Legendre functions and the method of random Bregman projections

Heinz H. Bauschke* and Jonathan M. Borwein†
Centre for Experimental & Constructive Mathematics,
Simon Fraser University, Burnaby, B.C., Canada V5A 1S6

March 2, 1995

Abstract

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, now days 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.

1991 M.R. Subject Classification.
Primary 49M45; Secondary 47H15, 52A41, 65F10, 65K05, 90C25.

Key words and phrases. Bregman function, Bregman projection, convex feasibility problem, convex function, convex set, essentially smooth function, essentially strictly convex function, Hermitian matrix, Legendre function, projection.

*Electronic mail: bauschke@cem.sfu.ca.
†Electronic mail: jborwein@cem.sfu.ca. Research supported by NSERC and by the Shrum Endowment.
1 Introduction

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

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

Typically, the points in the intersection are the sought-after solutions of a given problem and the sets $C_1, \ldots, C_N$ 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 $r$th constraint set by $P_r$, 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

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

The sequence $(y_n)$ converges to a solution if the underlying space is some Euclidean space $\mathbb{R}^I$ 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 Iusen [27]. Further progress was made within last year: Censor and Reich ([10, Theorem 3.2]) 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 [18] 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 Teboule’s [11], Eckstein’s [12], Iusen’s [17], and Teboule’s [29]. 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 [25], which has become a standard tool in the field. Following Censor, Elfving, and Reich [6, 10], 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 [19, 23, 21, 22] – 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) “Breg 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 [10] and partially generalize earlier work [4, 6, 7, 27]. Whereas Censor and Reich allow operators more general than Bregman projections, we build on the class of Bregman/Legendre functions: again, the important “Breg entropy” is included in our analysis but excluded from the class of Bregman functions.

Throughout the paper, we assume that

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

Almost all the facts we use from Convex Analysis can be found in Rockafellar’s fundamental book [28]. 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 \( \text{dom} f \) (\( f^* \), \( f^0 \), \( \nabla f \), \( \partial f \), \( f \square g \), respectively). The indicator function of a set \( C \) is denoted \( 1_C \) and its interior (boundary, closure, respectively) is abbreviated by \( \text{int} C \) (\( \text{bd} C \), \( \text{cl} C \), respectively). Finally, \( I \) stands for the identity mapping or identity matrix and, for sequences, the symbol “\( \rightarrow \)” indicates convergence.

2 Tools

The concepts in this section are fundamental to our analysis.
**Essential smoothness**

**Definition 2.1** (Rockafellar’s [28, Section 26]) Suppose \( f \) is a closed convex proper function on \( E \) with \( \text{int}(\text{dom} f) \neq \emptyset \). Then \( f \) is **essentially smooth**, if \( f \) is differentiable on \( \text{int}(\text{dom} f) \) and

\[
\{ (x_n) \text{ in } \text{int}(\text{dom} f), \quad x_n \to x \in \text{bd}(\text{dom} f) \} \Rightarrow ||\nabla f(x_n)|| \to +\infty.
\]

**Fact 2.2** (Rockafellar’s [28, Theorem 26.1 and Lemma 26.2]) Suppose \( f \) is closed convex proper on \( E \). Then the following are equivalent.

1. \( f \) is essentially smooth.
2. \( f \) is differentiable on \( \text{int}(\text{dom} f) \) and

\[
\lim_{t \downarrow 0} (\nabla f(x + tf(y - x)), y - x) = -\infty, \quad \forall x \in \text{bd}(\text{dom} f), \forall y \in \text{int}(\text{dom} f).
\]
3. \( \partial f(x) = \emptyset, \forall x \in \text{bd}(\text{dom} f) \), and \( \partial f(y) = \{ \nabla f(y) \}, \forall y \in \text{int}(\text{dom} f) \).

**Essential strict convexity**

**Definition 2.3** (Rockafellar’s [28, Section 26]) Suppose \( f \) is closed convex proper on \( E \). Then \( f \) is **essentially strictly convex**, if \( f \) is strictly convex on every convex subset of \( \text{dom} \partial f \).

**Fact 2.4** (Rockafellar’s [28, Theorem 26.3]) Suppose \( f \) is closed convex proper on \( E \). Then the following are equivalent.

1. \( f \) is essentially strictly convex.
2. \( f^* \) is essentially smooth.
3. \( \partial f(x) \cap \partial f(y) = \emptyset, \forall x, y \in \text{dom} f, x \neq y \).

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.

1. \( f \) is essentially strictly convex.
2. \( \text{range} \partial f = \text{int}(\text{dom} f^*) = \text{dom} \nabla f^* \).
3. \( \{ y \} = \partial f^*(\partial f(y)), \forall y \in \text{dom} \partial f \).

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

\[
\nabla f(y) \in \text{int}(\text{dom} f^*) \quad \text{and} \quad \nabla f^* \nabla f(y) = y, \quad \forall y \in \text{int}(\text{dom} f).
\]

4
The following example shows that range of $f$ need not be equal to $\text{int}(\text{dom} f^\circ)$:

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

$$\nabla f(\text{int}(\text{dom} f)) = I([0, +\infty]) = ]0, +\infty[ \subseteq \mathbb{R} = \text{int}(\text{dom} f^\circ).$$

**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 [28, Section 26]) 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. $\text{int}(\text{dom} f) \neq \emptyset$.
- L1. $f$ is differentiable on $\text{int}(\text{dom} f)$.
- L2. $\lim_{t \downarrow 0} (\nabla f(x + t(y - x)), y - x) = -\infty$, $\forall x \in \text{bd}(\text{dom} f)$, $\forall y \in \text{int}(\text{dom} f)$.
- L3. $f$ is strictly convex on $\text{int}(\text{dom} f)$.

**Fact 2.9** (Rockafellar’s [28, Theorem 26.5]) A convex function $f$ is of Legendre type, if and only if its conjugate $f^* \circ$ is. In this case, the gradient mapping

$$\nabla f : \text{int}(\text{dom} f) \to \text{int}(\text{dom} f^\circ) : x \mapsto \nabla f(x)$$

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

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.

**Coercivities**

**Definition 2.10** A function $f$ on $E$ is **coercive** (or **0-coercive**; see [15, Definition X.1.3.7]) 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 [28, Corollary 14.2.2]) Suppose $f$ is closed convex proper on $E$ and $x^* \in E$. Then $x^* \in \text{int}(\text{dom} f^\circ)$ 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 [28, Theorem 8.5] for the definition).

Fact 2.13 (Rockafellar’s [28, Corollary 13.3.4.(c)]) Suppose \( f \) is closed convex proper on \( E \). Then
\[
\text{int}(\text{dom} f^*) = \{ x^* \in E : \langle x^*, e \rangle < (f^{0+})(e), \forall e \in E \setminus \{0\} \}.
\]

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 \( \lim_n f(x_n)/\|x_n\| = (f^{0+})(\lim_n x_n/\|x_n\|). \)

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 \( \hat{x} \in \text{dom} f \) 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 \( \rho \). Then \( \|d_n\| \geq \rho \) eventually; hence
\[
L = \lim_n \frac{f(\hat{x} + \|d_n\|/\|x_n\| \cdot \frac{d_n}{\|d_n\|}) - f(\hat{x})}{\|d_n\|} \geq \lim_n \frac{f(\hat{x} + \rho \cdot \frac{d_n}{\|d_n\|}) - f(\hat{x})}{\rho} \geq \frac{f(\hat{x} + \rho q) - f(\hat{x})}{\rho}.
\]

Now let tend \( \rho \) to \( +\infty \) to conclude \( L \geq (f^{0+})(q) \). \( \square \)

Definition 2.15 A function \( f \) on \( E \) is called super-coercive (or 1-coercive; see [15, Definition X.1.3.7]), if
\[
\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty.
\]

Evidently, any super-coercive function — for instance, \( \frac{1}{\|x\|} \cdot \|x\|^2 \) — 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 [28, Section 13]), i.e. \((f^{0+})(e) = +\infty, \forall e \neq 0\); equivalently,
\[
\lim_{\lambda \to +\infty} \frac{f(\lambda e)}{\lambda} = +\infty, \quad \forall e \neq 0.
\]
(iii) \( \text{dom} f^* = E \).
(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 \( M \geq (f^{0+})(e) \), which contradicts the co-finiteness of \( f \).
“(ii)⇔(iii)”: is Rockafellar’s [28, Corollary 13.3.1].
“(iii)⇔(iv)”: clear from Fact 2.11. \( \square \)

3 Bregman distances and Bregman projections

Bregman distances

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

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

Although commonly used, the term “distance” is misleading: neither is \( D_f \) symmetric (unless \( f \) is quadratic, as Lusen [16] proved) nor does \( D_f \) satisfy the triangle inequality (check \( \|\cdot\|_2^2 \)). 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 \( \text{int}(\text{dom} f) \neq \emptyset \). If \( f \) is differentiable on \( \text{int}(\text{dom} f) \), then:

(i) \( D_f(x, y) = f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x \rangle, \forall x \in E, \forall y \in \text{int}(\text{dom} f) \).

