_n).$$

The first term, $D_f(x, y_n)$, tends to $+\infty$ by Theorem 3.7 (vi). The second term, $-D_f(x, x_n)$, tends to 0 by Proposition 3.2.(ii). The third term, $-\nabla f(x_n)(x - x_n)$, tends also to 0. Now $\nabla f(y_n) = \ln y_n \to +\infty$ and $x - x_n \uparrow 0$. Hence we can adjust (x_n) *a posteriori* so that $D_f(x, y_n) + \nabla f(y_n)(x - x_n) \to 0$. Then altogether

$$x_n \to x \in \operatorname{int}(\operatorname{dom} f) \quad \text{and} \quad D_f(x_n, y_n) \to 0, \quad \text{but} \quad \|y_n\| \to +\infty.$$

Hence the assumption on convergence of (y_n) in the hypothesis of Theorem 3.9.(iii) is important.

Theorem 3.9 has an intriguing consequence.

Corollary 3.11. *Suppose f is Legendre on E. Then the following implications hold:*

$$\begin{aligned} & D_f(x, \cdot) \textit{ is coercive, for some } x \in \operatorname{dom} f. \\ \Rightarrow & \operatorname{dom} f^* \textit{ is open.} \\ \Rightarrow & D_f(x, \cdot) \textit{ is coercive, for all } x \in \operatorname{int}(\operatorname{dom} f). \end{aligned}$$

Consequently, the following are equivalent:
(i) $D_f(x, \cdot)$ *is coercive, for some* $x \in \operatorname{int}(\operatorname{dom} f)$.
(ii) $\operatorname{dom} f^*$ *is open.*
(iii) $D_f(x, \cdot)$ *is coercive, for all* $x \in \operatorname{int}(\operatorname{dom} f)$.

Proof. The first implication is Theorem 3.9.(ii), the second is Theorem 3.7.(vi); the "Consequently" part follows. □

3.3. Bregman projections

Having studied Bregman distances in some detail, we now turn to the associated Bregman projections. These are, of course, the key players in the methods investigated later.
Fix a closed convex proper function f that is differentiable on $\operatorname{int}(\operatorname{dom} f)$ and a set C with $\operatorname{int}(\operatorname{dom} f) \cap C \neq \emptyset$ (the usual constraint qualification). Pick $y \in \operatorname{int}(\operatorname{dom} f)$. We wish to define a "projection of y onto C w.r.t. f", denoted $P_C y$ or P_C^f, by

$$P_C y = \operatorname*{argmin}_{x \in C \cap \operatorname{dom} f} D_f(x, y).$$

To really speak of a projection, we must require

- existence: the argmin should be nonempty; and
- uniqueness: the argmin should be singleton.

Loosely speaking, this is guaranteed by essential strict convexity.
In addition, note that y has to lie in $\operatorname{int}(\operatorname{dom} f)$ to make even sense of the argmin. Moreover, to be able to project the point $P_C y$ again (onto another constraint set perhaps), we have to impose

- interiority: the argmin should lie in $\operatorname{int}(\operatorname{dom} f)$.

(This shows *a posteriori* that the constraint qualification "$\operatorname{int}(\operatorname{dom} f) \cap C \neq \emptyset$" is *necessary.*) The interiority condition appears in the literature (for instance, [6]) under the name *"zone consistency"*. Surprisingly, in all the papers we are aware of, we have nowhere found a non-trivial sufficient condition for interiority/zone consistency. Fortunately, there is one very natural property guaranteeing precisely this: essential smoothness.
Altogether, Legendreness is the most natural property guaranteeing "good" Bregman projections.

Theorem 3.12. Suppose f is closed convex proper on E and differentiable on $\operatorname{int}(\operatorname{dom} f)$. Suppose further C is closed convex with $C \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and y is an arbitrary point in $\operatorname{int}(\operatorname{dom} f)$. Then:

(i) If f is essentially smooth, then $\operatorname{arginf}_{x \in C \cap \operatorname{dom} f} D_f(x, y)$ is nonempty.

(ii) If f is strictly convex on $\operatorname{dom} f$, then $\operatorname{arginf}_{x \in C \cap \operatorname{dom} f} D_f(x, y)$ is at most singleton.

(iii) If f is Legendre, then $\operatorname{arginf}_{x \in C \cap \operatorname{dom} f} D_f(x, y)$ is singleton and contained in $\operatorname{int}(\operatorname{dom} f)$.

Proof. (i): On the one hand, $D_f(\cdot, y)$ is closed and coercive (Theorem 3.7). On the other hand, $C \cap \operatorname{cl}(\operatorname{dom} f)$ is closed. Altogether, $\operatorname{arginf}_{x \in C \cap \operatorname{cl}(\operatorname{dom} f)} D_f(x, y)$ is nonempty; of course, this arginf is equal to $\operatorname{arginf}_{x \in C \cap \operatorname{dom} f} D_f(x, y)$.
(ii): Since f is strictly convex on $\operatorname{dom} f$, so is $D_f(\cdot, y)$.
(iii): By (i), $\operatorname{arginf}_{x \in C \cap \operatorname{dom} f} D_f(x, y)$ is nonempty.

Claim: $\operatorname{argmin}_{x \in C \cap \operatorname{dom} f} D_f(x, y) \subseteq \operatorname{int}(\operatorname{dom} f)$.

Assume to the contrary $\bar{x} \in \operatorname{argmin}_{x \in C \cap \operatorname{dom} f} D_f(x, y) \cap (\operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f))$. Fix any $z \in C \cap \operatorname{int}(\operatorname{dom} f)$ and define a closed convex proper function Φ by

$$\Phi: \ [0, 1] \to [0, +\infty[\,: t \mapsto D_f((1-t)\bar{x} + tz, y).$$

Then $\Phi'(t) = \langle \nabla f(\bar{x} + t(z - \bar{x})), z - \bar{x} \rangle - \langle \nabla f(y), z - \bar{x} \rangle$, for all $t \in]0, 1[$. By essential smoothness, $\lim_{t \downarrow 0} \Phi'(t) = -\infty$. This implies $\Phi(t) < \Phi(0)$, for all small positive t. However, $(1-t)\bar{x} + tz \in C \cap \operatorname{int}(\operatorname{dom} f)$ for these small t; hence we have contradicted the choice of $\bar{x}$. The claim is verified.

Finally, since $D_f(\cdot, y)$ is essentially strictly convex (Theorem 3.7.(i)) and thus strictly convex on $\operatorname{int}(\operatorname{dom} f)$, we conclude that the argmin is singleton. $\square$

Definition 3.13. (see also Censor and Lent's [7], [6]) Suppose f is closed convex proper on E and differentiable on $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. We say that f is *zone consistent*, if for every closed convex set C with $C \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and every $y \in \operatorname{int}(\operatorname{dom} f)$, the arginf

$$\operatorname*{arginf}_{x \in C \cap \operatorname{dom} f} D_f(x, y)$$

is singleton and contained in $\operatorname{int}(\operatorname{dom} f)$; that is, f is zone consistent with respect to every C. We denote this point by $P_C y$ or $P_C^f y$ and call the mapping

$$P_C : \operatorname{int}(\operatorname{dom} f) \to C \cap \operatorname{int}(\operatorname{dom} f) : y \mapsto P_C y$$

the *Bregman projection w.r.t. f*.

Theorem 3.12.(iii) now becomes:

Theorem 3.14. *Every Legendre function is zone consistent.*

Example 3.15. ("strict convexity alone is not enough") Consider the "positive energy" $f(x) := \frac{1}{2}\|x\|^2$ on $\operatorname{dom} f := \{x \in \mathbb{R}^2 : x \geq 0\}$ and $C := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 1\}$. Then $C \cap \operatorname{int}(\operatorname{dom} f) \ni (\frac{1}{2}, \frac{1}{2})$ is nonempty. Let further $y := (2, 1) \in \operatorname{int}(\operatorname{dom} f)$. One easily checks that

$$P_C^f(y) = (1, 0) \notin \operatorname{int}(\operatorname{dom} f).$$

Hence f is not zone consistent.

The following proposition is useful for calculating Bregman projections.

Proposition 3.16. *Suppose f is Legendre on E and C is closed convex with $C \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Suppose further $y \in \operatorname{int}(\operatorname{dom} f)$. Then the Bregman projection $P_C y$ is characterized by*

$$P_C y \in C \cap \operatorname{int}(\operatorname{dom} f) \quad \text{and} \quad \langle \nabla f(y) - \nabla f(P_C y), C - P_C y \rangle \leq 0.$$

In addition,

$$D_f(P_C y, y) \leq D_f(c, y) - D_f(c, P_C y), \qquad \text{for all } c \in C \cap \operatorname{dom} f.$$

Proof. Convex calculus time!

$$\begin{aligned}
& \bar{x} = P_C y \\
\overset{\text{Theorem 3.12.(iii)}}{\Leftrightarrow} \quad & \bar{x} = \operatorname*{argmin}_{x \in C \cap \operatorname{int}(\operatorname{dom} f)} D_f(x, y) \\
\Leftrightarrow \quad & \bar{x} = \operatorname*{argmin}_{x \in C \cap \operatorname{int}(\operatorname{dom} f)} f(x) - \langle \nabla f(y), x \rangle \\
\Leftrightarrow \quad & \bar{x} = \operatorname*{argmin}_{x \in E} f(x) + \langle -\nabla f(y), x \rangle + \iota_C(x) + \iota_{\operatorname{int}(\operatorname{dom} f)}(x) \\
\Leftrightarrow \quad & 0 \in \partial \big(f(\cdot) + \langle -\nabla f(y), \cdot \rangle + \iota_C(\cdot) + \iota_{\operatorname{int}(\operatorname{dom} f)}(\cdot) \big) (\bar{x}) \\
\overset{\text{[30,Theorem 23.8]}}{\Leftrightarrow} \quad & 0 \in \nabla f(\bar{x}) - \nabla f(y) + \partial \iota_C(\bar{x}) \text{ and } \bar{x} \in \operatorname{int}(\operatorname{dom} f),
\end{aligned}$$

which gives the desired characterization. The "In addition" part is a trivial expansion. □

Remark 3.17. If $f = \frac{1}{2}\|\cdot\|^2$, then the characterization of $P_C y$ becomes the well-known characterization of orthogonal projections:

$$P_C y \in C \quad \text{and} \quad \langle C - P_C y, y - P_C y \rangle \leq 0.$$

In this section, it has become obvious that essential smoothness or essential strict convexity guarantees many desirable properties of Bregman distances. Most importantly, Legendreness gives rise to well-defined and well-behaved Bregman projections.

4. Bregman functions

Bregman functions were introduced and utilized by Censor and Lent in [7]. The notion rests on Bregman's fundamental work [4] from 1967.

Definition 4.1. Suppose f is closed convex proper on E. Then f is *Bregman* (or a *Bregman function*), if the following properties (B0)–(B5) hold:

(B0) $\operatorname{dom} f$ is closed and $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$.
(B1) f is continuously differentiable on $\operatorname{int}(\operatorname{dom} f)$.
(B2) f is strictly convex and continuous on $\operatorname{dom} f$.
(B3) (i) $D_f(\cdot, y)$ is coercive, $\forall y \in \operatorname{int}(\operatorname{dom} f)$.
(ii) $D_f(x, \cdot)$ is coercive, $\forall x \in \operatorname{dom} f$.
(B4) (y_n) in $\operatorname{int}(\operatorname{dom} f)$, $y_n \to y \quad \Rightarrow \quad D_f(y, y_n) \to 0$.

$$\text{(B5)} \qquad \left.\begin{array}{r} (x_n) \text{ in } \operatorname{dom} f, (x_n) \text{ bounded}, \\ (y_n) \text{ in } \operatorname{int}(\operatorname{dom} f), y_n \to y, \\ D_f(x_n, y_n) \to 0 \end{array}\right\} \Rightarrow x_n \to y.$$

Remarks 4.2.

- (B0) is quite restrictive: "Burg's entropy", $-\ln x$, is automatically excluded since its domain is not closed. However, this function is known to be an extremely well-behaved convex function.
- In (B1), it suffices to require differentiability of f throughout $\operatorname{int}(\operatorname{dom} f)$: if f is differentiable throughout $\operatorname{int}(\operatorname{dom} f)$, then it is actually continuously differentiable (see Rockafellar's Corollary 25.5.1 in [30]).
- (B2) implies essential strict convexity of f.
- (B3)(i) is *redundant*: indeed, by (B2), f is essentially strictly convex and (B3)(i) follows from Theorem 3.7.(iii).
- If f is also essentially smooth (and hence Legendre), then (B3)(ii) simplifies (via Corollary 3.11) as follows:
  - •• $\operatorname{dom} f$ open, i.e. $\operatorname{dom} f = E$: (B3).(ii) $\Leftrightarrow$ $\operatorname{dom} f^*$ open.
  - •• $\operatorname{dom} f$ not open: (B3).(ii) $\Leftrightarrow$ $D_f(x, \cdot)$ is coercive, $\forall x \in \operatorname{bd}(\operatorname{dom} f)$.

Remark 4.3. One of the most important requirements of the function generating the Bregman distance is interiority/zone consistency. However: although the "positive energy" $f(x) = \frac{1}{2}\|x\|^2$ on $\operatorname{dom} f := \{x \in E : x_j \geq 0, \ \forall j\}$ in two (or more) dimensions is a Bregman function, *it is not zone consistent* (see Example 3.15). This is a serious shortcoming of Bregman functions. On the other hand, we have seen that Legendre functions guarantee zone consistency automatically. The path we follow is now obvious: in the next section, we introduce "Bregman/Legendre" functions which combine the best of both worlds: they are Legendre functions with "a little more"; this "little extra" is just enough to make the convergence analysis of the methods studied later work. Even better: in case of separable functions, it turns out extremely easy to *verify* "Bregman/Legendreness" and Burg's entropy "belongs to the club of Bregman/Legendre functions". The positive energy is *not* in this class; nonetheless, this is reasonable since this function is not zone consistent anyway.

In [19], Iusem discusses two additional useful properties of Bregman functions:

Definition 4.4. (Iusem [19]) Suppose f is Bregman on E.

(i) f is called *boundary coercive*, if

$$\left.\begin{array}{r}(x_n) \text{ in } \operatorname{int}(\operatorname{dom} f), y \in \operatorname{int}(\operatorname{dom} f),\\ x_n \to x \in \operatorname{bd}(\operatorname{dom} f)\end{array}\right\} \Rightarrow \langle \nabla f(x_n), y - x_n \rangle \to -\infty.$$

(ii) f is called *zone coercive*, if ∇f is onto.

It turns out that these concepts are just some old friends in disguise.

Theorem 4.5. *Suppose f is Bregman on E. Then:*

(i) *f is boundary coercive if and only if f is essentially smooth.*
(ii) *f is zone coercive if and only if f is essentially smooth and super-coercive.*

Proof. (i): "$\Rightarrow$": Fix $x \in \operatorname{bd}(\operatorname{dom} f)$, $y \in \operatorname{int}(\operatorname{dom} f)$, and $t_n \downarrow 0$. Set $x_n := x + t_n(y-x)$, for all n. By boundary coercivity,

$$-\infty \leftarrow \langle \nabla f(x_n), y - x_n \rangle = \langle \nabla f(x + t_n(y-x)), (1-t_n)(y-x) \rangle.$$

The last term is negative for all large n; hence $\langle \nabla f(x_n), y - x_n \rangle \geq \langle \nabla f(x + t_n(y-x)), y - x \rangle$, for all large n. It follows that $\langle \nabla f(x + t_n(y-x), y - x \rangle \to -\infty$, i.e. essential smoothness of f.
"$\Leftarrow$": Fix $y \in \operatorname{int}(\operatorname{dom} f)$ and (x_n) in $\operatorname{int}(\operatorname{dom} f)$ converging to $x \in \operatorname{bd}(\operatorname{dom} f)$. Since f and f^* are Legendre, we have $\|\nabla f(x_n)\| \to +\infty$ and $\nabla f(x_n)$, $\nabla f(y) \in \operatorname{int}(\operatorname{dom} f^*)$. Also, $D_{f^*}(\cdot, \nabla f(y))$ is coercive and hence

$$\begin{aligned} +\infty \leftarrow\ & D_{f^*}(\nabla f(x_n), \nabla f(y)) = D_f(y, x_n) \\ =\ & f(y) - f(x_n) - \langle \nabla f(x_n), y - x_n \rangle. \end{aligned}$$

By (B2), $f(x_n) \to f(x) \in \mathbb{R}$. Therefore, $\langle \nabla f(x_n), y - x_n \rangle \to -\infty$ and f is boundary coercive.
(ii): Fix a Bregman function f.
"$\Rightarrow$": By Theorem X.4.1.3 in [17], f^* is strictly convex on $\operatorname{dom} f^* = E$. In view of Fact 2.4 and Proposition 2.16, this means that f is essentially smooth and super-coercive.
"$\Leftarrow$": Super-coercivity is precisely $\operatorname{dom} f^* = E$; hence f^* is sub-differentiable everywhere: $\operatorname{dom} \partial f^* = \operatorname{range} \partial f = E$. Now ∂f is at most singleton (Fact 2.2); thus $\operatorname{range} \partial f = \operatorname{range} \nabla f = E$. The proof is complete. □

We now obtain Teboulle's generalization (see Burachik's Lemma 2.6 in [5]) of Iusem's Corollary 9.1 in [19].

Corollary 4.6. *For every Bregman function, the following implications hold:*

zone coercivity $\Rightarrow$ boundary coercivity $\Rightarrow$ Legendreness.

Of course, zone coercivity is genuinely stronger than boundary coercivity: The exponential function exp on $\mathbb{R}$ is boundary coercive but not zone coercive.
We end this section with a *verifiable* sufficient condition for "Bregmanness".

Theorem 4.7. *Suppose f is strictly convex and differentiable on* $\operatorname{int}(\operatorname{dom} f) = E$, *i.e.* f *is Legendre with* $\operatorname{dom} f = E$. *Then:*

(i) $\operatorname{dom} f^*$ *is open* $\Leftrightarrow$ f *is Bregman and boundary coercive.*
(ii) $\operatorname{dom} f^* = E \Leftrightarrow f$ *is Bregman and zone coercive.*

Proof. Fix f strictly convex and differentiable on $\operatorname{dom} f = E$. We first check that f is Bregman except for possibly (B3)(ii).
(B0): $\surd$. (B1): $\surd$ (see Remarks 4.2). (B2): $\surd$. (B3)(i): $\surd$ (see Remarks 4.2). (B3)(ii): Theorem 3.9 says: $D_f(x, \cdot)$ is coercive, $\forall x \in \operatorname{int}(\operatorname{dom} f) = \operatorname{dom} f = E$ if and only if $\operatorname{dom} f^*$ is open. (B4): $\surd$ (Proposition 3.2.(ii)). (B5): $\surd$ (Theorem 3.9.(iii)).
Hence: f is Bregman $\Leftrightarrow$ (B3)(ii) holds $\Leftrightarrow$ $\operatorname{dom} f^*$ is open. Consequently, (i) is true. But (ii) follows from Theorem 4.5 and Proposition 2.16. $\square$

Proposition 2.16 now implies:

Corollary 4.8. *Suppose f is strictly convex differentiable throughout E and super-coercive. Then f is Bregman and zone coercive.*

Remark 4.9. Corollary 4.8 improves upon De Pierro and Iusem's Theorem 5.1 in [29]. Their proof is quite complicated and different from the present, more conceptual proof.

5. Legendre functions and Bregman/Legendre functions

The last section underlined impressively the need for Legendre functions; we thus collect some basic facts.

Proposition 5.1. *Suppose f is Legendre on E.*

(i) *Suppose $\alpha > 0$. Then αf is Legendre with $D_{\alpha f} = \alpha D_f$.*
(ii) *Suppose g is closed convex proper on E and essentially smooth. If* $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \emptyset$, *then $f + g$ is Legendre with $D_{f+g} = D_f + D_g$.*
(iii) *Suppose T is an affine isomorphism of E. Then $f \circ T$ is Legendre with* $D_{f \circ T}(x, y) = D_f(Tx, Ty)$, $\forall x \in E, y \in \operatorname{int}(\operatorname{dom}(f \circ T)) = T^{-1}(\operatorname{int}(\operatorname{dom} f))$.
(iv) *Suppose s is convex, differentiable, and strictly increasing on $\mathbb{R}$. Then $s \circ f$ is Legendre with* $D_{s \circ f}(x, y) = D_s(f(x), f(y)) + \nabla s(f(y)) D_f(x, y)$, $\forall x \in E, y \in \operatorname{int}(\operatorname{dom}(s \circ f)) = \operatorname{int}(\operatorname{dom} f)$.
(v) *Suppose g is closed, convex, proper, essentially strictly convex on E with* $\operatorname{int}(\operatorname{dom} g^*) \cap \operatorname{int}(\operatorname{dom} f^*) \neq \emptyset$. *Then $f \Box g$ is Legendre.*

Proof. (i): is trivial.
(ii): $f + g$ is a closed convex proper function on E. We check L0 through L3.
L0: $\operatorname{int}(\operatorname{dom}(f + g)) = \operatorname{int}((\operatorname{dom} f) \cap (\operatorname{dom} g)) = \operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \emptyset$.
L1: $f + g$ is differentiable on $\operatorname{int}(\operatorname{dom}(f + g))$, since f and g are.
L2: Suppose $x \in \operatorname{bd}(\operatorname{dom}(f+g))$. Then $x \in \operatorname{bd}(\operatorname{dom} f) \cup \operatorname{bd}(\operatorname{dom} g)$. Fix $y \in \operatorname{int}(\operatorname{dom}(f + g)) = \operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g)$.

Case 1: $x \in \mathrm{bd}(\mathrm{dom}\, f)$. Then $\lim_{t\downarrow 0}\langle \nabla f(x+t(y-x)), y-x\rangle = -\infty$. If $x \in \mathrm{bd}(\mathrm{dom}\, g)$, then $\lim_{t\downarrow 0}\langle \nabla g(x+t(y-x)), y-x\rangle = -\infty$. Otherwise, $x \in \mathrm{int}(\mathrm{dom}\, g)$ and $\lim_{t\downarrow 0}\langle \nabla g(x+t(y-x)), y-x\rangle = \langle \nabla g(x), y-x\rangle \in \mathbb{R}$. For either alternative, $\lim_{t\downarrow 0}\langle \nabla (f+g)(x+t(y-x)), y-x\rangle = -\infty$, as sought for.
Case 2: $x \in \mathrm{int}(\mathrm{dom}\, f)$. Then $x \in \mathrm{bd}(\mathrm{dom}\, g)$ and we reason analogously.
L3: $f+g$ is strictly convex on $\mathrm{int}(\mathrm{dom}(f+g))$, since f is.
(iii): follows easily with results of Section 6 of [30].
(iv): L0, L1, and L3 are easy. For L2, recall $\nabla(s \circ f)(y) = \nabla s(f(y))\nabla f(y)$. Essential smoothness holds, since ∇s is increasing and strictly positive, and since f is minorized by affine functions.
(v): g^* is essentially smooth and f^* is Legendre. By (ii), $f^* + g^*$ is Legendre and so is $(f^*+g^*)^* = f\Box g$; see Theorem 16.4 of [30]. □

We now define Bregman/Legendre functions which form a subclass in the class of Legendre functions. However, they are more general than the class of functions that are both Bregman and Legendre.

Definition 5.2. Suppose f is Legendre on E. We say f *is Bregman/Legendre* (or a *Bregman/Legendre function*), if the following properties BL0-BL3 hold:
BL0. $\mathrm{dom}\, f^*$ is open.
BL1. $D_f(x,\cdot)$ is coercive, $\forall x \in \mathrm{dom}\, f \setminus \mathrm{int}(\mathrm{dom}\, f)$.
BL2.

$$\left.\begin{array}{r} x \in \mathrm{dom}\, f \setminus \mathrm{int}(\mathrm{dom}\, f), (y_n) \text{ in } \mathrm{int}(\mathrm{dom}\, f), \\ y_n \to y \in \mathrm{bd}(\mathrm{dom}\, f), (D_f(x,y_n)) \text{ bounded} \end{array}\right\} \Rightarrow \begin{array}{c} D_f(y,y_n) \to 0 \\ (\text{and hence } y \in \mathrm{dom}\, f). \end{array}$$

BL3.

$$\left.\begin{array}{r} (x_n), (y_n) \text{ in } \mathrm{int}(\mathrm{dom}\, f), x_n \to x \in \mathrm{dom}\, f \setminus \mathrm{int}(\mathrm{dom}\, f), \\ y_n \to y \in \mathrm{dom}\, f \setminus \mathrm{int}(\mathrm{dom}\, f), D_f(x_n,y_n) \to 0 \end{array}\right\} \Rightarrow x = y.$$

Remark 5.3.

- In view of Corollary 3.11, BL0 and BL1 together say that $D_f(x,\cdot)$ is coercive, $\forall x \in \mathrm{dom}\, f$. The split into BL0 and BL1 is on purpose and will allow us to make clear which part is used; see, for instance, the proof of Theorem 8.1. Also, BL0 and BL1 together are (B3)(ii) in Definition 4.1. Again, there is a nice split as in Remarks 4.2:
  •• $\mathrm{dom}\, f$ open: BL0 and BL1 $\Leftrightarrow$ $\mathrm{dom}\, f^*$ open $\Leftrightarrow$ $D_f(x,\cdot)$ is coercive, for some $x \in \mathrm{dom}\, f$.
  •• $\mathrm{dom}\, f$ not open: BL0 and BL1 $\Leftrightarrow$ $\mathrm{dom}\, f^*$ open and BL1 $\Leftrightarrow$ $D_f(x,\cdot)$ is coercive, for all $x \in \mathrm{dom}\, f \setminus \mathrm{int}(\mathrm{dom}\, f)$.
- BL1 is equivalent to

$$\left.\begin{array}{r} x \in \mathrm{dom}\, f \setminus \mathrm{int}(\mathrm{dom}\, f), (y_n) \text{ in } \mathrm{int}(\mathrm{dom}\, f), \\ (D_f(x,y_n)) \text{ bounded} \end{array}\right\} \Rightarrow (y_n) \text{ bounded.}$$

- BL2 is at least formally more general than (B4) in Definition 4.1.
- BL3 is equivalent to (B5) in Definition 4.1; however, BL3 is slightly easier to check.
- The "Boltzmann/Shannon entropy" $f(x) := x\ln x - x$ on $\mathrm{dom}\, f := [0,+\infty[$ is Bregman/Legendre. Given $0 \in \mathrm{bd}(\mathrm{dom}\, f)$ and $y_n \to +\infty$, we can find a sequence $x_n \downarrow 0$

such that $D_f(x_n, y_n) \to 0$ by reckoning similarly to Example 3.10 and by using the decomposition

$$D_f(x_n, y_n) = D_f(0, y_n) + f(x_n) - f(0) - \nabla f(y_n)(x_n - 0).$$

This shows the importance of the hypotheses in BL1 and BL3.

In Section 6, we see that Definition 5.2 is flexible enough to include a nice large set of examples.

Remark 5.4. For closed convex proper functions on E, the following strict implications hold:

$$\text{Legendre} \Leftarrow \text{Bregman/Legendre} \Leftarrow \text{Bregman and Legendre} \Rightarrow \text{Bregman}.$$

The strictness of the implications follows from the following examples: exp is Legendre but neither Bregman/Legendre nor Bregman. $-\ln$ is Bregman/Legendre but not Bregman. $f(x) = \frac{1}{2}|x|^2$, if $x \geq 0$; $+\infty$, else, is Bregman but neither Bregman/Legendre nor Legendre.