(ii) \( \{y_n\} \text{ in } \text{int}(\text{dom} f), y_n \to y \in \text{int}(\text{dom} f) \} \Rightarrow D_f(y, y_n) \to 0. \)

**Proof.** (i): By [28, Theorem 23.5], \( f(y) + f^*(\nabla f(y)) = \langle \nabla f(y), y \rangle \); now substitute.
(ii): This is clear from (i), [28, Corollary 25.5.1 and Theorem 23.5]. \( \square \)

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

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

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

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

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

Now apply [28, Corollary 7.5.1]. \qed

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 $\text{int}(\text{dom} f_1), \ldots, \text{int}(\text{dom} f_J) \neq \emptyset$, respectively. If every $f_j$ is differentiable on $\text{int}(\text{dom} f_j)$ and if $\lambda_1, \ldots, \lambda_J$ are strictly positive real numbers, then

$$f : E := \prod_j E_j \to ]-\infty, +\infty[ : x := (x_1, \ldots, x_J) \mapsto \sum_j \lambda_j f_j(x_j)$$

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

$$D_f(x, y) = \sum_j \lambda_j D_{f_j}(x_j, y_j), \quad \forall x \in E, \forall y \in \text{int}(\text{dom} f).$$

The following proposition will be useful later.

Proposition 3.6 Suppose $f$ is Legendre on $\mathbb{R}$ and $\text{dom} f = [a, b]$, where $-\infty < a < b \leq +\infty$. If $(y_n)$ is a sequence in $\text{int}(\text{dom} f)$ 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

$$D_f(a, y_n) = f(a) - f(y_n) + \nabla f(y_n)(y_n - a)$$

$$\geq f(a) - f(y_n) + \frac{f(y_n) - f(y)}{y_n - y}(y_n - a)$$

$$= \left\{ f(a) - f(y) \frac{y_n - a}{y_n - y}\right\} + f(y_n) \left\{ \frac{y_n - a}{y_n - y}\right\}. $$

8
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 \(b \not\in \text{dom} f\). 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
\[
D_f(a, y_n) = f(a) - f(y_n) + \nabla f(y_n)(y_n - a)
\geq f(a) - f(y_n) + \frac{f(y_n) - f(y)}{y_n - y}(y_n - a)
= \left\{ f(a) - f(y) \frac{y_n - a}{y_n - y} \right\} + \frac{f(y_n)}{y_n - a} \left\{ \frac{y_n - a}{y_n - y} (y - a) \right\}.
\]
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 \(\text{int}(\text{dom} f) \neq \emptyset\), and essentially strictly convex.
Suppose further \(y \in \text{int}(\text{dom} f)\). Then:

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

Proof. (i): is clear by Proposition 3.2.(i).
(ii),(iii): \(f\) essentially strictly convex \(\Rightarrow \nabla f(y) \in \text{int}(\text{dom} f^*)\) (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 \(D_f(x, y) = 0 \iff f(x) + f^*(\nabla f(y)) = \langle \nabla f(y), x \rangle \Leftrightarrow x = \nabla f^*(\nabla f(y)) \Leftrightarrow x = y\) (Corollary 2.6).
(v): \(D_f(\nabla f(y), \nabla f(x)) = f^*(\nabla f(y)) + f^*\langle \nabla f^*(\nabla f(x)) - \langle \nabla f^*(\nabla f(x)), \nabla f(y) \rangle = f^*(\nabla f(y)) + f(x) - \langle x, \nabla f(y) \rangle = D_f(x, y)\).
(vi): Suppose $x \in \text{int}(\text{dom} f)$, $(y_n)$ in $\text{int}(\text{dom} f)$ with $(D_f(x, y_n)) = (f(x) + f^*(\langle \nabla f(y_n), \cdot \rangle))$ bounded. Since $x \in \text{int}(\text{dom} f)$, the function $f^*(\cdot) - \langle \cdot, \cdot \rangle$ is coercive. Hence $(\nabla f(y_n))$ is bounded. $f^*$ is closed, thus all cluster points of $(\nabla f(y_n))$ lie in $\text{dom} f^* = \text{int}(\text{dom} f^*)$. It follows that $(y_n) = (\nabla f^*(\nabla f(y_n)))$ is bounded, too (Corollary 2.6 and [28, Corollary 25.5.1]). \hfill \Box

**Theorem 3.8** ("essential smoothness helps") Suppose $f$ is closed convex proper on $E$, differentiable on int(\text{dom} f) ≠ 0, and essentially smooth. Then:

(i) \[ x \in \text{int}(\text{dom} f), (y_n) \text{ in } \text{int}(\text{dom} f), y_n \rightarrow y \in \text{bd}(\text{dom} f) \Rightarrow D_f(x, y_n) \rightarrow +\infty. \]

(ii) \[ x \in \text{int}(\text{dom} f), (y_n) \text{ in } \text{int}(\text{dom} f), y_n \rightarrow y \in \text{cl}(\text{dom} f), D_f(x, y_n) \text{ bounded} \Rightarrow y \in \text{int}(\text{dom} f) \]

(iii) \[ (x_n) \text{ in } \text{dom} f, x_n \rightarrow x \in \text{dom} f, (y_n) \text{ in } \text{int}(\text{dom} f), y_n \rightarrow y \in \text{dom} f, \{x, y\} \cap \text{int}(\text{dom} f) \neq \emptyset, D_f(x_n, y_n) \rightarrow 0 \Rightarrow D_f(x, y) = 0 \]

**Proof.** (i): Assume to the contrary that $\lim D_f(x, y_n) < +\infty$. Without loss (subsequences!), we assume $\|\nabla f(y_n)\| \rightarrow +\infty$ (by essential smoothness), $\|\nabla f(y_n)\| \|\nabla f(y_n)\| \rightarrow q \neq 0$, and $(D_f(x, y_n)) = (f(x) + f^*(\langle \nabla f(y_n), \cdot \rangle))$ is bounded. Dividing the last sequence by $\|\nabla f(y_n)\|$ yields $f^*(\langle \nabla f(y_n), \cdot \rangle)/\|\nabla f(y_n)\| \rightarrow \langle q, \cdot \rangle$. By Proposition 2.14, $\langle q, x \rangle \geq (f^*0^+)(q)$; thus (Fact 2.13) $x \not\in \text{int}(\text{dom} f)$ which is absurd.

(ii): is equivalent to (i). The "(and $D_f(y_n, y_n) \rightarrow 0$)" part follows from Proposition 3.2(ii).

(iii): Claim: $y \in \text{int}(\text{dom} f)$.

Otherwise, $y \in \text{dom} f \setminus \text{int}(\text{dom} f)$. Then $x \in \text{int}(\text{dom} f)$ and $\|\nabla f(y_n)\| \rightarrow +\infty$. We assume (subsequence!) that $\langle \nabla f(y_n), \cdot \rangle$ is convergent, say to $q \neq 0$. Now $0 \leftarrow D_f(x_n, y_n) = f(x_n) + f^*(\langle \nabla f(y_n), \cdot \rangle) - \langle \nabla f(y_n), x_n \rangle$: thus division by $\|\nabla f(y_n)\|$ yields $\lim \frac{\langle \nabla f(y_n), \cdot \rangle}{\|\nabla f(y_n)\|} = \langle q, x \rangle$. Proposition 2.14 implies $\langle q, x \rangle \geq (f^*0^+)(q)$. Hence, by Fact 2.13, $x \not\in \text{int}(\text{dom} f^*) = \text{int}(\text{dom} f)$. This is the desired contradiction and the claim thus holds.