We now develop the basic facts on Bregman/Legendre functions. The following Proposition will be useful later.

Proposition 5.5. *Suppose f is Bregman/Legendre on E. Then:*

$$\left.\begin{array}{r} (y_n) \text{ in } \operatorname{int}(\operatorname{dom} f),\ y_n \to y \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f), \\ x \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f),\ D_f(x, y_n) \to 0 \end{array}\right\} \Rightarrow\ x = y.$$

Proof. Suppose the hypothesis of the implication holds. Let $x_n := (1-1/n)x + (1/n)y_n$, for all $n \geq 2$. Then the sequence (x_n) lies in $\operatorname{int}(\operatorname{dom} f)$ and $x_n \to x$. The convexity of $D_f(\cdot, y_n)$ yields $D_f(x_n, y_n) \leq D_f(x, y_n) \to 0$. Now apply BL3 and conclude $x = y$. □

Checking "Bregman/Legendreness" can be very easy:

Theorem 5.6. *Suppose f is Legendre on E with* dom f *open. Then:*

$$f \textit{ is Bregman/Legendre} \quad \Leftrightarrow \quad \operatorname{dom} f^* \textit{ is open.}$$

Proof. BL0 and BL1 $\Leftrightarrow$ dom f^* open, as observed in Remarks 5.3.
Now BL2 and BL3 hold trivially. □

Remark 5.7. De Pierro and Iusem (Theorem 5.1 in [29]) proved that if f is twice continuously differentiable, strictly convex on E and super-coercive, then f is Bregman. Theorem 5.6 can be viewed as a very potent generalization of their result; see also Corollary 4.8.

On the real line, we only have to check the domain of the conjugate:

Theorem 5.8. *Suppose f is Legendre on* $\mathbb{R}$. *Then:*

(i) BL0 *and* BL1 $\Leftrightarrow$ dom f^* *is open.*
(ii) BL2 *always holds.*
(iii) BL3 *always holds.*

Consequently,

$$f \text{ is Bregman/Legendre} \quad \Leftrightarrow \quad f \text{ is Legendre and } \operatorname{dom} f^* \text{ is open.}$$

Proof. The theorem is clear if $\operatorname{dom} f$ is open (Theorem 5.6). So we assume without loss that $\operatorname{dom} f$ is not open.

(i): "$\Rightarrow$": follows from Corollary 3.11.
"$\Leftarrow$": In view of Corollary 3.11, it is enough to show that BL1 holds. We can assume that $\operatorname{dom} f$ is unbounded, say $\operatorname{dom} f = [x, +\infty[$ (the remaining case $\operatorname{dom} f =]-\infty, x]$ is treated analogously). Now let us assume to the contrary that BL1 fails, i.e. there is a sequence (y_n) in $\operatorname{int}(\operatorname{dom} f) =]x, +\infty[$ with $y_n \uparrow +\infty$ but $(D_f(x, y_n))$ bounded.
Claim: $\nabla f(y_n) \to +\infty$.
Because f is convex, ∇f is increasing. If the claim does not hold, then we have $\nabla f(y_n) \to y^*$. Now

$$(f(x) + f^*(\nabla f(y_n)) - \langle \nabla f(y_n), x \rangle)$$

is a bounded sequence. Since f^* is closed, we get $y^* \in \operatorname{dom} f^* = \operatorname{int}(\operatorname{dom} f^*)$ and further $y_n = \nabla f^*(\nabla f(y_n)) \to \nabla f^*(y^*)$ which is absurd. The claim thus holds.
Apply Proposition 3.6 to obtain $D_f(x, y_n) \to +\infty$, the desired contradiction.

(ii): Suppose x, y, and the sequence (y_n) are as in the hypothesis of BL2.
In view of Proposition 3.3, we only have to show that $y \in \operatorname{dom} f$. This is obviously true if $\operatorname{dom} f$ is closed. Hence we can assume that $\operatorname{dom} f$ is neither open nor closed, i.e. of the form $[a, b[$ (or $]a, b]$ but this is again treated similarly). The only (potentially) "critical" case is therefore $\operatorname{dom} f = [x, y[$. Since $y_n \to y \in \operatorname{bd}(\operatorname{dom} f)$, the sequence $(\nabla f(y_n))$ has to tend to $+\infty$. By Proposition 3.6, $D_f(x, y_n) \to +\infty$ which contradicts our assumption. Thus the "critical" case never occurs and BL2 is established, i.e. (ii) holds.

(iii): Suppose (x_n), (y_n), x, and y are as in the hypothesis of BL3. Now BL3 holds trivially unless $\operatorname{dom} f$ has *two* boundary points. We assume to the contrary that BL3 fails, i.e. $\operatorname{dom} f = [x, y]$.
Case 1: $x < y$.
Note that $\nabla f(y_n) \to +\infty$. Now

$$0 \leftarrow D_f(x_n, y_n) = f(x_n) - f(y_n) - \nabla f(y_n)(x_n - y_n);$$

hence $\nabla f(y_n) - (f(x) - f(y))/(x - y) \to 0$, which is absurd.
Case 2: $x > y$ is proved similarly. □

Remark 5.9.

- Theorem 5.8 is powerful and extends easily to separable functions (see Corollary 5.13 below). Having used so heavily the one-dimensionality in the proof of Theorem 5.8, it doesn't come as a surprise that the non-separable multi-dimensional case is much more involved; see Section 6.
- The previous results yield easily the following characterization:

  f is Bregman and Legendre on $\mathbb{R}$ if and only if
  f is Legendre on $\mathbb{R}$, $\operatorname{dom} f$ is closed, and $\operatorname{dom} f^*$ is open.

Proposition 5.10. *Suppose f is Bregman/Legendre on E.*

(i) *If $\alpha > 0$, then αf is Bregman/Legendre.*

(ii) *Suppose g is essentially smooth on E with*

$$\left.\begin{array}{r} x \in \operatorname{dom} g, (y_n) \textit{ in } \operatorname{int}(\operatorname{dom} g), \\ y_n \to y \in \operatorname{cl}(\operatorname{dom} g), (D_g(x, y_n)) \textit{ bounded} \end{array}\right\} \Rightarrow D_g(y, y_n) \to 0. \tag{S2}$$

If $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \emptyset$, then $f + g$ is Bregman/Legendre.

(iii) *If T is an affine isomorphism of E, then $f \circ T$ is Bregman/Legendre.*

Proof. (i): is trivial.
(ii): Let $h := f + g$. Then h is Legendre with Bregman distance $D_h = D_f + D_g$ (Proposition 5.1.(ii)). We have to check BL0 through BL3 for h.
BL0: Fix a point $x \in \operatorname{int}(\operatorname{dom} h)$ and a sequence (y_n) in $\operatorname{int}(\operatorname{dom} h)$ with $(D_h(x, y_n))$ bounded. Then $x \in \operatorname{int}(\operatorname{dom} f)$, (y_n) lies in $\operatorname{int}(\operatorname{dom} f)$, and $(D_f(x, y_n))$ is bounded. Now $\operatorname{dom} f^*$ is open, thus (y_n) is bounded (Corollary 3.11 for f). Since x and (y_n) were chosen arbitrarily, Corollary 3.11 applies once more and yields the openness of $\operatorname{dom} h^*$.
BL1: Fix $x \in \operatorname{dom} h \setminus \operatorname{int}(\operatorname{dom} h)$ and a sequence (y_n) in $\operatorname{int}(\operatorname{dom} h)$ with $(D_h(x, y_n))$ bounded. Note that (y_n) is in $\operatorname{int}(\operatorname{dom} f)$ and that $(D_f(x, y_n))$ is bounded. If $x \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$, then, by BL1 for f, the sequence (y_n) is bounded. Otherwise, $x \in \operatorname{int}(\operatorname{dom} f)$ and the boundedness of (y_n) follows from BL0 and Corollary 3.11 (for f).
BL2: Fix $x \in \operatorname{dom} h \setminus \operatorname{int}(\operatorname{dom} h)$ and (y_n) in $\operatorname{int}(\operatorname{dom} h)$ with $y_n \to y \in \operatorname{bd}(\operatorname{dom} h)$ and $(D_h(x, y_n))$ bounded. Then $(D_f(x, y_n))$ and $(D_g(x, y_n))$ are bounded. S2 implies $y \in \operatorname{dom} g$ and $D_g(y, y_n) \to 0$. It suffices to show that $y \in \operatorname{dom} f$ and $D_f(y, y_n) \to 0$. *Case 1:* $x \in \operatorname{int}(\operatorname{dom} f)$. Employ Theorem 3.8.(ii). *Case 2:* $x \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$. If $y \in \operatorname{int}(\operatorname{dom} f)$, then use Proposition 3.2.(ii). Otherwise, $y \in \operatorname{bd}(\operatorname{dom} f)$, and we apply BL2.
BL3: Fix sequences (x_n), (y_n) in $\operatorname{int}(\operatorname{dom} h)$ with $x_n \to x \in \operatorname{dom} h \setminus \operatorname{int}(\operatorname{dom} h)$, $y_n \to y \in \operatorname{dom} h \setminus \operatorname{int}(\operatorname{dom} h)$, and $D_h(x_n, y_n) \to 0$. If $\{x, y\} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, then we make use of Theorem 3.9.(iii). Otherwise, $x, y \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$ and BL3 does it.
(iii): Denote $f \circ T$ by g and the linear part of T by L (i.e. $Tx = Lx + T0$, for all $x \in E$). Then one checks that $g^*(x^*) = f^*((L^*)^{-1}x^*) - \langle (L^*)^{-1}, T0 \rangle$ so that $\operatorname{dom} g^* = L^* \operatorname{dom} f^*$ is open. Hence BL0 holds. BL1 through BL3 for g follow straight-forwardly from the corresponding properties for f. □

Corollary 5.11. *Suppose f is Bregman/Legendre on E. Suppose further g is affine or Bregman/Legendre on E. If $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \emptyset$, then $f + g$ is Bregman/Legendre.*

Proof. It is enough to show that g satisfies the hypothesis of Proposition 5.10.(ii). If g is affine, this is done easily. Otherwise, g is Bregman/Legendre with $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \emptyset$ and we only have to check S2. So fix $x \in \operatorname{dom} g$, (y_n) in $\operatorname{int}(\operatorname{dom} g)$ with $y_n \to y \in \operatorname{cl}(\operatorname{dom} g)$ and $(D_g(x, y_n))$ bounded. Goal: $y \in \operatorname{dom} g$ and $D_g(y, y_n) \to 0$. *Case 1:* $x \in \operatorname{int}(\operatorname{dom} g)$. Apply Theorem 3.8.(ii). *Case 2:* $x \in \operatorname{dom} g \setminus \operatorname{int}(\operatorname{dom} g)$. If $y \in \operatorname{int}(\operatorname{dom} g)$, then use Proposition 3.2.(ii). Else $y \in \operatorname{bd}(\operatorname{dom} g)$ and BL2 (for g) does the job. □

Theorem 5.12. *Suppose $f_1, .., f_J$ are Bregman/Legendre on Euclidean spaces $E_1, .., E_J$, respectively. If $\lambda_1, \dots, \lambda_J$ are strictly positive real numbers, then*

$$\mathbf{f} \quad : \quad \mathbf{E} := \prod_j E_j \to \,]-\infty, +\infty] \quad : \quad \mathbf{x} := (x_1, \dots, x_J) \mapsto \sum_j \lambda_j f_j(x_j)$$

is Bregman/Legendre on $\mathbf{E}$.

Proof. Note that $\nabla \mathbf{f}(\mathbf{x}) = (\nabla f_1(x_1), \dots, \nabla f_J(x_J))$, $\mathrm{int}(\mathrm{dom}\,\mathbf{f}) = \prod_j \mathrm{int}(\mathrm{dom}\, f_j)$, and $D_{\mathbf{f}}(\mathbf{x}, \mathbf{y}) = \sum_j D_{f_j}(x_j, y_j)$ (Proposition 3.5). We first check that $\mathbf{f}$ is Legendre; see Definition 2.8.
L0: $\surd$. L1: $\surd$. L2: Take $(\mathbf{x}_n)$ in $\mathrm{int}(\mathrm{dom}\, f)$ with $\mathbf{x}_n \to \mathbf{x} = (x_1, \dots, x_J) \in \mathrm{bd}(\mathrm{dom}\,\mathbf{f})$. Then there is a j such that $x_j \in \mathrm{bd}(\mathrm{dom}\, f_j)$. Hence, by Fact 2.2,

$$\|\nabla \mathbf{f}(\mathbf{x}_n)\| \geq \|\nabla f_j((x_n)_j)\| \to +\infty;$$

thus L2 holds. L3: $\surd$.
Next, we check BL0-BL3.
BL0,BL1: Since each f_j are Bregman/Legendre, we have that each $D_{f_j}(x_j, \cdot)$ is coercive, $\forall x_j \in \mathrm{dom}\, f_j$. It easily follows that $D_{\mathbf{f}}(\mathbf{x}, \cdot)$ is coercive, $\forall \mathbf{x} \in \mathrm{dom}\,\mathbf{f}$.
BL2: Pick $\mathbf{x}$, $\mathbf{y}$, and $(\mathbf{y}_n)$ as in the hypothesis of BL2 for $\mathbf{f}$. If $y_j \in \mathrm{int}(\mathrm{dom}\, f_j)$, apply Proposition 3.2.(ii) to conclude $D_{f_j}(y_j, (\mathbf{y}_n)_j) \to 0$. Else $y_j \in \mathrm{bd}(\mathrm{dom}\, f_j)$ and, depending on the location of x_j, either Theorem 3.8.(ii) or BL2 for f_j applies.
BL3: Fix $\mathbf{x}$, $\mathbf{y}$, $(\mathbf{x}_n)$, and $(\mathbf{y}_n)$ as in the hypothesis of BL3 for $\mathbf{f}$. Then for every j, we obtain $x_j = y_j$ either by Theorem 3.9.(iii) or by BL3 for f_j. $\square$

Corollary 5.13. *Suppose $f_1, \dots, f_J$ are Legendre on $\mathbb{R}$ with $\mathrm{dom}\, f_1^*, \dots, \mathrm{dom}\, f_J^*$ open. Then*

$$f \quad : \quad E = \mathbb{R}^J \to \,]-\infty, +\infty] \quad : \quad x := (x_1, \dots, x_J) \mapsto \sum_j f_j(x_j)$$

is Bregman/Legendre type on E.