Because $f$ is closed, we get $0 = \lim_n D_f(x_n, y_n) = \lim_n f(x_n) + f^*(\langle \nabla f(y_n), \cdot \rangle) - \langle \nabla f(y_n), x_n \rangle \geq f(x) + f^*(\langle \nabla f(y), \cdot \rangle) - \langle \nabla f(y), x \rangle = D_f(x, y) \geq 0$. \hfill \Box

**Legendreness**

**Theorem 3.9** ("Legendreness helps a lot") Suppose $f$ is Legendre on $E$. Then:

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

(ii) If dom$f^*$ is not open, then $D_f(x, \cdot)$ is not coercive, $\forall x \in \text{dom} f$.  

10
\[(x_n) \text{ in } \text{dom} f, x_n \to x \in \text{dom} f, \]

\[(y_n) \text{ in } \text{int}(\text{dom} f), y_n \to y \in \text{dom} f, \]

\[\{x, y\} \cap \text{int}(\text{dom} f) \neq \emptyset, D_f(x_n, y_n) \to 0\}
\[\Rightarrow x = y.\]

**Proof.** (i): \(f^*\) is essentially smooth and essentially strictly convex, since \(f\) is Legendre. Hence Theorem 3.7.(v) applies.

(ii): Fix \(y^* \in \text{dom} f^* \setminus \text{int}(\text{dom} f^*), y_n^* \in \text{int}(\text{dom} f^*),\) and let \(y_n^* := (1 - 1/n)y^* + (1/n)y_n^*,\) for all \(n \geq 2.\) Then the sequence \((y_n^*)\) lies in \(\text{int}(\text{dom} f^*)\) and converges to \(y^*\) along the segment between \(y_n^*\) and \(y.\) Hence \(f^*(y_n^*) \to f^*(y^*)\) ([28, Corollary 7.5.1]). Let \(y_n := \nabla f^*(y_n^*),\) for all \(n;\) then \(\|y_n\| \to +\infty.\) Now fix an arbitrary \(x \in \text{dom} f.\) The sequences \((y_n^*, x), (f^*(y_n^*))\) 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 \(\text{dom} f = \mathbb{R}.\) 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 [17, Proposition 9.1] or Chen and Teboule’s [11, Lemma 3.1]) gives

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

The first term, \(D_f(x_n, y_n),\) tends to \(+\infty\) by Theorem 3.7.(vi). The second term, \(-D_f(x_n, x),\) 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 \((x_n)\) \textit{a posteriori} so that \(D_f(x_n, y_n) + \nabla f(y_n)(x - x_n) \to 0.\) Then altogether

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

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:

\[D_f(x, \cdot) \text{ is coercive, for some } x \in \text{dom} f.\]

\[\Rightarrow \text{ dom} f^* \text{ is open.}\]

\[\Rightarrow D_f(x, \cdot) \text{ is coercive, for all } x \in \text{int}(\text{dom} f).\]

Consequently, the following are equivalent:

(i) \(D_f(x, \cdot) \text{ is coercive, for some } x \in \text{int}(\text{dom} f).\)

(ii) \(\text{ dom} f^* \text{ is open.}\)

(iii) \(D_f(x, \cdot) \text{ is coercive, for all } x \in \text{int}(\text{dom} f).\)

**Proof.** The first implication is Theorem 3.9.(ii), the second is Theorem 3.7.(vi); the “Consequently” part follows. \(\square\)
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 $\text{int}(\text{dom}f)$ and a set $C$ with $\text{int}(\text{dom}f) \cap C \neq \emptyset$ (the usual constraint qualification). Pick $y \in \text{int}(\text{dom}f)$. We wish to define a “projection of $y$ onto $C$ w.r.t. $f$”, denoted $P_Cy$ or $P^f_C$, by

$$P_Cy = \arg\min_{x \in C \cap \text{dom}f} D_f(x, y).$$

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 $\text{int}(\text{dom}f)$ to make even sense of the argmin. Moreover, to be able to project the point $P_Cy$ again (onto another constraint set perhaps), we have to impose

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

(This shows \textit{a posteriori} that the constraint qualification “$\text{int}(\text{dom}f) \cap C \neq \emptyset$" is \textit{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.

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

(i) If $f$ is essentially smooth, then $\arg\inf_{x \in C \cap \text{dom}f} D_f(x, y)$ is nonempty.

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

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

\textbf{Proof.} (i): On the one hand, $D_f(\cdot, y)$ is closed and coercive (Theorem 3.7). On the other hand, $C \cap \text{cl}(\text{dom}f)$ is closed. Altogether, $\arg\min_{x \in C \cap \text{cl}(\text{dom}f)} D_f(x, y)$ is nonempty; of course, this argmin is equal to $\arg\min_{x \in C \cap \text{dom}f} D_f(x, y)$. 

12
(ii): Since $f$ is strictly convex on $\text{dom} f$, so is $D_f(\cdot, y)$.

(iii): By (i), $\operatorname{arginf}_{x \in C \cap \text{dom} f} D_f(x, y)$ is nonempty.

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

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

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

Then $\Phi'(t) = \langle \nabla f((1-t)\bar{x} + tz), z - \bar{x} \rangle - \langle \nabla f(y), z - \bar{x} \rangle$, for all $t \in [0, 1]$. By essential smoothness, $\lim_{t \to 0} \Phi'(t) = -\infty$. This implies $\Phi(t) < \Phi(0)$, for all small positive $t$. However, $(1-t)\bar{x} + tz \in C \cap \text{int}(\text{dom} f)$ 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 $\text{int}(\text{dom} f)$, we conclude that the argmin is singleton. \hfill \square

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

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

is singleton and contained in $\text{int}(\text{dom} f)$; 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

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

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 $\text{dom} f := \{ x \in \mathbb{R}^2 : x \geq 0 \}$ and $C := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 1 \}$. Then $C \cap \text{int}(\text{dom} f) \ni (\frac{1}{2}, \frac{1}{2})$ is nonempty. Let further $y := (2, 1) \in \text{int}(\text{dom} f)$. One easily checks that

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

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 \( C \cap \text{int}(\text{dom} f) \neq \emptyset \). Suppose further \( y \in \text{int}(\text{dom} f) \). Then the Bregman projection \( P_C y \) is characterized by

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

In addition,

\[
D_f(P_C y, y) \leq D_f(c, y) - D_f(c, P_C y), \quad \text{for all} \; c \in C \cap \text{dom} f.
\]

Proof. Convex calculus time!

\[
\bar{x} = P_C y
\]

Theorem 3.12 (iii) \[
\bar{x} = \arg\min_{x \in C \cap \text{int}(\text{dom} f)} D_f(x, y)
\]

\[
\bar{x} = \arg\min_{x \in C \cap \text{int}(\text{dom} f)} f(x) - \langle \nabla f(y), x \rangle
\]

\[
\bar{x} = \arg\min_{x \in E} f(x) + \langle - \nabla f(y), x \rangle + \iota_C(x) + \iota_{\text{int}(\text{dom} f)}(x)
\]

\[
0 \in \partial \left( f(\cdot) + \langle - \nabla f(y), \cdot \rangle + \iota_C(\cdot) + \iota_{\text{int}(\text{dom} f)}(\cdot) \right)(\bar{x})
\]

[28, Theorem 23.8] \[
0 \in \nabla f(\bar{x}) - \nabla f(y) + \partial \iota_C(\bar{x}) \quad \text{and} \; \bar{x} \in \text{int}(\text{dom} f),
\]

which gives the desired characterization. The “In addition” part is a trivial expansion. \( \square \)