Proof. Combine Theorem 5.8 with Theorem 5.12. $\square$

6. Examples

6.1. Bregman/Legendre functions on the real line

All convex functions in the examples of this subsection are Legendre. Deciding whether or not they are Bregman/Legendre is, thanks to Theorem 5.8, extremely easy: simply check the openness of $\mathrm{dom}\, f^*$. Examples 6.1 to 6.7 are standard; see Rockafellar's Section 12 of [30].

Example 6.1. Suppose $1 < p < +\infty$ and $f(x) = \frac{1}{p}|x|^p$ on $\mathrm{dom}\, f = \mathbb{R}$. Then $f^*(x^*) = \frac{1}{q}|x^*|^q$ on $\mathrm{dom}\, f^* = \mathbb{R}$, where $\frac{1}{p} + \frac{1}{q} = 1$.
Hence f and f^* are Bregman/Legendre.

We state an important special instance of the preceding example:

Example 6.2. (norm^2) $f(x) = \frac{1}{2}|x|^2$ is Bregman/Legendre on $\mathbb{R}$.

Example 6.3. Suppose $0 < p < 1$ and $f(x) = -\frac{1}{p}x^p$ on $\operatorname{dom} f = [0, +\infty[$. Then $f^*(x^*) = -\frac{1}{q}(-x^*)^q$ on $\operatorname{dom} f^* =]-\infty, 0[$, where $\frac{1}{p} + \frac{1}{q} = 1$.
Hence f is Bregman/Legendre whereas f^* is not.

Example 6.4. ("Hellinger") Suppose $f(x) = -\sqrt{1 - x^2}$ on $\operatorname{dom} f = [-1, +1]$. Then $f^*(x^*) = \sqrt{1 + (x^*)^2}$ on $\operatorname{dom} f^* = \mathbb{R}$.
Hence f is Bregman/Legendre whereas f^* is not.

Example 6.5. ("Boltzmann/Shannon") Suppose $f(x) = x \ln x - x$ on $\operatorname{dom} f = [0, +\infty[$. Then $f^*(x^*) = \exp x^*$ on $\operatorname{dom} f^* = \mathbb{R}$.
Hence f is Bregman/Legendre whereas f^* is not.

Example 6.6. ("Fermi/Dirac") Suppose $f(x) = x \ln x + (1 - x) \ln(1 - x)$ on $\operatorname{dom} f = [0, 1]$. Then $f^*(x^*) = \ln(1 + \exp x^*)$ on $\operatorname{dom} f^* = \mathbb{R}$.
Hence f is Bregman/Legendre whereas f^* is not.

Example 6.7. ("Burg") Suppose $f(x) = -\frac{1}{2} - \ln x$ on $\operatorname{dom} f =]0, +\infty[$. Then $f^*(x^*) = -\frac{1}{2} - \ln(-x^*)$ on $\operatorname{dom} f^* =]-\infty, 0[$.
Hence f and f^* are Bregman/Legendre.

Example 6.8. (De Pierro & Iusem's Example on page 438 in [29]) Suppose $f(x) = \frac{1}{2}(x^2 - 4x + 3)$, if $x \leq 1$; $-\ln x$, otherwise; on $\operatorname{dom} f = \mathbb{R}$. Then $f^*(x^*) = \frac{1}{2}(x^*)^2 + 2x^* + \frac{1}{2}$, if $x^* \leq -1$; $-1 - \ln(-x^*)$, if $-1 \leq x^* < 0$; on $\operatorname{dom} f^* =]-\infty, 0[$.
Hence f and f^* are Bregman/Legendre.

We summarize these examples in table 1 on page 23.

6.2. Multi-dimensional (Bregman/)Legendre functions

On the real line or in the separable case, Theorem 5.8 and Theorem 5.12 (see also Corollary 5.13) provide an extremely easy and elegant check for the somewhat cumbersome conditions in the definition of a Bregman/Legendre function (Definition 5.2). The question arises if BL1 through BL3 hold "for free" as soon as BL0 holds. This is, however, false as the following example illustrates:

Example 6.9. Suppose

$$f : \mathbb{R}^2 \to]-\infty, +\infty] : (x, r) \mapsto \begin{cases} r \ln(r^2/x), & \text{if } x, r > 0; \\ 0, & \text{if } x \geq 0,\ r = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then f is Legendre and

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

Hence $\operatorname{dom} f^*$ is open, i.e. BL0 holds. Nonetheless, f is not Bregman/Legendre, because BL1 through BL3 fail.

Some Bregman/Legendre functions on $\mathbb{R}$		
$f(x)$	dom f	Remarks
$\frac{1}{p}\lvert x\rvert^p$	$\mathbb{R}$	$1 < p < +\infty$
$\frac{1}{2}x^2$	$\mathbb{R}$	norm^2
$-\frac{1}{p}x^p$	$[0,+\infty[$	$0 < p < 1$
$-\sqrt{1-x^2}$	$[-1,+1]$	"Hellinger"
$x\ln x - x$	$[0,+\infty[$	"Boltzmann/Shannon"
$x\ln x + (1-x)\ln(1-x)$	$[0,1]$	"Fermi/Dirac"
$-\frac{1}{2}\ln x$	$]0,+\infty[$	"Burg"
$\frac{1}{2}(x^2-4x+3),\ x \le 1;$ $-\ln x,\ x \ge 1.$	$\mathbb{R}$	"De Pierro & Iusem"

Table 1.

Proof. Verifying that f is of Legendre type and calculating f^* is straight-forward but somewhat tedious. Thus we indicate only where BL1 through BL3 go wrong.

BL1: Set $x := (1,0)$ and $y_n := (n,1)$, for all n. Then $D_f(x,y_n) = 1 + 1/n \to 1$, but (y_n) is clearly unbounded.

BL2: Choose $x := (0,0)$, $y := (0,1)$, and $y_n := (1/n,1)$, for all n. Then $y_n \to y \notin \operatorname{dom} f$, although $D(x,y_n) \equiv 1$.

BL3: Let $x := (2,0)$, $y := (1,0)$, $x_n := (2,1/n)$, and $y_n := (1,1/n)$, for all n. Then $D_f(x_n,y_n) = (1-\ln(2n))/n \to 0$; however, $x \neq y$. $\square$

Remark 6.10. It might appear that we pulled Example 6.9 out of a hat. This is not true; in fact, a nice systematic way to generate interesting convex non-separable functions is as follows:

Let g be defined on $\mathbb{R}^{J-1}$ with $\operatorname{int}(\operatorname{dom} g) = \{x \in \mathbb{R}^{J-1} : x > 0\}$ and h be defined on $\mathbb{R}$ with $\operatorname{int}(\operatorname{dom} h) = \{r \in \mathbb{R} : r > 0\}$. Construct

$$f : \mathbb{R}^J \to \,]-\infty,+\infty] : (x,r) \mapsto \begin{cases} rg(x/r) + h(r), & \text{if } x \in \operatorname{dom} g \text{ and } r > 0; \\ (g0^+)(x) + h(0), & \text{if } x \in \operatorname{dom} g \text{ and } r = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then f is closed convex proper; see Rockafellar's Remark following Corollary 8.5.1 on page 67 in [30]. The reader will enjoy discovering nice patterns such as

$$D_f((x,r),(y,s)) = rD_g(x/r,y/s) + D_h(r,s),$$

for $(x,r),(y,s) \in \operatorname{int}(\operatorname{dom} f)$. Example 6.9 arises by choosing

$$g(x) = -\ln x \quad \text{and} \quad h(r) = r \ln r.$$

6.3. *Bregman projections*

Having built a stock of examples of Bregman/Legendre functions, we now consider Bregman projections.

Proposition 6.11. (see also Censor and Elfving's Lemma 6.1 in [6]) *Suppose f is Legendre on E and C is an affine subspace of E, say $C = \{x \in E : Ax = b\}$, for $A : E \to \mathbb{R}^M : x \mapsto (\langle a^{(m)}, x\rangle)_{m=1}^M$ and $b \in \mathbb{R}^M$. Suppose further $C \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and $y \in \operatorname{int}(\operatorname{dom} f)$. Then $z = P_C y$ exactly when*

$$z \in \operatorname{int}(\operatorname{dom} f), \quad Az = b, \quad \textit{and} \quad \nabla f(z) = \nabla f(y) + \textstyle\sum_{m=1}^M \mu_m a^{(m)},$$

for some parameters $\mu_1, \ldots, \mu_M \in \mathbb{R}$ which will be unique whenever A is onto (equivalently, the vectors $a^{(1)}, \ldots, a^{(M)}$ are linearly independent).

Proof. This is clear from Proposition 3.16 and the fact that $(\operatorname{kernel} A)^\perp = \operatorname{range} A^* = \operatorname{span}(a^{(1)}, \ldots, a^{(M)})$. □

Corollary 6.12. (see also Bregman's Theorem 3 in [4]) *Suppose f is a convex function of Legendre type on E and H is a hyperplane in E, say $H = \{x \in E : \langle a, x\rangle = b\}$, for some $a \in E \setminus \{0\}$ and $b \in \mathbb{R}$. Suppose further $H \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and $y \in \operatorname{int}(\operatorname{dom} f)$. Then $z = P_H y$ exactly when*

$$z \in \operatorname{int}(\operatorname{dom} f), \quad \langle a, z\rangle = b, \quad \textit{and} \quad \nabla f(z) = \nabla f(y) + \mu a,$$

for some (unique) parameter $\mu \in \mathbb{R}$.

Remark 6.13. In practice, the projection $P_H y$ (in the setting of Corollary 6.12) can be computed as follows:

1. Get z as a function of μ by Fact 2.9:

$$z(\mu) = (\nabla f^*)(\nabla f(y) + \mu a).$$

2. Estimate μ by solving $\langle a, z(\mu)\rangle = b$ subject to $z(\mu) \in \operatorname{int}(\operatorname{dom} f)$.
3. Compute $z(\mu)$ through 1.

Table 2 contains the function $z(\mu) = (z_j(\mu))_{j=1}^J$ of Step 1 for some convex functions of Bregman/Legendre type of the form $f(x) = \sum_j f_j(x_j)$

Step 2 is quite hard: only for $f(x) = \frac{1}{2}\|x\|^2$ can one solve $\langle a, z(\mu)\rangle = b$ explicitly for μ (and one then recovers the well-known formula for the orthogonal projection onto a hyperplane). However, μ can be estimated by iterative methods — in some cases, even *a one-step approximation by the secant method* is enough to guarantee convergence of a particular case of the method of Bregman projections (defined in Section 8); this nice observation is due to Censor et al. [9].

$f_j(x_j)$	$z_j(\mu)$
$\frac{1}{2}x_j^2$	$y_j + \mu a_j$
$x_j \ln x_j - x_j$	$y_j \exp(\mu a_j)$
$x_j \ln x_j + (1-x_j)\ln(1-x_j)$	$\dfrac{\exp(\mu a_j) y_j/(1-y_j)}{1+\exp(\mu a_j) y_j/(1-y_j)}$
$-\frac{1}{2} - \ln x_j$	$y_j/(1-\mu a_j y_j)$
$\frac{1}{p}\lvert x_j\rvert^p$	$(\operatorname{sign}(y_j)\lvert y_j\rvert^{p-1} + \mu a_j)^{1/(p-1)}$
$-\sqrt{1-x_j^2}$	$\dfrac{\mu a_j + y_j/(1-y_j^2)^{1/2}}{\sqrt{(\mu a_j + y_j/(1-y_j^2)^{1/2})^2+1}}$
$-4\sqrt{x_j}$	$y_j/(1-\sqrt{y_j}\mu a_j/2)^2$

Table 2.

Remark 6.14. Frequently, the interior of domain of the Legendre function on E is the strictly positive cone $E^+ := \{x \in E : x_j > 0,\ \forall j\}$. Let a hyperplane in E be given by $H := \{x \in E : \langle a, x\rangle = b\}$, for some $a \in E \setminus \{0\}$ and $b \in \mathbb{R}$. Then $H \cap E^+ = \emptyset$ if and only

$$a \geq 0 \text{ and } b \leq 0 \qquad \text{or} \qquad a \leq 0 \text{ and } b \geq 0.$$

In applications, typically $a \geq 0$, $a \neq 0$, and $b > 0$; therefore, $H \cap E^+$ is nonempty and Corollary 6.12 applies.

Example 6.15. Let $f(x) = \sum_j x_j \ln x_j - x_j$ on $\operatorname{dom} f = \{x \in \mathbb{R}^J : x_j \geq 0,\ \forall j\}$ and H be the "probabilistic hyperplane" $H = \{x \in \mathbb{R}^J : \sum_j x_j = 1\}$. By Remark 6.14 (or by considering $(1/J)(1,1,\ldots,1) \in \mathbb{R}^J$), we have $H \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Moreover, the projection onto H has the beautiful form

$$P_H y = \frac{1}{\sum_j y_j} y;$$

as is readily verified using Remark 6.13.

7. More Examples

7.1. Product spaces and Pythagorean means

Throughout this subsection, we assume that $f_1, \ldots, f_N$ are Bregman/Legendre functions on some Euclidean space $E = \mathbb{R}^J$, that $C_1, \ldots, C_N$ are closed convex nonempty subsets of E, and that $\lambda_1, \ldots, \lambda_N$ are strictly positive real numbers.