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:
\[ B0. \text{ dom} f \text{ is closed and } \text{int}(\text{dom} f) \neq \emptyset. \]
\[ B1. \text{ } f \text{ is continuously differentiable on } \text{int}(\text{dom} f). \]
\[ B2. \text{ } f \text{ is strictly convex and continuous on } \text{dom} f. \]
\[ B3. \begin{align*}
(i) & \quad D_f(x, \cdot) \text{ is coercive, } \forall y \in \text{int}(\text{dom} f). \\
(ii) & \quad D_f(x, \cdot) \text{ is coercive, } \forall x \in \text{dom} f.
\end{align*} \]
\[ B4. \begin{align*}
& \begin{cases}
(y_n) \text{ in } \text{int}(\text{dom} f), y_n \to y \\
(x_n) \text{ in } \text{dom } f, (x_n) \text{ bounded},
\end{cases}
\Rightarrow \quad D_f(y, y_n) \to 0.
\end{align*} \]
\[ B5. \begin{align*}
& \begin{cases}
(y_n) \text{ in } \text{int}(\text{dom} f), y_n \to y, \\
D_f(x_n, y_n) \to 0
\end{cases}
\Rightarrow x_n \to y.
\end{align*} \]

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 \( \text{int}(\text{dom} f) \): if \( f \) is differentiable throughout \( \text{int}(\text{dom} f) \), then it is actually continuously differentiable (see Rockafellar’s [28, Corollary 25.5.1]).
- 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:
  - dom\( f \) open, i.e. dom\( f = E \) : BL3.(ii) \( \Leftrightarrow \) dom\( f^* \) open.
  - dom\( f \) not open: BL3.(ii) \( \Leftrightarrow D_f(x, \cdot) \) is coercive, \( \forall x \in \text{bd}(\text{dom} f) \).

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 dom\( f := \{ x \in E : x_j \geq 0 \text{ for all } j \} \) 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 [17], Iusem discusses two additional useful properties of Bregman functions:

**Definition 4.4** (Iusem [17]) Suppose $f$ is Bregman on $E$.

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

$$\begin{align*}
(x_n) & \text{ in } \text{int} \left(\text{dom} f \right), \ y \in \text{int} \left(\text{dom} f \right), \\
 x_n & \to x \in \text{bd} \left(\text{dom} f \right) \\
\end{align*}$$

$$\Rightarrow \quad \langle \nabla f(x_n), y - x_n \rangle \to -\infty.$$

(ii) $f$ is called **zone coercive**, if $\nabla f$ 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 $x \in \text{bd} \left(\text{dom} f \right)$, $y \in \text{int} \left(\text{dom} f \right)$, 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 $y \in \text{int} \left(\text{dom} f \right)$ and $(x_n)$ in $\text{int} \left(\text{dom} f \right)$ converging to $x \in \text{bd} \left(\text{dom} f \right)$. Since $f$ and $f^*$ are Legendre, we have $||\nabla f(x_n)|| \to +\infty$ and $\nabla f(x_n), \nabla f(y) \in \text{int} \left(\text{dom} f^* \right)$. Also, $D_{f^*}(\cdot, \nabla f(y))$ is coercive and hence

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

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 [15, Theorem X.4.1.3], $f^*$ is strictly convex on $\text{dom} f^* = E$. 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 $\text{dom} f^* = E$; hence $f^*$ is sub-differentiable everywhere: $\text{dom} \partial f^* = \text{range} \partial f = E$. Now $\partial f$ is at most singleton (Fact 2.2); thus $\text{range} \partial f = \text{range} \nabla f = E$. The proof is complete. $\square$

We now obtain Teboulle's generalization (see Burachik's [5, Lemma 2.6]) of Iusem's [17, Corollary 9.1].

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

$$\text{zone coercivity } \Rightarrow \text{ boundary coercivity } \Rightarrow \text{Legendreness}.$$
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 “Bregmaness”.

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

(i) $\text{dom} f^* \text{ is open} \iff f \text{ is Bregman and boundary coercive.}$

(ii) $\text{dom} f^* = E \iff f \text{ is Bregman and zone coercive.}$

**Proof.** Fix $f$ strictly convex and differentiable on $\text{dom} f = E$. We first check that $f$ is Bregman except for possibly B3.(ii). B0: $\checkmark$. B1: $\checkmark$ (see Remarks 4.2). B2: $\checkmark$. B3.(i): $\checkmark$ (see Remarks 4.2). B3.(ii): Theorem 3.9 says: $D_f(x, \cdot)$ is coercive, $\forall x \in \text{int} (\text{dom} f) = \text{dom} f = E$ if and only if $\text{dom} f^*$ is open. B4: $\checkmark$ (Proposition 3.2.(ii)). B5: $\checkmark$ (Theorem 3.3.(iii)). Hence: $f$ is Bregman $\iff$ B3.(ii) holds $\iff \text{dom} f^*$ is open. Consequently, (i) is true. But (ii) follows from Theorem 4.5 and Proposition 2.16. $\Box$

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 [27, Theorem 5.1]. 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 $\alpha f$ is Legendre with $D_{\alpha f} = \alpha D_f$.

(ii) Suppose $g$ is closed convex proper on $E$ and essentially smooth. If $\text{int}(\text{dom} f) \cap \text{int}(\text{dom} g) \neq \emptyset$, 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 $D_{f \circ T}(x, y) = D_f(Tx, Ty), \forall x \in E, y \in \text{int}(\text{dom}(f \circ T)) = T^{-1}(\text{int}(\text{dom} f))$.

(iv) Suppose $s$ is convex, differentiable, and strictly increasing on $\mathbb{R}$. Then $s \circ f$ is Legendre with $D_{s \circ f}(x, y) = D_s(f(x), f(y)) + s(f(y))D_f(x, y), \forall x \in E, y \in \text{int}(\text{dom}(s \circ f)) = \text{int}(\text{dom} f)$.

(v) Suppose $g$ is closed, convex, proper, essentially strictly convex on $E$ with $\text{int} (\text{dom} g^*) \cap \text{int} (\text{dom} f^*) \neq \emptyset$. Then $f \square g$ is Legendre.
Proof. (i): is trivial.
(ii): \( f + g \) is a closed convex proper function on \( E \). We check L0 through L3.
L0: \( \text{int}(\text{dom}(f + g)) = \text{int}((\text{dom} f) \cap (\text{dom} g)) = \text{int}(\text{dom} f) \cap \text{int}(\text{dom} g) \neq \emptyset \).
L1: \( f + g \) is differentiable on \( \text{int}(\text{dom}(f + g)) \), since \( f \) and \( g \) are.
L2: Suppose \( x \in \text{bd}(\text{dom}(f + g)) \). Then \( x \in \text{bd}(\text{dom} f) \cup \text{bd}(\text{dom} g) \). Fix \( y \in \text{int}(\text{dom}(f + g)) = \text{int}(\text{dom} f) \cap \text{int}(\text{dom} g) \).
  Case 1: \( x \in \text{bd}(\text{dom} f) \). Then \( \lim_{t \to 0} \langle \nabla f(x + t(y - x)), y - x \rangle = -\infty \). If \( x \in \text{bd}(\text{dom} g) \), then \( \lim_{t \to 0} \langle \nabla g(x + t(y - x)), y - x \rangle = -\infty \). Otherwise, \( x \in \text{int}(\text{dom} g) \) and \( \lim_{t \to 0} \langle \nabla g(x + t(y - x)), y - x \rangle = \langle \nabla g(x), y - x \rangle \in \mathbb{R} \). For either alternative, \( \lim_{t \to 0} \langle \nabla (f + g)(x + t(y - x)), y - x \rangle = -\infty \), as sought for. Case 2: \( x \in \text{int}(\text{dom} f) \). Then \( x \in \text{bd}(\text{dom} g) \) and we reason analogously.
L3: \( f + g \) is strictly convex on \( \text{int}(\text{dom}(f + g)) \), since \( f \) is.
(iii): follows easily with results of [28, Section 6].
(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 \( \nabla s \) 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 \( (f^* + g^*)^* = f \cap g \); see [28, Theorem 16.4].

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.** \( \text{dom} f^* \) is open.