By Theorem 5.12,

$$\mathbf{f} \;:\; \mathbf{E} := \prod_i E \to \,]-\infty, +\infty] \;:\; \mathbf{x} := (x_1, \dots, x_N) \mapsto \sum_i \lambda_i f_i(x_i)$$

is Bregman/Legendre on $\mathbf{E}$.
Define the product set

$$\mathbf{C} := \prod_i C_i = \{(x_1, \dots, x_N) \in \mathbf{E} : x_i \in C_i, \ \forall i\}$$

and the diagonal set

$$\boldsymbol{\Delta} := \{(e, \dots, e) \in \mathbf{E} : e \in E\}.$$

Then $\mathbf{C} \cap \boldsymbol{\Delta} \neq \emptyset$ if and only if $\bigcap_i C_i \neq \emptyset$; this reduction to two sets in the product space $\mathbf{E}$ goes back at least as far as Pierra [28].
Now what do the Bregman projections onto $\mathbf{C}$ and $\boldsymbol{\Delta}$ with respect to $\mathbf{f}$ look like? They are well-defined (Theorem 3.12) as soon as $\mathbf{C} \cap \operatorname{int}(\operatorname{dom} \mathbf{f}) \neq \emptyset$ and $\boldsymbol{\Delta} \cap \operatorname{int}(\operatorname{dom} \mathbf{f}) \neq \emptyset$; equivalently:

$$C_i \cap \operatorname{int}(\operatorname{dom} f_i) \neq \emptyset, \quad \text{for all } i; \quad \text{and} \quad \bigcap_i \operatorname{int}(\operatorname{dom} f_i) \neq \emptyset. \tag{$*$}$$

The next proposition is easily verified; see also the closely related Lemmata 4.1 and 4.2 in [6] by Censor and Elfving.

Proposition 7.1. *If* $(*)$ *holds, then for every* $\mathbf{y} = (y_1, \dots, y_N) \in \prod_i \operatorname{int}(\operatorname{dom} f_i)$*:*

(i) $P^{\mathbf{f}}_{\mathbf{C}}(\mathbf{y}) = (P^{f_1}_{C_1}(y_1), \dots, P^{f_N}_{C_N}(y_N))$.

(ii) $\mathbf{z} = (z, \dots, z) = P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y})$ *if and only if*

$$z \in \bigcap_i \operatorname{int}(\operatorname{dom} f_i) \quad \textit{and} \quad \sum_i \lambda_i \nabla f_i(z) = \sum_i \lambda_i \nabla f_i(y_i).$$

Proof. (i): Obvious, since $D_{\mathbf{f}}(\mathbf{x}, \mathbf{y}) = \sum_i \lambda_i D_{f_i}(x_i, y_i)$ is separable.
(ii): By Proposition 3.16, $\mathbf{z} = (z, \dots, z) = P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y})$ if and only if $z \in \bigcap_i \operatorname{int}(\operatorname{dom} f_i)$ and

$$\langle \nabla \mathbf{f}(\mathbf{y}) - \nabla \mathbf{f}(\mathbf{z}), \boldsymbol{\Delta} - \mathbf{z} \rangle \leq 0.$$

Now $\boldsymbol{\Delta} - \mathbf{z} = \boldsymbol{\Delta}$ is a subspace and $\boldsymbol{\Delta}^{\perp} = \{(x_1, \dots, x_N) \in \mathbf{E} : \sum_i x_i = 0\}$; the result follows. ☐

Corollary 7.2. *Suppose* f *is Bregman/Legendre on* E *and* $\lambda_1, \dots, \lambda_N$ *are strictly positive weights:* $\sum_i \lambda_i = 1$. *Let* $\mathbf{f}(\mathbf{x}) := \sum_i \lambda_i f(x_i)$, *for all* $\mathbf{x} = (x_1, \dots, x_N) \in \mathbf{E}$. *If* $\mathbf{y} = (y_1, \dots, y_N) \in \prod_i \operatorname{int}(\operatorname{dom} f)$, *then*

$$P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y}) = (z, \dots, z), \quad \textit{where} \quad z = \nabla f^*\Big(\sum_i \lambda_i \nabla f(y_i)\Big).$$

Proof. Clear from Proposition 7.1.(ii) and Fact 2.9. Note that z is indeed in $\operatorname{int}(\operatorname{dom} f)$ by Fact 2.9 and the fact that the interior of a convex set is convex: $y_i \in \operatorname{int}(\operatorname{dom} f)$, $\forall i$

$\Rightarrow \nabla f(y_i) \in \operatorname{int}(\operatorname{dom} f^*),\ \forall i \Rightarrow \sum_i \lambda_i \nabla f(y_i) \in \operatorname{int}(\operatorname{dom} f^*) \Rightarrow \nabla f^*(\sum_i \lambda_i \nabla f(y_i)) = z \in \operatorname{int}(\operatorname{dom} f)$. $\square$

Corollary 7.2 allows explicit calculation of Bregman projections onto the diagonal. It is pleasing that these projections turn out to be Pythagorean means if we use the best known Bregman/Legendre functions. Until the end of this subsection, let $\lambda_1, \ldots, \lambda_N$ be strictly positive weights. We denote the j^{th} coordinate of a vector $x \in E$ by $x(j)$. The following examples are readily verified with Corollary 7.2:

Example 7.3. (p-norm and the $(p-1)$-Hölder mean) Suppose $1 < p < +\infty$ and $f(x) = \frac{1}{p}\|x\|_p^p = \frac{1}{p}\sum_j |x(j)|^p$ on $\operatorname{dom} f = E = \mathbb{R}^J$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = \mathbf{E} = (\mathbb{R}^J)^N$. Then $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y}) = (z, \ldots, z)$, where

$$z(j) = (\lambda_1 |y_1(j)|^{p-1} + \cdots + \lambda_N |y_N(j)|^{p-1})^{\frac{1}{p-1}}, \quad \text{for all } j;$$

i.e. the j^{th} coordinate of $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y})$ is the $(p-1)$-*Hölder mean* of $y_1(j), \ldots, y_N(j)$.

Two special cases deserve further attention; for $p = 2$ we obtain:

Example 7.4. (2-norm and the arithmetic mean) Suppose $f(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}\sum_j |x(j)|^2$ on $\operatorname{dom} f = E = \mathbb{R}^J$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = \mathbf{E} = (\mathbb{R}^J)^N$. Then

$$P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y}) = \lambda_1 y_1 + \cdots + \lambda_N y_N;$$

i.e. the j^{th} coordinate of $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y})$ is the *arithmetic mean* of $y_1(j), \ldots, y_N(j)$.

And $p = 3$ gives:

Example 7.5. (3-norm and the quadratic mean) Suppose $f(x) = \frac{1}{3}\|x\|_3^3 = \frac{1}{3}\sum_j |x(j)|^3$ on $\operatorname{dom} f = E = \mathbb{R}^J$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = \mathbf{E} = (\mathbb{R}^J)^N$. Then $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y}) = (z, \ldots, z)$, where

$$z(j) = (\lambda_1 |y_1(j)|^2 + \cdots + \lambda_N |y_N(j)|^2)^{\frac{1}{2}}, \quad \text{for all } j;$$

i.e. the j^{th} coordinate of $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y})$ is the *quadratic mean* of $y_1(j), \ldots, y_N(j)$.

Example 7.6. ("Boltzmann/Shannon" and the geometric mean) Suppose $f(x) = \sum_j x(j)\ln x(j) - x(j)$ on $\operatorname{dom} f = \{x \in E = \mathbb{R}^J : x(j) \geq 0\ \forall j\}$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = \prod_i \{x \in E = \mathbb{R}^J : x(j) > 0,\ \forall j\}$. Then $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y}) = (z, \ldots, z)$, where

$$z(j) = \textstyle\prod_i (y_i(j))^{\lambda_j}, \quad \text{for all } j;$$

i.e. the j^{th} coordinate of $P_{\mathbf{\Delta}}^{\mathbf{f}}(\mathbf{y})$ is the *geometric mean* of $y_1(j), \ldots, y_N(j)$.

Remark 7.7. Example 7.6 is the limiting case $p = 1$ of Example 7.3 in the sense that

$$\lim_{p \downarrow 1} P_{\mathbf{\Delta}}^{p\textbf{-norm}} = P_{\mathbf{\Delta}}^{\textbf{Boltzmann/Shannon}}$$

point-wise on the interior of the domain of **Boltzmann/Shannon**.

The formula for the next example is precisely the one from Example 7.3.

Example 7.8. (p-root and the $(p-1)$-Hölder mean – again!) Suppose $0 < p < 1$ and $f(x) = -\frac{1}{p}\sum_j |x(j)|^p$ on $\operatorname{dom} f = \{x \in E = \mathbb{R}^J : x(j) \geq 0,\ \forall j\}$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = \prod_i \{x \in E = \mathbb{R}^J : x(j) > 0,\ \forall j\}$. Then $P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y}) = (z, \ldots, z)$, where

$$z(j) = (\lambda_1 (y_1(j))^{p-1} + \cdots + \lambda_N (y_N(j))^{p-1})^{\frac{1}{p-1}}, \quad \text{for all } j;$$

i.e. the j^{th} coordinate of $P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y})$ is the $(p-1)$-*Hölder mean* of $y_1(j), \ldots, y_N(j)$.

Example 7.9. ("Burg" and the harmonic mean) Suppose $f(x) = -\sum_j \ln x(j)$ on $\operatorname{dom} f = \{x \in E = \mathbb{R}^J : x(j) > 0,\ \forall j\}$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = \prod_i \{x \in E = \mathbb{R}^J : x(j) > 0,\ \forall j\}$. Then $P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y}) = (z, \ldots, z)$, where

$$z(j) = \frac{1}{\frac{\lambda_1}{y_1(j)} + \cdots + \frac{\lambda_N}{y_N(j)}}, \quad \text{for all } j;$$

i.e. the j^{th} coordinate of $P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y})$ is the *harmonic mean* of $y_1(j), \ldots, y_N(j)$.

Remark 7.10. Example 7.9 can be viewed as the limiting case $p = 0$ of Example 7.8 in the sense that

$$\lim_{p \downarrow 0} P^{\textbf{\textit{p}-root}}_{\boldsymbol{\Delta}} = P^{\textbf{Burg}}_{\boldsymbol{\Delta}}$$

point-wise on the interior of the domain of **Burg**.

Example 7.11. ("Fermi/Dirac") Suppose $f(x) = \sum_j x(j) \ln x(j) + (1-x(j)) \ln(1-x(j))$ on $\operatorname{dom} f = [0,1]^J$. Let $\mathbf{f}(\mathbf{x}) = \sum_i \lambda_i f(x_i)$ and $\mathbf{y} \in \prod_i \operatorname{int}(\operatorname{dom} f) = (]0,1[^J)^N$. Then $P^{\mathbf{f}}_{\boldsymbol{\Delta}}(\mathbf{y}) = (z, \ldots, z)$, where

$$z(j) = \frac{\prod_i (x_i(j))^{\lambda_i}}{\prod_i (x_i(j))^{\lambda_i} + \prod_i (1 - x_i(j))^{\lambda_i}}, \quad \text{for all } j.$$

Remark 7.12.

- It is not too surprising that the geometric mean (resp. the harmonic mean) appears as limiting case of the Hölder mean for $p = 1$ in Remark 7.7 (resp. $p = 0$ in Remark 7.10), since in fact the "Boltzmann/Shannon entropy" (resp. the "Burg entropy") is the "limiting" case of "an affine pertubation" of the p-norm (resp. the p-root) in the sense that

$$\lim_{p\downarrow 1}(t^p - t)/(p-1) - pt = t \ln t - t \quad (\text{resp. } \lim_{p \downarrow 0}(1 - t^p)/p = -\ln t\ ).$$

- The projection onto the diagonal Δ shares "mean"-like properties since the j^{th} coordinate of $P^{\mathbf{f}}_{\Delta}$ is of the form $\varphi^{-1}(\lambda_1 \varphi(a_1) + \cdots + \lambda_N \varphi(a_N))$ (Fact 2.9 and Corollary 7.2), where φ is strictly increasing by the strict convexity of f.
- Examples 7.4, 7.6, and 7.9 also appear in Censor and Reich's Example 4.1 in [12]. However, Example 7.9 is given for purely formal reasons – essentially because the

function $-\ln$ is not Bregman and thus not covered by their framework (see also Remarks 4.2).

7.2. The Hermitian matrices

Lewis recently demonstrated [22, 25, 24, 26] that many parts of classical matrix analysis can be very satisfactorily studied within the framework of convex analysis. This viewpoint has provided numerous insights and examples; a glimpse is provided in this subsection where we assume throughout that

$$\mathcal{H} \text{ is the real vector space of } J \times J \text{ Hermitian matrices.}$$

(Recall that $X \in \mathbb{C}^{J\times J}$ is *Hermitian*, if $X^* := \overline{X^T} = X$.) Denoting the *trace* of a matrix by tr, the vectorspace $\mathcal{H}$ becomes a Hilbert space through

$$\langle X, Y\rangle = \mathrm{tr}(XY^*), \quad \forall X, \forall Y \in \mathcal{H}.$$

For brevitiy, we denote the $J \times J$ permutation matrices by $\mathcal{P}$ and the $J \times J$ unitary matrices by $\mathcal{U}$. (Recall that $U \in \mathbb{C}^{J\times J}$ is *unitary*, if $U^* = U^{-1}$.) Let $\lambda(X) := (\lambda_1(X), \lambda_2(X), \ldots, \lambda_J(X))$ be the *eigenvalues* of X in decreasing order, so that λ is a mapping from $\mathcal{H}$ to $\mathbb{R}^J$. Assume further that

$$f : \mathbb{R}^J \to\,]-\infty, +\infty]$$

is closed, convex, permutation-symmetric, and proper. (f is called *permutation-symmetric*, if $f(Px) = f(x)$, $\forall x \in \mathbb{R}^J$, $\forall P \in \mathcal{P}$.)