**BL1.** \( D_f(x, \cdot) \) is coercive, \( \forall x \in \text{dom} f \setminus \text{int}(\text{dom} f) \).

**BL2.** \( x \in \text{dom} f \setminus \text{int}(\text{dom} f), (y_n) \) in \( \text{int}(\text{dom} f) \), \( y_n \to y \in \text{bd}(\text{dom} f) \), \( (D_f(x, y_n)) \) bounded \( \Rightarrow \)
(\( D_f(y_n) \to 0 \) (and hence \( y \in \text{dom} f \)).

**BL3.** \( (x_n), (y_n) \) in \( \text{int}(\text{dom} f), x_n \to x \in \text{dom} f \setminus \text{int}(\text{dom} f), y_n \to y \in \text{dom} f \setminus \text{int}(\text{dom} f), D_f(x_n, y_n) \to 0 \) \( \Rightarrow \) \( x = y \).

**Remarks 5.3**

- In view of Corollary 3.11, BL0 and BL1 together say that \( D_f(x, \cdot) \) is coercive, \( \forall x \in \text{dom} f \). 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:

- **domf open:** BL0 and BL1 \( \leftrightarrow \) domf open \( \leftrightarrow \) \( D_f(x, \cdot) \) is coercive, for some \( x \in \text{dom} f \).
- **domf not open:** BL0 and BL1 \( \leftrightarrow \) domf open and BL1 \( \leftrightarrow \) \( D_f(x, \cdot) \) is coercive, for all \( x \in \text{dom} f \setminus \text{int}(\text{dom} f) \).
• BL1 is equivalent to
\[
x \in \text{dom } f \setminus \text{int}(\text{dom } f), \quad (y_n) \text{ in } \text{int}(\text{dom } f), \\
(D_f(x, y_n)) \text{ bounded}
\] \Rightarrow \quad (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 \( \text{dom } f := [0, +\infty] \)
is Bregman/Legendre. Given \( 0 \in \text{bd}(\text{dom } f) \) 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:

\[
\text{Legendre } \Leftrightarrow \text{Bregman/Legendre } \Leftrightarrow \text{Bregman and Legendre } \Rightarrow \text{Bregman.}
\]

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:
\[
(y_n) \text{ in } \text{int}(\text{dom } f), \quad y_n \to y \in \text{dom } f \setminus \text{int}(\text{dom } f), \\
x \in \text{dom } f \setminus \text{int}(\text{dom } f), \quad D_f(x, y_n) \to 0
\] \Rightarrow \quad 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 \( \text{int}(\text{dom } f) \) 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/Legendreeness” can be very easy:

**Theorem 5.6** Suppose \( f \) is Legendre on \( E \) with \( \text{dom } f \) open. Then:
\[
f \text{ is Bregman/Legendre } \Leftrightarrow \text{ dom } f^* \text{ is open.}
\]
Proof. BL0 and BL1 ⇔ dom$f^*$ open, as observed in Remarks 5.3.
Now BL2 and BL3 hold trivially. \( \square \)

Remark 5.7 De Pierro and Iusem ([27, Theorem 5.1]) 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 ⇔ dom$f^*$ is open.
(ii) BL2 always holds.
(iii) BL3 always holds.

Consequently,

\( f \) is Bregman/Legendre ⇔ \( f \) is Legendre and dom$f^*$ is open.

Proof. The theorem is clear if dom\( f \) is open (Theorem 5.6). So we assume without loss that dom\( f \) is not open.
(i): “⇒” follows from Corollary 3.11.
“⇐”: In view of Corollary 3.11, it is enough to show that BL1 holds. We can assume that dom\( f \) is unbounded, say dom\( f = [x, +\infty[ \) (the remaining case 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 \( \text{int}(\text{dom} f) = ]x, +\infty[ \) with \( y_n \uparrow +\infty \)
but \( (D_f(x, y_n)) \) bounded.
Claim: \( \nabla f(y_n) \to +\infty \).
Because \( f \) is convex, \( \nabla f \) is increasing. If the claim does not hold, then we have
\[ (f(x) + f^*(\nabla f(y_n))) - \langle \nabla f(y_n), x \rangle \]
is a bounded sequence. Since \( f^* \) is closed, we get \( y^* \in \text{dom} f^* = \text{int}(\text{dom} f^*) \)
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 \( y \in \text{dom} f \). This is obviously true if \( \text{dom} f \) is closed. Hence we can assume that \( \text{dom} f \) 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 \( \text{dom} f = [x, y[ \). Since \( y_n \to y \in \text{bd}(\text{dom} f) \), 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

20
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 $\text{dom} f$ has two boundary points. We assume to the contrary that BL3 fails, i.e. $\text{dom} f = [x, y]$.

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. \qed

Remarks 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:

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

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

(i) If $\alpha > 0$, then $\alpha f$ is Bregman/Legendre.
(ii) Suppose $g$ is essentially smooth on $E$ with

\[ \begin{align*}
S2. \\
x \in \text{dom} g, (y_n) \in \text{int}(\text{dom} g), \\
y_n \rightarrow y \in \text{cl}(\text{dom} g), (D_g(x, y_n)) \text{ bounded} \\
& \implies D_g(y, y_n) \to 0.
\end{align*} \]

If $\text{int}(\text{dom} f) \cap \text{int}(\text{dom} g) = \emptyset$, 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(iii)). We have to check BL0 through BL3 for $h$.

BL0: Fix a point $x \in \text{int}(\text{dom} h)$ and a sequence $(y_n)$ in $\text{int}(\text{dom} h)$ with $(D_h(x, y_n))$ bounded. Then $x \in \text{int}(\text{dom} f), (y_n)$ lies in $\text{int}(\text{dom} f)$, and $(D_f(x, y_n))$ is bounded. Now 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 dom $h^*$.

BL1: Fix $x \in \text{dom} h \setminus \text{int}(\text{dom} h)$ and a sequence $(y_n)$ in $\text{int}(\text{dom} h)$ with $(D_h(x, y_n))$ bounded. Note that $(y_n)$ is in $\text{int}(\text{dom} f)$ and that $(D_f(x, y_n))$ is bounded. If $x \in \text{dom} f \setminus \text{int}(\text{dom} f)$, then, by BL1 for $f$, the sequence $(y_n)$ is bounded. Otherwise, $x \in \text{int}(\text{dom} f)$ and the boundedness of $(y_n)$ follows from BL0 and
Corollary 3.11 (for $f$).

**BL2:** Fix $x \in \text{dom} h \setminus \text{int}(\text{dom} h)$ and $(y_n)$ in $\text{int}(\text{dom} h)$ with $y_n \to y \in \text{bd}(\text{dom} h)$ and $(D_h(x, y_n))$ bounded. Then $(D_f(x, y_n))$ and $(D_g(y, y_n))$ are bounded. S2 implies $y \in \text{dom} g$ and $D_f(y, y_n) \to 0$. It suffices to show that $y \in \text{dom} f$ and $D_f(y, y_n) \to 0$. **Case 1:** $x \in \text{int}(\text{dom} f)$. Employ Theorem 3.8.(ii). **Case 2:** $x \in \text{dom} f \setminus \text{int}(\text{dom} f)$. If $y \in \text{int}(\text{dom} f)$, then use Proposition 3.2.(ii). Otherwise, $y \in \text{bd}(\text{dom} f)$, and we apply BL2.

**BL3:** Fix sequences $(x_n)$, $(y_n)$ in $\text{int}(\text{dom} h)$ with $x_n \to x \in \text{dom} h \setminus \text{int}(\text{dom} h)$, $y_n \to y \in \text{dom} h \setminus \text{int}(\text{dom} h)$, and $D_h(x_n, y_n) \to 0$. If $(x_n, y_n) \cap \text{int}(\text{dom} f) \neq \emptyset$, then we make use of Theorem 3.9.(iii). Otherwise, $x, y \in \text{dom} f \setminus \text{int}(\text{dom} f)$ 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 $g^*(x^*) = f^*((L^*)^{-1} x^*) - ((L^*)^{-1}, T0)$ so that $\text{dom} g^* = L^* \text{dom} f^*$ 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 $\text{int}(\text{dom} f) \cap \text{int}(\text{dom} g) \neq \emptyset$, 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 $\text{int}(\text{dom} f) \cap \text{int}(\text{dom} g) \neq \emptyset$ and we only have to check S2. So fix $x \in \text{dom} g$, $(y_n)$ in $\text{int}(\text{dom} g)$ with $y_n \to y \in \text{cl}(\text{dom} g)$ and $(D_g(y, y_n))$ bounded. Goal: $y \in \text{dom} g$ and $D_g(y, y_n) \to 0$. **Case 1:** $x \in \text{int}(\text{dom} g)$. Apply Theorem 3.8.(ii). **Case 2:** $x \in \text{dom} g \setminus \text{int}(\text{dom} g)$. If $y \in \text{int}(\text{dom} g)$, then use Proposition 3.2.(ii). Else $y \in \text{bd}(\text{dom} g)$ and BL2 (for $g$) does the job. □

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

$$f : E := \prod_j E_j \to [-\infty, +\infty] : x := (x_1, \ldots, x_J) \mapsto \sum_j \lambda_j f_j(x_j)$$

is Bregman/Legendre on $E$.