Fact 7.13. (Lewis' Corollary 2.7 and Theorem 2.6 in [22])
The induced function

$$f \circ \lambda \quad : \quad \mathcal{H} \to\,]-\infty, +\infty] \quad : \quad X \mapsto f(\lambda(X))$$

is unitarily equivalent, i.e. $(f \circ \lambda)(U^*XU) = (f \circ \lambda)(X)$, $\forall X \in \mathcal{H}$, $\forall U \in \mathcal{U}$. *Moreover,*

$$(f \circ \lambda)^* = f^* \circ \lambda.$$

The punchline is that there is a nice relationship between properties of f and of $f \circ \lambda$. Often, it is enough to study the much simpler function f and still possible to get useful information on $f \circ \lambda$; for instance:

Fact 7.14. (Lewis)

(i) $\mathrm{dom}(f \circ \lambda) = \lambda^{-1}(\mathrm{dom}\, f)$.
(ii) $\mathrm{int}(\mathrm{dom}(f \circ \lambda)) = \lambda^{-1}(\mathrm{int}(\mathrm{dom}\, f))$.
(iii) $f \circ \lambda$ *is Legendre if and only if* f *is.*

Proof. (i) and (ii) follow from Theorem 5.4 in [26].
Corollary 3.3 in [22] in tandem with Corollary 3.5 in [22] implies (iii). □

Of course, we want to know when $f \circ \lambda$ is a Bregman/Legendre function. Theorem 5.6 gives an easy criterion:

Proposition 7.15. *If* $\mathrm{dom}\, f$ *and* $\mathrm{dom}\, f^*$ *are open, then* $f \circ \lambda$ *is Bregman/Legendre.*

Proof. By Fact 7.14.(i),(ii),

$$\operatorname{int}(\operatorname{dom}(f \circ \lambda)) = \lambda^{-1}(\operatorname{int}(\operatorname{dom} f)) = \lambda^{-1}(\operatorname{dom} f) = \operatorname{dom}(f \circ \lambda);$$

hence $\operatorname{dom}(f \circ \lambda)$ is open. Now $(f \circ \lambda)^* = f^* \circ \lambda$; thus (by similar reasoning), $\operatorname{dom}(f \circ \lambda)^*$ is open. Apply Theorem 5.6. $\square$

Example 7.16. (p-norm) Suppose $1 < p < +\infty$ and $f(x) = \sum_j \frac{1}{p}|x_j|^p = \frac{1}{p}\|x\|_p^p$ on $\operatorname{dom} f = \operatorname{int}(\operatorname{dom} f) = E = \mathbb{R}^J$. Then $f \circ \lambda$ is Bregman/Legendre and

$$(f \circ \lambda)(X) = \tfrac{1}{p}\|\lambda(X)\|_p^p = \textstyle\sum_j \tfrac{1}{p}|\lambda_j(X)|^p,$$

for all $X \in \operatorname{dom}(f \circ \lambda) = \mathcal{H}$.

The next example shows that the famous "logarithmic barrier function", also known as Burg's entropy on $\mathcal{H}$, is Bregman/Legendre:

Example 7.17. Suppose $f(x) = \sum_j - \ln x_j$ on $\operatorname{dom} f = \{x \in E : x_j > 0,\ \forall j\}$. Then $f \circ \lambda$ is Bregman/Legendre and

$$(f \circ \lambda)(X) = -\ln \det X,$$

for all $X \in \operatorname{dom}(f \circ \lambda) = \{X \in \mathcal{H} : X \text{ positive semidefinite}\}$.
The best we could possibly hope for would be a result like

"$f \circ \lambda$ is a convex function of Bregman/Legendre type if and only if f is";

this will, however, turn out to be false even for the "Boltzmann/Shannon entropy". For this counter-example, some machinery has to be developed; along the way, we will obtain some interesting positive results.
Given a vector $x = (x_j) \in \mathbb{R}^J$, we write $\Delta(x)$ or Δx for the $J \times J$ diagonal matrix with diagonal entries $x_1, \dots, x_J$. Similarly, given a $J \times J$ matrix X with diagonal entries $X_{11}, \dots, X_{JJ}$, we write $\Delta(X)$ or ΔX for the diagonal matrix with diagonal entries $X_{11}, \dots, X_{JJ}$ or for the vector in $\mathbb{R}^J$ with components $X_{11}, \dots, X_{JJ}$. (It will be clear from the context which object is meant.)
We give a chain of useful facts and omit or comment only briefly on their proofs.

Proposition 7.18. *Suppose $X \in \mathcal{H}$ is diagonal: $X = \Delta(X)$. Then*

$$\langle X, Y \rangle = \langle X, \Delta Y \rangle, \quad \forall Y \in \mathcal{H}.$$

Proposition 7.19. (Lewis)
Suppose f is a Legendre function and $Y \in \operatorname{int}(\operatorname{dom}(f \circ \lambda))$. Then
(i) $\nabla(f \circ \lambda)(Y) = V(\Delta \nabla f(\lambda(Y)))V^*$, $\forall V \in \mathcal{U}$ *with* $V^*YV = \Delta\lambda(Y)$.
(ii) $\nabla(f \circ \lambda)(U^*YU) = U^*\nabla(f \circ \lambda)(Y)U$, $\forall U \in \mathcal{U}$.
(iii) $\langle \nabla(f \circ \lambda)(U^*YU), U^*(X - Y)U \rangle = \langle \nabla(f \circ \lambda)(Y), X - Y \rangle$, $\forall U \in \mathcal{U}$, $\forall X \in \mathcal{H}$.

Proof. (i) follows from Lewis' Corollary 3.3 in [22].
(i) implies (ii), which in turn implies (iii). $\square$

Remark 7.20. If $Y \in \mathrm{int}(\mathrm{dom}(f \circ \lambda))$ is *diagonal*, then so is $\nabla(f \circ \lambda)(Y)$: simply pick V as a permutation matrix in Proposition 7.19.(i).

Corollary 7.21. *$D_{f\circ\lambda}(X,Y) = D_{f\circ\lambda}(U^*XU, U^*YU)$, $\forall X \in \mathcal{H}$, $\forall Y \in \mathrm{int}(\mathrm{dom}(f \circ \lambda))$, $\forall U \in \mathcal{U}$.*

Proof.

$$\begin{aligned} D_{f\circ\lambda} &= (f\circ\lambda)(X) - (f\circ\lambda)(Y) - \langle \nabla(f\circ\lambda)(Y), X - Y\rangle \\ &= (f\circ\lambda)(U^*XU) - (f\circ\lambda)(U^*YU) - \langle \nabla(f\circ\lambda)(U^*YU), U^*(X-Y)U\rangle \\ &= D_{f\circ\lambda}(U^*XU, U^*YU). \end{aligned}$$

□

The last corollary allows reduction to the case when one matrix is diagonal.

Proposition 7.22. (Lewis [23]) *Suppose $X \in \mathcal{H}$ and ΔX is decreasing: $X_{11} \geq X_{22} \geq \cdots \geq X_{JJ}$. Then*

$$\Delta X \in \mathrm{conv}\{P\lambda(X) : P \in \mathcal{P}\}.$$

Proof. Since $\mathcal{P}$ is finite, so is $\{P\lambda(X) : P \in \mathcal{P}\}$. Hence $\mathrm{conv}\{P\lambda(X) : P \in \mathcal{P}\}$ is compact. Assume to the contrary that $\Delta X \notin \mathrm{conv}\{P\lambda(X) : P \in \mathcal{P}\}$. Separation yields $z \in \mathbb{R}^J$ such that $\langle z, \Delta X\rangle > \sup\langle z, \mathrm{conv}\{P\lambda(X) : P \in \mathcal{P}\}\rangle$. Now

$$\begin{aligned} \sup\langle z, \mathrm{conv}\{P\lambda(X) : P \in \mathcal{P}\}\rangle &= \sup\langle z, \{P\lambda(X) : P \in \mathcal{P}\}\rangle \\ &= \sup\{\langle Pz, \lambda(X)\rangle : P \in \mathcal{P}\}. \end{aligned}$$

The last supremum is attained for some $\tilde{P} \in \mathcal{P}$. By Lemma 2.1 in [22], $\tilde{P}z$ has precisely the components of z, *but arranged decreasingly.* Invoking Lemma 2.1 in [22], Proposition 7.18, and a result due to von Neumann (see Theorem 2.2 in [22]), we obtain

$$\begin{aligned} \langle \tilde{P}z, \lambda(X)\rangle &= \sup\langle z, \mathrm{conv}\{P\lambda(X) : P \in \mathcal{P}\}\rangle \\ &< \langle z, \Delta X\rangle \\ &\leq \langle \tilde{P}z, \Delta X\rangle \\ &= \langle \Delta(\tilde{P}z), \Delta(X)\rangle \\ &= \langle \Delta(\tilde{P}z), X\rangle \\ &\leq \langle \lambda(\Delta(\tilde{P}z)), \lambda(X)\rangle \\ &= \langle \tilde{P}z, \lambda(X)\rangle, \end{aligned}$$

which is the desired contradiction. □

Corollary 7.23. *Suppose $C \subseteq \mathbb{R}^J$ is convex and permutation-symmetric: $PC = C$, $\forall P \in \mathcal{P}$. If $Y \in \lambda^{-1}(C)$, then $\Delta Y \in C$.*

Proof. Note that $Y \in \lambda^{-1}(C) \Leftrightarrow \lambda(Y) \in C \Leftrightarrow P\lambda(Y) \in C$, $\forall P \in \mathcal{P}$. Also, there is some $\tilde{P} \in \mathcal{P}$ such that $\Delta(\tilde{P}Y\tilde{P})$ is a decreasing re-arrangement of ΔY. Thus, using Proposition 7.22,

$$\Delta(\tilde{P}Y\tilde{P}) \in \mathrm{conv}\{P\lambda(\tilde{P}Y\tilde{P}) : P \in \mathcal{P}\} = \mathrm{conv}\{P\lambda(Y) : P \in \mathcal{P}\} \subseteq C;$$

hence $\Delta Y \in C$. $\square$

Theorem 7.24. *Suppose $X \in \operatorname{dom}(f \circ \lambda)$, $Y \in \operatorname{int}(\operatorname{dom}(f \circ \lambda))$, and X is diagonal: $X = \Delta X$. Then:*

(i)

$$\begin{aligned} D_{f\circ\lambda}(X,Y) =& D_{f\circ\lambda}(X, \nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y)) \\ &+ (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) \\ &- (f^* \circ \lambda)(\Delta\nabla(f \circ \lambda)(Y)). \end{aligned}$$

(ii)

$$\begin{aligned} D_{f\circ\lambda}(\nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y), Y) =& (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) \\ &- (f^* \circ \lambda)(\Delta\nabla(f \circ \lambda)(Y) \\ =& D_{f^*\circ\lambda}(\nabla(f \circ \lambda)(Y), \Delta\nabla(f \circ \lambda)(Y)). \end{aligned}$$

(iii)

$$\begin{aligned} D_{f\circ\lambda}(X,Y) =& D_{f\circ\lambda}(X, \nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y)) \\ &+ D_{f\circ\lambda}(\nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y), Y). \end{aligned}$$

Proof. $Y \in \operatorname{int}(\operatorname{dom}(f\circ\lambda))$ implies $\nabla(f\circ\lambda)(Y) \in \operatorname{int}(\operatorname{dom}(f\circ\lambda)^*) = \operatorname{int}(\operatorname{dom}(f^*\circ\lambda)) = \lambda^{-1}(\operatorname{int}(\operatorname{dom} f^*))$ by Fact 2.9, Fact 7.13, and Fact 7.14.(ii). Hence $\Delta\nabla(f \circ \lambda)(Y) \in \lambda^{-1}(\operatorname{int}(\operatorname{dom} f^*)) = \operatorname{int}(\operatorname{dom}(f \circ \lambda)^*)$, by Corollary 7.23. Thus $\nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y) \in \operatorname{int}(\operatorname{dom}(f \circ \lambda))$, by Fact 2.9 and Fact 7.13; in other words: all terms appearing in the statement of the theorem make sense. Using Proposition 7.18, we deduce (i):

$$\begin{aligned} D_{f\circ\lambda}(X,Y) &= (f \circ \lambda)(X) + (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - \langle \nabla(f \circ \lambda)(Y), X \rangle \\ &= (f \circ \lambda)(X) + (f^* \circ \lambda)(\Delta\nabla(f \circ \lambda)(Y)) - \langle \Delta\nabla(f \circ \lambda)(Y), X \rangle \\ &\quad + (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - (f^* \circ \lambda)(\Delta\nabla(f \circ \lambda)(Y)) \\ &= D_{f\circ\lambda}(X, \nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y)) \\ &\quad + (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - (f^* \circ \lambda)(\Delta\nabla(f \circ \lambda)(Y)). \end{aligned}$$

By Remark 7.20, the matrix $\nabla(f^* \circ \lambda)\Delta\nabla(f \circ \lambda)(Y)$ is diagonal. Setting X equal to the last matrix yields the first equation of (ii); the second one is just Theorem 3.7.(v). Now (iii) follows. $\square$

Remark 7.25. Theorem 7.24 is quite useful when investigating the method of random projections on the Hermitian matrices, because the nonnegative Bregman distance is further broken up into two nonnegative parts.

"Deconjugating" Theorem 7.24.(ii) yields

Corollary 7.26. *For all $Y \in \operatorname{int}(\operatorname{dom}(f \circ \lambda))$:*

$$D_{f\circ\lambda}(Y, \Delta Y) = (f \circ \lambda)(Y) - (f \circ \lambda)(\Delta Y) \geq 0;$$

equality holds if and only if $Y = \Delta Y$, i.e. Y is diagonal.