**Proof.** Note that $\nabla f(x) = (\nabla f_1(x_1), \ldots, \nabla f_J(x_J))$, $\text{int}(\text{dom} f) = \prod_j \text{int}(\text{dom} f_j)$, and $D_f(x, y) = \sum_j D_{f_j}(x_j, y_j)$ (Proposition 3.5). We first check that $f$ is Legendre; see Definition 2.8.

L0: $\checkmark$. L1: $\checkmark$. L2: Take $(x_n)$ in $\text{int}(\text{dom} f)$ with $x_n \to x = (x_1, \ldots, x_J) \in \text{bd}(\text{dom} f)$. Then there is a $j$ such that $x_j \in \text{bd}(\text{dom} f_j)$. Hence, by Fact 2.2,

$$\|\nabla f(x_n)\| \geq \|\nabla f_j(x_n)\| \to +\infty;$$

thus L2 holds. L3: $\checkmark$.

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, $\forall x_j \in \text{dom} f_j$. It easily follows that $D_f(x; \cdot)$ is coercive, $\forall x \in \text{dom} f$.

BL2: Pick $x$, $y$, and $(y_n)$ as in the hypothesis of BL2 for $f$. If $y_j \in \text{int}(\text{dom} f_j)$, apply Proposition 3.2(ii) to conclude $D_{f_j}(y_j, (y_n)_j) \to 0$. Else $y_j \in \text{bd}(\text{dom} f_j)$ and, depending on the location of $x_j$, either Theorem 3.8(ii) or BL2 for $f_j$ applies.

BL3: Fix $x$, $y$, $(x_n)$, and $(y_n)$ as in the hypothesis of BL3 for $f$. 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 $f_1, \ldots, f_J$ are Legendre on $\mathbb{R}$ with $\text{dom} f_1^*, \ldots, \text{dom} f_J^*$ open. Then

\[ f : E = \mathbb{R}^d \to ] - \infty, +\infty[ : x := (x_1, \ldots, x_J) \mapsto \sum_j f_j(x_j) \]

is Bregman/Legendre type on $E$.

**Proof.** Combine Theorem 5.8 with Theorem 5.12. \[ \square \]

6 Examples

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 $\text{dom} f^*$. Examples 6.1 to 6.7 are standard; see Rockafellar’s [28, Section 12].

**Example 6.1** Suppose $1 < p < +\infty$ and $f(x) = \frac{1}{p}|x|^p$ on $\text{dom} f = \mathbb{R}$. Then $f^*(x^*) = \frac{1}{q}|x^*|^q$ on $\text{dom} f^* = \mathbb{R}$, 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 \( \text{dom} f = [0, +\infty[ \). Then \( f^*(x^*) = -\frac{1}{p} (-x^*)^p \) on \( \text{dom} f^* = ]-\infty, 0[ \), 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 \( \text{dom} f = [-1, +1] \). Then \( f^*(x^*) = \sqrt{1 + (x^*)^2} \) on \( \text{dom} f^* = \mathbb{R} \).

Hence \( f \) is Bregman/Legendre whereas \( f^* \) is not.

Example 6.5 ("Boltzmann/Shannon") Suppose \( f(x) = x \ln x - x \) on \( \text{dom} f = [0, +\infty[ \). Then \( f^*(x^*) = \exp x^* \) on \( \text{dom} f^* = \mathbb{R} \).
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 $\text{dom} f = [0, 1]$. Then $f^*(x^*) = \ln(1 + \exp x^*)$ on $\text{dom} f^* = \mathbb{R}$.

Hence $f$ is Bregman/Legendre whereas $f^*$ is not.

**Example 6.7** ("Burg") Suppose $f(x) = -\frac{1}{2} \ln x$ on $\text{dom} f = ]0, +\infty[$. Then $f^*(x^*) = -\frac{1}{2} \ln(-x^*)$ on $\text{dom} f^* = ]-\infty, 0[.$

Hence $f$ and $f^*$ are Bregman/Legendre.
Example 6.8 (De Pierro & Iusem’s [27, Example on page 438]) Suppose \( f(x) = \frac{1}{2}(x^2 - 4x + 3) \), if \( x \leq 1; -\ln x \), otherwise; on \( \text{dom} f = \mathbb{R} \). 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 \( \text{dom} f^* = [-\infty, 0[ \).

Hence \( f \) and \( f^* \) are Bregman/Legendre.

We summarize these examples in the following table:

<table>
<thead>
<tr>
<th>Some Bregman/Legendre functions on ( \mathbb{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(x) )</td>
</tr>
<tr>
<td>( \frac{1}{p}</td>
</tr>
<tr>
<td>( \frac{1}{2}x^2 )</td>
</tr>
<tr>
<td>( -\frac{1}{p}x^p )</td>
</tr>
<tr>
<td>( -\sqrt{1 - x^2} )</td>
</tr>
<tr>
<td>( x \ln x - x )</td>
</tr>
<tr>
<td>( x \ln x + (1 - x) \ln(1 - x) )</td>
</tr>
<tr>
<td>( -\frac{1}{2}\ln x )</td>
</tr>
<tr>
<td>( \frac{1}{2}(x^2 - 4x + 3), \ x \leq 1; -\ln x, \ x \geq 1 )</td>
</tr>
</tbody>
</table>

**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.

26
(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

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

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

**Proof.** Verifying that \( f \) is of Legendre type and calculating \( f^* \) is straightforward 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 \( y_n \to y \not\in \text{dom} f \), 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, \( x \neq y \).

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

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

Then \( f \) is closed convex proper; see Rockafellar’s [28, Remark following Corollary 8.5.1 on page 67]. 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 \((x, r), (y, s) \in \text{int(dom}f\)\. Example 6.9 arises by choosing

\[
g(x) = -\ln x \quad \text{and} \quad h(r) = r \ln r.
\]
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 [6, Lemma 6.1]) Suppose \( f \) is Legendre on \( E \) and \( C \) is an affine subspace of \( E \), say \( C = \{ x \in E : Ax = b \} \), for \( A : E \rightarrow \mathbb{R}^M : x \mapsto (\langle a^{(m)}, x \rangle)_{m=1}^M \) and \( b \in \mathbb{R}^M \). Suppose further \( C \cap \text{int}(\text{dom} f) \neq \emptyset \) and \( y \in \text{int}(\text{dom} f) \). Then \( z = P_C y \) exactly when

\[
z \in \text{int}(\text{dom} f), \quad Az = b, \quad \text{and} \quad \nabla f(z) = \nabla f(y) + \sum_{m=1}^M \mu_m a^{(m)},
\]

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 \( (\text{kernel} A)^{\perp} = \text{range} A^* = \text{span}(a^{(1)}, \ldots, a^{(M)}) \). \( \square \)

Corollary 6.12 (see also Bregman’s [4, Theorem 3]) 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 \( H \cap \text{int}(\text{dom} f) \neq \emptyset \) and \( y \in \text{int}(\text{dom} f) \). Then \( z = P_H y \) exactly when

\[
z \in \text{int}(\text{dom} f), \quad \langle a, z \rangle = b, \quad \text{and} \quad \nabla f(z) = \nabla f(y) + \mu a,
\]

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 \( \mu \) by Fact 2.9:

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

2. Estimate \( \mu \) by solving \( \langle a, z(\mu) \rangle = b \) subject to \( z(\mu) \in \text{int}(\text{dom} f) \).