We can thus interpret $D_{f\circ\lambda}(Y, \Delta Y)$ as a "measure of non-diagonality" of Y, which turns out to be well-known for certain instances of f:

Example 7.27. (p-trace)

(i) Suppose $1 < p < +\infty$. Then for all $Y \in \mathcal{H}$:

$$\|\lambda(Y)\|_p \geq \|\lambda(\Delta Y)\|_p;$$

equality holds if and only if $Y = \Delta Y$.

(ii) Suppose $0 < p < 1$. Then for all $Y \in \mathcal{H}$ that are positive definite:

$$-\tfrac{1}{p}\textstyle\sum_j (\lambda_j(Y))^p \geq -\tfrac{1}{p}\sum_j Y_{jj}^p;$$

equality holds if and only if $Y = \Delta Y$.

Proof. (i): Consider $f(x) = \sum_j \frac{1}{p}|x_j|^p$ on $\operatorname{dom} f = \mathbb{R}^J$.
(ii): Consider $f(x) = -\sum_j \frac{1}{p}x_j^p$ on $\operatorname{dom} f = \{x \in \mathbb{R}^J : x_j \geq 0,\ \forall j\}$. Note that $\operatorname{int}(\operatorname{dom}(f \circ \lambda)) = \lambda^{-1}(\operatorname{int}(\operatorname{dom} f))$ = the positive definite matrices in $\mathcal{H}$. ◻

Example 7.28. (Hadamard's inequality; see, e.g., Chapter 9, Section B of [27]) For all $Y \in \mathcal{H}$ that are positive definite:

$$\det Y \leq \textstyle\prod_j Y_{jj};$$

equality holds if and only if $Y = \Delta Y$.

Proof. Consider $f(x) = \sum_j -\ln x_j$ on $\operatorname{dom} f = \{x \in \mathbb{R}^J : x_j > 0\ \forall j\}$. Then $(f \circ \lambda)(X) = -\ln\det X$ on the positive definite matrices in $\mathcal{H}$; see Lewis' Section 4 in [22]. ◻

We now give the counter-example announced at the beginning of this subsection.

Example 7.29. ("Boltzmann/Shannon $\circ\lambda$" is not Bregman/Legendre on $\mathcal{H}$)
Suppose $f(x) = \sum_{j=1}^2 x_j \ln x_j - x_j$ on $\operatorname{dom} f = \{x \in \mathbb{R}^2 : x_j \geq 0,\ \forall j\}$. Let

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix};$$

then $B \in \operatorname{dom}(f\circ\lambda)\setminus\operatorname{int}(\operatorname{dom}(f\circ\lambda))$. Suppose further (λ_n), (μ_n), (s_n), (c_n) are sequences of strictly positive real numbers with $\lambda_n \uparrow 1$, $\mu_n < \lambda_n$, $\mu_n \downarrow 0$, $s_n^2+c_n^2 = 1$, $s_n \to 0$, $c_n \to 1$. Let

$$V_n = \begin{pmatrix} c_n & -s_n \\ s_n & c_n \end{pmatrix} \quad\text{and}\quad Y_n = V_n \begin{pmatrix} \lambda_n & 0 \\ 0 & \mu_n \end{pmatrix} V_n^*.$$

Then V_n is unitary, $V_n \to I$, $B_n \in \operatorname{int}(\operatorname{dom}(f \circ \lambda))$ = the positive definite matrices in $\mathcal{H}$, and $B_n \to B$. Also, for $X \in \operatorname{dom}(f \circ \lambda)$, $Y \in \operatorname{int}(\operatorname{dom}(f \circ \lambda))$,

$$\begin{aligned} D_{f\circ\lambda} &= (f \circ \lambda)(X) + (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - \langle\nabla(f \circ \lambda)(Y), X\rangle \\ &= \textstyle\sum_{j=1}^2 \lambda_j(X)\ln\lambda_j(X) - \lambda_j(X) + \sum_{j=1}^2 \lambda_j(Y) - \langle\nabla(f \circ \lambda)(Y), X\rangle. \end{aligned}$$

Thus

$$D_{f\circ\lambda}(0, B_n) = \textstyle\sum_{j=1}^2 \lambda_j(B_n) = \operatorname{tr}(B_n) = \lambda_n + \mu_n \to 1,$$

and all hypotheses of BL2 hold. By Proposition 7.19.(i),

$$\nabla(f\circ\lambda)(B_n) = V_n \begin{pmatrix} \ln\lambda_n & 0 \\ 0 & \ln\mu_n \end{pmatrix} V_n^*$$

$$= \begin{pmatrix} c_n^2\ln\lambda_n + s_n^2\ln\mu_n & c_n s_n(\ln\lambda_n - \ln\mu_n) \\ c_n s_n(\ln\lambda_n - \ln\mu_n) & s_n^2\ln\lambda_n + c_n^2\ln\mu_n \end{pmatrix};$$

so

$$D_{f\circ\lambda}(B,B_n) = -1 + \lambda_n + \mu_n - (c_n^2\ln\lambda_n + s_n^2\ln\mu_n).$$

Thus, for large n,

$$D_{f\circ\lambda}(B,B_n) \approx -s_n^2\ln\mu_n.$$

A posteriori, it is easy to arrange that the sequence $(-s_n^2\ln\mu_n)$ does not converge to 0 (as would be required to satisfy BL2). Take, for instance, $s_n = 1/\sqrt{n}$ and $\mu_n = \exp(-n^2)$; then the sequence $(-s_n^2\ln\mu_n) = (n)$ even tends to $+\infty$.

Remark 7.30. Using Theorem 7.24, one can show that "Boltzmann/Shannon$\circ\lambda$" satisfies BL1; see also Remark 7.25.

Remark 7.31. It is clear that all results in this subsection on unitarily invariant matrix functions defined on the Hermitian matrices have counter-parts for *orthogonally invariant matrix functions defined on the symmetric matrices.*
In particular, Example 7.29 can be interpreted as an example on $\mathbb{R}^3$, since $\mathbb{R}^3$ and the symmetric 2×2 matrices are isomorphic. This opens another avenue for constructing interesting non-separable convex functions. Specifically, Example 7.29 translates to the following:

Example 7.32. $\mathbb{R}^3$ is isomorphic to the symmetric 2×2 matrices via

$$Tx = T(x_1,x_2,x_3) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix}, \qquad \forall x = (x_1,x_2,x_3)\in\mathbb{R}^3.$$

The matrix Tx is positive semi-definite if and only if $x_1, x_3 \geq 0$ and $x_1x_3 - x_2^2 \geq 0$. The eigenvalues of Tx in decreasing order are

$$\lambda(Tx) = \tfrac{1}{2}(x_1 + x_3 + \sqrt{(x_1-x_3)^2 + 4x_2^2}, x_1 + x_3 - \sqrt{(x_1-x_3)^2+4x_2^2}).$$

Now let $f(y) = y_1\ln y_1 - y_1 + y_2\ln y_2 - y_2$ on $\mathbb{R}^2$. Then $F = f\circ\lambda\circ T$ is a Legendre function but F is not Bregman/Legendre. For instance, this interpretation makes the convexity of F a triviality. On the other hand, it seems not to be easy to check convexity directly. Also, this example induces the counter-example announced in Remark 3.4.
We conclude this section with a truly matrix-based example.

Example 7.33. (Doubly stochastic constraints) Let $\mathcal{S}$ be the real Hilbert space of $J\times J$ symmetric matrices with $\langle X,Y\rangle = \mathrm{tr}(XY^T)$, $\forall X, \forall Y\in\mathcal{S}$ (see Remark 7.31). The elements of the i^{th} row (or column) of a given matrix $Y \in \mathcal{S}$ add up to 1 exactly when

$$\langle A_i, Y \rangle = 1, \qquad \text{where} \quad (A_i)_{mn} = \begin{cases} 1, & \text{if } m = n = i; \\ \frac{1}{2}, & \text{if } m \neq n \text{ and } i \in \{m, n\}; \\ 0, & \text{otherwise.} \end{cases}$$

Hence Y is doubly-stochastic, if $Y \in \bigcap_i H_i$, where $H_i = \{X \in \mathcal{S} : \langle A_i, X \rangle = 1\}$. The orthogonal projection (i.e. the Bregman projection w.r.t. $\frac{1}{2}\|\cdot\|^2$) is explicitly given through

$$P_{H_i} Y = Y - \frac{\sum_j Y_{ij} - 1}{(J+1)/2} A_i.$$

Bregman projections with respect to other (Bregman/)Legendre functions can be approximated by the procedure described in Remark 6.13.

8. The method of random Bregman projections

It is convenient to abbreviate the following assumption (A):

$$\begin{gathered} f \text{ is a Legendre function on } E, \\ C_1, \ldots, C_N \text{ are closed convex sets with } \bigcap_i C_i \cap \operatorname{dom} f \neq \emptyset, \\ C_i \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset, \quad \forall i. \end{gathered} \tag{A}$$

Note that it may happen that $\bigcap_i C_i \cap \operatorname{int}(\operatorname{dom} f) = \emptyset$.

Let r be a *random mapping for* $\{1, \ldots, N\}$, i.e. a mapping from $\mathbb{N}$ onto $\{1, \ldots, N\}$ that takes each value in $\{1, \ldots, N\}$ infinitely often.

The *method of random Bregman projections* generates a sequence (y_n) by

$$y_0 \in \operatorname{int}(\operatorname{dom} f) \quad \text{and} \quad y_{n+1} := P_{C_{r(n+1)}} y_n, \quad \forall n \geq 0. \tag{M}$$

For instance, we could consider the random function $r(n) \equiv n \bmod N$ (where we let the $\bmod N$ function take values in $\{1, \ldots, N\}$) and thus obtain the well-known method of cyclic Bregman projections. We refer to the point y_0 as the *starting point.*

In view of Theorem 3.14, assumption (A) guarantees that the sequence generated by method (M) lies in $\operatorname{int}(\operatorname{dom} f)$ and is thus well-defined (the interiority/zone consistency condition!).

Theorem 8.1. *Suppose assumption* (A) *and (at least) one of the following conditions hold:*

(i) *f is a Bregman/Legendre function.*

(ii) $\bigcap_i C_i \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ *and* $\operatorname{dom} f^*$ *is open.*

(iii) $\operatorname{dom} f$ *and* $\operatorname{dom} f^*$ *are open.*

Then for an arbitrary starting point $y_0 \in \operatorname{int}(\operatorname{dom} f)$*, the sequence* (y_n) *generated by method* (M) *converges to some point* y *in* $\bigcap_i C_i \cap \operatorname{dom} f$ *and* $D_f(y, y_n) \to 0$*. If* (ii) *or* (iii) *holds, then* y *actually belongs to* $\bigcap_i C_i \cap \operatorname{int}(\operatorname{dom} f)$.

Proof. (i): Proposition 3.16 yields

1. $D_f(y_{n+1}, y_n) \leq D_f(c, y_n) - D_f(c, y_{n+1}), \quad \forall n \geq 0,\ \forall c \in C_{r(n+1)} \cap \operatorname{dom} f.$

Hence

2. $D_f(y_{n+1}, y_n) \leq D_f(c, y_n) - D_f(c, y_{n+1}), \quad \forall n \geq 0,\ \forall c \in \bigcap_i C_i \cap \operatorname{dom} f.$

Fix any $c \in \bigcap_i C_i \cap \operatorname{dom} f$ and observe that $(D_f(c, y_n))$ is decreasing and hence bounded. By Corollary 3.11 (i.e. BL0) or BL1 (depending on the location of c), the sequence

3. (y_n) is bounded.

Suppose now that $\bar{y}$ is an arbitrary cluster point of (y_n), say $y_{k_n} \to \bar{y}$. Step by step, we collect properties on $\bar{y}$.

Property 1: $\bar{y} \in \operatorname{dom} f$ and $D_f(\bar{y}, y_{k_n}) \to 0$.

If $\bar{y} \in \operatorname{int}(\operatorname{dom} f)$, then apply Proposition 3.2.(ii). If $\bigcap_i C_i \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, then pick any c in this intersection; Property 1 then follows from 2 and Theorem 3.8.(ii). Otherwise, $\bar{y} \in \operatorname{bd}(\operatorname{dom} f)$ and we make use of BL2. Property 1 is established for all cases.

Property 2: $\bar{y} \in \bigcap_i C_i$.

We can assume without loss that $r(k_n) \equiv \rho \in \{1, \ldots, N\}$ (after passing to a subsequence if necessary). Let's define

$$I_{\text{in}} := \{i \in \{1, \ldots, N\} : \bar{y} \in C_i\} \quad \text{and} \quad I_{\text{out}} := \{1, \ldots, N\} \setminus I_{\text{in}}.$$

We want to show that $I_{\text{out}} = \emptyset$.

So let us assume to the contrary that $I_{\text{out}} \neq \emptyset$. Since r is a random mapping, we can also assume (subsequence!) that $\{r(k_n), r(k_n + 1), \ldots, r(k_{n+1} - 1)\} = \{1, \ldots, N\}$. For every n, pick m_n maximal in $\{k_n, k_n + 1, \ldots, k_{n+1} - 1\}$ such that $r(m_n) \in I_{\text{in}}$. This is possible, since $\rho \in I_{\text{in}}$ and I_{out} is assumed to be nonempty. Then, by definition of m_n, for every $k_n \leq \nu \leq m_n$, $r(\nu) \in I_{\text{in}}$; hence, by using 1 successively, $D_f(\bar{y}, y_{m_n}) \leq D_f(\bar{y}, y_{k_n})$. It follows with Property 1 that

3. $D_f(\bar{y}, y_{m_n}) \to 0.$

Claim 1: $y_{m_n} \to \bar{y}$.

We can assume without loss (subsequence!) that $y_{m_n} \to \bar{z} \in \operatorname{dom} f$ with $D_f(\bar{z}, y_{m_n}) \to 0$ (Property 1). If $\{\bar{y}, \bar{z}\} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, then 3 and Theorem 3.9.(iii) imply $\bar{y} = \bar{z}$. Otherwise $\bar{y}, \bar{z} \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$ and Proposition 5.5 applies. Claim 1 thus holds.