3. Compute \( z(\mu) \) through 1.

The following table 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 \( f(x) = \sum J f_j(x_j) \):
\[
\begin{array}{|l|l|}
\hline
f_j(x_j) & z_j(\mu) \\
\hline
\frac{1}{2}x_j^2 & y_j + \mu a_j \\
x_j \ln x_j - x_j & y_j \exp(\mu a_j) \\
x_j \ln x_j + (1 - x_j) \ln(1 - x_j) & \frac{\exp(\mu a_j) y_j/(1 - y_j)}{1 + \exp(\mu a_j) y_j/(1 - y_j)} \\
-\frac{1}{2} - \ln x_j & y_j/(1 - \mu a_j y_j) \\
\frac{1}{p}|x_j|^p & (\text{sign}(y_j)|y_j|^{p-1} + \mu a_j)^{1/(p-1)} \\
-\sqrt{1 - x_j^2} & \frac{\mu a_j + y_j/(1 - y_j^2)^{1/2}}{\sqrt{\left(\mu a_j + y_j/(1 - y_j^2)^{1/2}\right)^2 + 1}} \\
-4\sqrt{x_j} & y_j/(1 - \sqrt{\mu a_j/2})^2 \\
\hline
\end{array}
\]

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 \( \mu \) (and one then recovers the well-known formula for the orthogonal projection onto a hyperplane). However, \( \mu \) 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].

**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

\[
a \geq 0 \quad \text{and} \quad b \leq 0 \quad \text{or} \quad a \leq 0 \quad \text{and} \quad b \geq 0. \]

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** (“probabilistic Boltzmann/Shannon”) Let \( f(x) = \sum_j x_j \ln x_j - x_j \) on \( \text{dom} f = \{ x \in \mathbb{R}^d : x_j \geq 0, \forall j \} \) and \( H \) be the “probabilistic hyperplane” \( H = \{ x \in \mathbb{R}^d : \sum_j x_j = 1 \} \). By Remark 6.14, \( H \cap \text{int}(\text{dom} f) \neq \emptyset \). 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

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,

\[
f : E := \prod_i E_i \rightarrow [0, +\infty] \quad : \quad x := (x_1, \ldots, x_N) \mapsto \sum_i \lambda_i f_i(x_i)
\]
is Bregman/Legendre on \( E \).

Define the product set

\[
C := \prod_i C_i = \{(x_1, \ldots, x_N) \in E : x_i \in C_i, \ \forall i\}
\]
and the diagonal set

\[
\Delta := \{(e, \ldots, e) \in E : e \in E\}.
\]

Then \( C \cap \Delta \neq \emptyset \) if and only if \( \bigcap_i C_i \neq \emptyset \); this reduction to two sets in the product space \( E \) goes back at least as far as Pierra [25].

Now what do the Bregman projections onto \( C \) and \( \Delta \) with respect to \( f \) look like? They are well-defined (Theorem 3.12) as soon as \( C \cap \text{int}(\text{dom} f) \neq \emptyset \) and \( \Delta \cap \text{int}(\text{dom} f) \neq \emptyset \); equivalently:

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

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

**Proposition 7.1** If \( (*) \) holds, then for every \( y = (y_1, \ldots, y_N) \in \prod_i \text{int}(\text{dom} f_i) \):

(i) \( P_f^C(y) = (P_{C_1}^f(y_1), \ldots, P_{C_N}^f(y_N)) \).

(ii) \( z = (z_1, \ldots, z) = P_{\Delta}^f(y) \) if and only if

\[
z \in \bigcap_i \text{int}(\text{dom} f_i) \quad \text{and} \quad \sum_i \lambda_i \nabla f_i(z) = \sum_i \lambda_i \nabla f_i(y_i).
\]

**Proof.** (i): Obvious, since \( D_f(x, y) = \sum_i \lambda_i D_{f_i}(x_i, y_i) \) is separable.

(ii): By Proposition 3.16, \( z = (z_1, \ldots, z) = P_{\Delta}^f(y) \) if and only if \( z \in \bigcap_i \text{int}(\text{dom} f_i) \) and

\[
\langle \nabla f(y) - \nabla f(z), \Delta - z \rangle \leq 0.
\]

Now \( \Delta - z = \Delta \) is a subspace and \( \Delta^\perp = \{(x_1, \ldots, x_N) \in E : \sum_i x_i = 0\} \); the result follows. \( \square \)
Corollary 7.2 Suppose $f$ is Bregman/Legendre on $E$ and $\lambda_1, \ldots, \lambda_N$ are strictly positive weights: $\sum_i \lambda_i = 1$. Let $f(x) := \sum_i \lambda_i f(x_i)$, for all $x = (x_1, \ldots, x_N) \in E$. If $y = (y_1, \ldots, y_N) \in \prod_i \text{int}(\text{dom} f)$, then

$$P^f_{\Delta}(y) = (z, \ldots, z), \quad \text{where} \quad z = \nabla f^*(\sum_i \lambda_i \nabla f(y_i)).$$

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

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

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 $\text{dom} f = E = \mathbb{R}^d$. Let $f(x) = \sum_i \lambda_i f(x_i)$ and $y \in \prod_i \text{int}(\text{dom} f) = E = (\mathbb{R}^d)^N$. Then $P^f_{\Delta}(y) = (z, \ldots, z)$, 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 $P^f_{\Delta}(y)$ 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 $f(x) = \frac{1}{2} ||x||_2^2 = \frac{1}{2} \sum_j |x(j)|^2$ on $\text{dom} f = E = \mathbb{R}^d$. Let $f(x) = \sum_i \lambda_i f(x_i)$ and $y \in \prod_i \text{int}(\text{dom} f) = E = (\mathbb{R}^d)^N$. Then

$$P^f_{\Delta}(y) = \lambda_1 y_1 + \cdots + \lambda_N y_N;$$

i.e. the $j$th coordinate of $P^f_{\Delta}(y)$ 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 $\text{dom} f = E = \mathbb{R}^d$. Let $f(x) = \sum_i \lambda_i f(x_i)$ and $y \in \prod_i \text{int}(\text{dom} f) = E = (\mathbb{R}^d)^N$. Then $P^f_{\Delta}(y) = (z, \ldots, z)$, 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 $P^f_{\Delta}(y)$ 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 \( \text{dom } f = \{ x \in E = \mathbb{R}^d : x(j) \geq 0 \ \forall j \} \). Let \( f(x) = \sum_i \lambda_i f(x_i) \) and \( y \in \prod_i \text{int(dom } f) = \prod_i \{ x \in E = \mathbb{R}^d : x(j) > 0, \ \forall j \} \). Then \( P_{\Delta}^f(y) = (z, \ldots, z) \), where
\[
z(j) = \prod_i (y_i(j))^{\lambda_i}, \quad \text{for all } j;
\]
i.e. the \( j \)th coordinate of \( P_{\Delta}^f(y) \) 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
\[
\lim_{p \to 1} P_{\Delta}^{p \text{-norm}} = P_{\Delta}^{\text{Boltzmann/Shannon}}
\]
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 \( \text{dom } f = \{ x \in E = \mathbb{R}^d : x(j) \geq 0, \ \forall j \} \). Let \( f(x) = \sum_i \lambda_i f(x_i) \) and \( y \in \prod_i \text{int(dom } f) = \prod_i \{ x \in E = \mathbb{R}^d : x(j) > 0, \ \forall j \} \). Then \( P_{\Delta}^f(y) = (z, \ldots, z) \), where
\[
z(j) = \left( \frac{\lambda_1 (y_1(j))^p + \cdots + \lambda_N (y_N(j))^p}{\prod_i y_i(j)} \right)^{1/p}, \quad \text{for all } j;
\]
i.e. the \( j \)th coordinate of \( P_{\Delta}^f(y) \) 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 \( \text{dom } f = \{ x \in E = \mathbb{R}^d : x(j) > 0, \ \forall j \} \). Let \( f(x) = \sum_i \lambda_i f(x_i) \) and \( y \in \prod_i \text{int(dom } f) = \prod_i \{ x \in E = \mathbb{R}^d : x(j) > 0, \ \forall j \} \). Then \( P_{\Delta}^f(y) = (z, \ldots, z) \), 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 \( P_{\Delta}^f(y) \) 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
\[
\lim_{p \to 0} P_{\Delta}^{p \text{-root}} = P_{\Delta}^{\text{Burg}}
\]
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 \( \text{dom } f = [0, 1]^d \). Let \( f(x) = \sum_i \lambda_i f(x_i) \) and \( y \in \prod_i \text{int(dom } f) = ([0, 1]^d)^N \). Then \( P_{\Delta}^f(y) = (z, \ldots, z) \), where
\[
z(j) = \frac{\prod_i (x_i(j))^{\lambda_i}}{\prod_i (x_i(j))^{\lambda_i} + \prod_i (1 - x_i(j))^{\lambda_i}}, \quad \text{for all } j.
\]
Remarks 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 perturbation" of the $p$-norm (resp. the $p$-root) in the sense that

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

- The projection onto the diagonal $\Delta$ shares "mean"-like properties since the $j$th coordinate of $P^f_\Delta$ 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 $\varphi$ is strictly increasing by the strict convexity of $f$.