After passing to yet another subsequence if necessary, we assume without loss that $r(m_n + 1) \equiv i$, for some $i \in I_{\text{out}}$; and that $y_{m_n+1} \to \bar{z} \in \operatorname{dom} f \cap C_i$. Note that by 2,

4. $D_f(y_{m_n+1}, y_{m_n}) \to 0.$

Claim 2: $\bar{y} = \bar{z}$.

If $\bar{y}$ or $\bar{z}$ is in the interior of $\operatorname{dom} f$, then use 4 and Theorem 3.9.(iii); otherwise, use 4 and BL3. Claim 2 is verified.

Claim 2 now yields the contradiction $i \in I_{\text{in}} \cap I_{\text{out}}$. Consequently, Property 2 does hold. Property 1, Property 2, and 1 imply

Property 3: $D_f(\bar{y}, y_n) \to 0$.

It remains to show that the entire sequence converges to $\bar{y}$. Let $\bar{z}$ be a (possibly different) cluster point of (y_n), say $y_{l_n} \to \bar{z}$. By Properties 1 through 3 (for (y_{l_n}) and $\bar{z}$),

5. $\quad \bar{z} \in \bigcap_i C_i \cap \operatorname{dom} f \quad$ and $\quad D_f(\bar{z}, y_n) \to 0$.

If $\{\bar{y}, \bar{z}\} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, then $\bar{y} = \bar{z}$ by Theorem 3.9.(iii). Otherwise, $\bar{y}, \bar{z} \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$ and $D_f(\bar{y}, y_{l_n}) \to 0$ (Property 3). Then Proposition 5.5 applies and yields $\bar{y} = \bar{z}$. The proof for (i) is complete.
(ii): is proved as (i), with the exceptions that the stronger $\bar{y} \in \operatorname{int}(\operatorname{dom} f)$ is derived and that BL1 through BL3 are not needed.
(iii): is then clear, since (iii) implies (ii). □

Remark 8.2. Let us see how the conditions (i), (ii), and (iii) are related.

- "(iii)⇒(ii)" and "(iii)⇒(i)" but not vice versa: this follows from assumption (A), Theorem 5.6, and the example $x \ln x - x$ on $E = \mathbb{R} = C_1$ with $N = 1$.
- On the real line, "(i)⇔(ii)": use assumption (A) and a direct interval argument for proving "⇒" and Theorem 5.8 for "⇐".
- In general, (i) and (ii) are independent: if $N = 1$, $E = C_1 = \mathbb{R}^2$, and f is as in Example 6.9, then (ii) holds but (i) does not. And if $N = 2$, $f(x) = \sum_j x_j \ln x_j - x_j$ on $E = \mathbb{R}^2$, then (i) holds but (ii) fails.

To summarize, we can say that the method of random Bregman projections always works for two quite general situations:

- the function f is Bregman/Legendre – which is easy to check for separable functions.
- the constraint qualification $\bigcap_i C_i \cap \operatorname{dom} f \neq \emptyset$ holds and f is just Legendre with $\operatorname{dom} f^*$ open. Here, the conditions on f are readily verifiable whereas the constraint qualification requires some a priori knowledge.

Remark 8.3.

- The proof of Theorem 8.1 is an extension of the proof of the authors' (Theorem 3.10 of [3]). We want to remark that the latter proof in turn relies on an idea developed almost simultaneously by Flåm and Zowe [16], by Tseng [32], and by Elsner et al. [15] around the beginning of the decade; this idea is also present in Censor and Reich's analysis [12].
- Though similar, the assumptions in Censor and Reich's work [12] differ from ours: they allow one to draw operators from a possibly infinite pool whereas our underlying distance function is more general. For instance, "Burg's entropy", $-\ln$, is excluded in [12] but included here.
- It is clear that Theorem 8.1, Theorem 5.12, and the results in the first half of Section 7 imply convergence results for simultaneous Bregman projection methods. This procedure is straight-forward; the development follows along the lines of Censor and Elfving work [6]. Our present analysis has the advantage that interiority/zone consistency of the Bregman projections in the product space is guaranteed automatically.
- A pointer to some references on the method of random *orthogonal* projections (in Hilbert space) is the first author's [2].

- In August 1994, during the Mathematical Programming conference in Ann Arbor, K. Kiwiel announced results that appear to be related to Theorem 8.1.

We can say more on the limit in the important case of hyperplanes:

Theorem 8.4. *Suppose assumption* (A) *holds for hyperplanes* $C_i = \{x \in E : \langle a_i, x\rangle = b_i\}$, *where* $a_i \in E \setminus \{0\}$, $b_i \in \mathbb{R}$. *Suppose also (at least) one of the following conditions is satisfied:*
(i) f *is a Bregman/Legendre function.*
(ii) $\bigcap_i C_i \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ *and* $\operatorname{dom} f^*$ *is open.*
(iii) $\operatorname{dom} f$ *and* $\operatorname{dom} f^*$ *are open.*
Suppose further the set $\{z \in E : \nabla f(z) \in \operatorname{span}(a_1, \ldots, a_N)\}$ *is nonempty and* y_0 *is an arbitrary element of it. Then the sequence* (y_n) *generated by method* (M) *with starting point* y_0 *converges to some point in*

$$\operatorname*{argmin}_{x \in \operatorname{dom} f \cap \bigcap_i C_i} f(x).$$

In case of (ii) *or* (iii), *the argmin is singleton and an element of* $\operatorname{int}(\operatorname{dom} f)$.

Proof. Theorem 8.1 yields the convergence of the sequence (y_n) to some point y in $\operatorname{dom} f \cap \bigcap_i C_i$ with $D_f(y, y_n) \to 0$. Define $A : E \to \mathbb{R}^N : x \mapsto (\langle a_i, x\rangle)_i$. Then, by Corollary 6.12, the entire sequence (y_n) belongs to $Z := \{z \in E : \nabla f(z) \in \operatorname{span}(a_1, \ldots, a_N)\} = \{z \in E : \nabla f(z) \in \operatorname{range} A^*\}$. Now fix an arbitrary element x in $\bigcap_i C_i \cap \operatorname{dom} f$. Then $y - x \in \operatorname{kernel} A = (\operatorname{range} A^*)^{\perp}$ so that

$$\begin{aligned} f(y) - f(x) &= f(y) - f(x) - \langle \nabla f(y_n), y - x\rangle \\ &= D_f(y, y_n) - D_f(x, y_n) \\ &\leq D_f(y, y_n) \\ &\to 0. \end{aligned}$$

Thus y is contained in the argmin. Finally, if (ii) or (iii) holds, then $y \in \operatorname{int}(\operatorname{dom} f)$ by Theorem 8.1. But f is strictly convex on $\operatorname{int}(\operatorname{dom} f)$ and the argmin is therefore singleton. $\square$

Remark 8.5.

- Theorem 8.4 remains true if we replace “hyperplane” by “affine subspace”. The proof is the same, but notationally much less convenient. Classical results on the method of orthogonal projections are obtained for the choice $f(x) = \frac{1}{2}\|x\|^2$; for more, see [3] and the references therein.
- Assuming the hypothesis of Theorem 8.4, suppose the argmin $_{x \in \operatorname{dom} f} f(x)$ is actually contained in $\operatorname{int}(\operatorname{dom} f)$. By essential strict convexity of f, the argmin is singleton, say x_0. We now can simply choose $y_0 = x_0$ as starting point for the method (M), because $\nabla f(y_0) = 0 \in \operatorname{span}(a_1, \ldots, a_N)$. The following table contains a selection of Legendre functions on $\mathbb{R}$ for which this technique applies; of course, this extends to separable Legendre functions.

$f(x)$	$\underset{x \in \operatorname{dom} f}{\operatorname{argmin}} f(x)$
$\frac{1}{2}\lvert x\rvert^2$	0
$x \ln x - x$	1
$x \ln x + (1-x)\ln(1-x)$	$\frac{1}{2}$
$\frac{1}{p}\lvert x\rvert^p$	0
$-\sqrt{1-x^2}$	0

- Related though somewhat different to Theorem 8.4 are the following:
  - •• Bregman's [4] (cyclic control).
  - •• Censor and Lent's [7] (cyclic control).
  - •• De Pierro and Iusem's [29] (cyclic control, but "relaxed" projections).
  - •• Censor and Lent's [8] (cyclic control for "Burg's entropy", $-\ln$, with "bare hands").
  - •• Censor and Reich's [12] (more general operators but the function that induces the Bregman distance has to be Bregman; see the Remarks in Section 4).

Acknowledgment. It is our pleasure to thank Adrian Lewis for extremely helpful discussions on the subsection on Hermitian matrices in Section 7.2, for the proof of Proposition 7.22, and for sending us [22, 25, 24, 26]. We wish to thank Regina Burachik, Dan Butnariu, Yair Censor, Alfredo Iusem, and Krzysztof Kiwiel for providing us with [5], [1], [6], [19], and [21]. Thanks are also due to two anonymous referees who made several valuable comments and suggestions; one referee pointed out that Censor et al. considered in [11], [10] slight extensions of Bregman functions which do include Burg's entropy.

References

[1] Y. Alber and D. Butnariu: Convergence of Bregman-projection methods for solving consistent convex feasibility problems in reflexive Banach spaces. Journal of Optimization Theory and Applications (1994) Accepted for Publication.

[2] H.H. Bauschke: A norm convergence result on random products of relaxed projections in Hilbert space. Transactions of the American Mathematical Society, 347(4) (1995) 1365–1374.

[3] H.H. Bauschke and J.M. Borwein: On projection algorithms for solving convex feasibility problems. SIAM Review 38(3) (1996) 367–426.

[4] 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 7(3) (1967) 200–217.

[5] R.S. Burachik: Generalized Proximal Point Methods of the Variational Inequality Problem. PhD thesis, Rio de Janeiro, 1995.

[6] Y. Censor and T. Elfving: A multiprojection algorithm using Bregman projections in a product space. Numerical Algorithms 8 (1994) 221–239.

[7] Y. Censor and A. Lent: An iterative row-action method for interval convex programming. Journal of Optimization Theory and Applications 34(3) (1981) 321–353.

[8] Y. Censor and A. Lent: Optimization of "$\log x$" entropy over linear equality constraints. SIAM Journal on Control and Optimization 25(4) (1987) 921–933.

[9] Y. Censor, A.R. de Pierro, T. Elfving, G.T. Herman, and A.N. Iusem: On iterative methods for linearly constrained entropy maximization. In A. Wakulicz, editor, Numerical Analysis and Mathematical Modelling, pages 145–163, Warsaw, 1990. Polish Scientific Publishers. Banach Center Publications Volume 24.

[10] Y. Censor, A.R. de Pierro, and A.N. Iusem: On maximization of entropies and a generalization of Bregman's method for convex programming. Technical Report MIPG 113, Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, U.S.A., 1986.

[11] Y. Censor, A.R. de Pierro, and A.N. Iusem: Optimization of Burg's entropy over linear constraints. Applied Numerical Mathematics 7 (1991) 151–165.

[12] Y. Censor and S. Reich: Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization. Optimization 37(4) (1996) 323–339.

[13] G. Chen and M. Teboulle: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM Journal on Optimization 3(3) (1993) 538–543.

[14] J. Eckstein: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Mathematics of Operations Research 18(1) (1993) 202–226.

[15] L. Elsner, I. Koltracht, and M. Neumann: Convergence of sequential and asynchronous nonlinear paracontractions. Numerische Mathematik 62 (1992) 305–319.

[16] S.D. Flåm and J. Zowe: Relaxed outer projections, weighted averages and convex feasibility. BIT 30 (1990) 289–300.

[17] J.-B. Hiriart-Urruty and C. Lemaréchal: Convex Analysis and Minimization Algorithms II, volume 306 of Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1993.

[18] A.N. Iusem, 1994: Personal communication.

[19] A.N. Iusem: Proximal point methods in optimization, May 1994. Unpublished manuscript.

[20] K.C. Kiwiel: conference talk in Ann Arbor, Michigan, August 1994.

[21] K.C. Kiwiel: Free-steering relaxation methods for problems with strictly convex costs and linear constraints. WP-94-89, International Institute for Applied Systems Analysis, Laxenburg, Austria, September 1994. Revised July 1995.

[22] A.S. Lewis: Convex analysis on the Hermitian matrices. SIAM Journal on Optimization 6(1) (1996) 164–177.

[23] A.S. Lewis, 1994: Personal communication.

[24] A.S. Lewis: The convex analysis of unitarily invariant matrix norms. Journal of Convex Analysis 2 (1995) 173–183.

[25] A.S. Lewis: Derivatives of spectral functions. Mathematics of Operations Research 21(3) (1996) 576–588.

[26] A.S. Lewis: Group invariance and convex matrix analysis. SIAM Journal on Matrix Analysis 17(4) (1996) 927–949.

[27] A.W. Marshall and I. Olkin: Inequalities: Theory of Majorization and Its Applications, volume 143 of Mathematics in Science and Engineering. Academic Press, New York, 1979.

[28] G. Pierra: Decomposition through formalization in a product space. Mathematical Programming 28 (1984) 96–115.

[29] A.R. de Pierro and A.N. Iusem: A relaxed version of Bregman's method for convex programming. Journal of Optimization Theory and Applications 51(3) (1986) 421–440.

[30] R.T. Rockafellar: Convex Analysis. Princeton University Press, Princeton, NJ, 1970.

[31] M. Teboulle: Entropic proximal mappings with applications to nonlinear programming. Mathematics of Operations Research 17(3) (1992) 670–690.

[32] P. Tseng: On the convergence of the products of firmly nonexpansive mappings. SIAM Journal on Optimization 2(3) (1992) 425–434.

HIER :

Leere Seite
68