Examples 7.4, 7.6, and 7.9 also appear in Censor and Reich’s [10, Example 4.1]. 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).

The Hermitian matrices

Lewis recently demonstrated [19, 23, 21, 22] 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

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

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

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

For brevity, 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 $\lambda$ 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' [19, Corollary 2.7 and Theorem 2.6])
The induced function

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

is closed and convex.
is unitarily equivalent, i.e. \((f \circ \lambda)(U^*XU) = (f \circ \lambda)(X)\), \(\forall X \in H\), \(\forall U \in \mathcal{U}\). 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) \(\text{dom}(f \circ \lambda) = \lambda^{-1}(\text{dom } f)\).

(ii) \(\text{int}(\text{dom}(f \circ \lambda)) = \lambda^{-1}(\text{int}(\text{dom } f))\).

(iii) \(f \circ \lambda\) is Legendre if and only if \(f\) is.

**Proof.** (i) and (ii) follow from [22, Theorem 5.4]. \([19, \text{Corollary 3.3}]\) in tandem with \([19, \text{Corollary 3.5}]\) implies (iii). \(\square\)

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 \(\text{dom } f\) and \(\text{dom } f^*\) are open, then \(f \circ \lambda\) is Bregman/Legendre.

**Proof.** By Fact 7.14(i),(ii),
\[\text{int}(\text{dom}(f \circ \lambda)) = \lambda^{-1}(\text{int}(\text{dom } f)) = \lambda^{-1}(\text{dom } f) = \text{dom}(f \circ \lambda);\]
hence \(\text{dom}(f \circ \lambda)\) is open. Now \((f \circ \lambda)^* = f^* \circ \lambda\); thus (by similar reasoning), \(\text{dom}(f \circ \lambda)^*\) is open. Apply Theorem 5.6. \(\square\)

**Example 7.16** \((p\text{-norm})\) Suppose \(1 < p < +\infty\) and \(f(x) = \sum_j \frac{1}{p} |x_j|^p = \frac{1}{p}||x||_p^p\) on \(\text{dom } f = \text{int}(\text{dom } f) = E = \mathbb{R}^t\). Then \(f \circ \lambda\) is Bregman/Legendre and
\[(f \circ \lambda)(X) = \frac{1}{p}||\lambda(X)||_p^p = \sum_j \frac{1}{p} |\lambda_j(X)|^p,\]
for all \(X \in \text{dom}(f \circ \lambda) = H\).

The next example shows that the famous “logarithmic barrier function”, aka Burg’s entropy on \(H\), is Bregman/Legendre:

**Example 7.17** \(\text{("Logarithmic Barrier/Burg")}\) Suppose \(f(x) = \sum_j -\ln x_j\) on \(\text{dom } f = \{x \in E : x_j > 0, \forall j\}\). Then \(f \circ \lambda\) is Bregman/Legendre and
\[(f \circ \lambda)(X) = -\ln \det X,\]
for all \(X \in \text{dom}(f \circ \lambda) = \{X \in H : X \text{ positive semidefinite}\}\).
The best we could possibly hope for would be a result like

\[ f \circ \lambda \text{ is a convex function of Bregman/Legendre type if and only if } f \text{ is} \];

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 \( \Delta x \) for the \( J \times J \) diagonal matrix with diagonal entries \( x_1, \ldots, x_J \). Similarly, given a \( J \times J \) matrix \( X \) with diagonal entries \( X_{11}, \ldots, X_{JJ} \), we write \( \Delta(X) \) or \( \Delta X \) for the diagonal matrix with diagonal entries \( X_{11}, \ldots, X_{JJ} \) or for the vector in \( \mathbb{R}^J \) with components \( X_{11}, \ldots, X_{JJ} \). (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 \( Y \in \text{int}(\text{dom}(f \circ \lambda)) \). Then

\[ \begin{align*}
(\text{i}) \quad & \nabla(f \circ \lambda)(Y) = V(\Delta \nabla f(\lambda(Y)))V^*, \quad \forall V \in \mathcal{U} \text{ with } V^*V = \Delta \lambda(Y). \\
(\text{ii}) \quad & \nabla(f \circ \lambda)(U^*YU) = U^*\nabla(f \circ \lambda)(Y)U, \quad \forall U \in \mathcal{U}.
\end{align*} \]

\[ \text{(iii)} \quad \langle \nabla(f \circ \lambda)(U^*YU), U^*(X - Y)U \rangle = \langle \nabla(f \circ \lambda)(Y), X - Y \rangle, \quad \forall U \in \mathcal{U}, \; \forall X \in \mathcal{H}. \]

**Proof.** (i) follows from Lewis’ [19, Corollary 3.3].

(i) implies (ii), which in turn implies (iii). \( \square \)

**Remark 7.20** If \( Y \in \text{int}(\text{dom}(f \circ \lambda)) \) 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** \( D_{f \circ \lambda}(X, Y) = D_{f \circ \lambda}(U^*XU, U^*YU), \forall X \in \mathcal{H}, \forall Y \in \text{int}(\text{dom}(f \circ \lambda)), \forall U \in \mathcal{U}. \)

**Proof.**

\[ D_{f \circ \lambda} = (f \circ \lambda)(X) - (f \circ \lambda)(Y) - \langle \nabla(f \circ \lambda)(Y), X - Y \rangle \]

\[ = (f \circ \lambda)(U^*XU) - (f \circ \lambda)(U^*YU) - \langle \nabla(f \circ \lambda)(U^*YU), U^*(X - Y)U \rangle \]

\[ = D_{f \circ \lambda}(U^*XU, U^*YU). \]

\( \square \)

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

**Proposition 7.22** (Lewis [20]) Suppose \( X \in \mathcal{H} \) and \( \Delta X \) is decreasing: \( X_{11} \geq X_{22} \geq \cdots \geq X_{JJ} \). Then

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

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

The last supremum is attained for some $\tilde{P} \in \mathcal{P}$. By [19, Lemma 2.1], $\tilde{P}_z$ has precisely the components of $z$, but arranged decreasingly. Invoking [19, Lemma 2.1], Proposition 7.18, and a result due to von Neumann (see [19, Theorem 2.2]), we obtain

$$\langle \tilde{P}_z, \lambda(X) \rangle = \sup\{\langle z, \conv\{P\lambda(X) : P \in \mathcal{P}\} \rangle\} < \langle z, \Delta X \rangle \leq \langle \tilde{P}_z, \Delta X \rangle = \langle \Delta(\tilde{P}_z), \Delta(X) \rangle = \langle \Delta(P\lambda), \Delta(X) \rangle \leq \langle \lambda(\Delta(\tilde{P}_z)), \lambda(X) \rangle = \langle \tilde{P}_z, \lambda(X) \rangle,$$

which is the desired contradiction. 

\[\square\]

Corollary 7.23 Suppose $C \subseteq \mathbb{R}^d$ 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 $\Delta Y$. Thus, using Proposition 7.22,

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

hence $\Delta Y \in C$. 

\[\square\]

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

(i) 

$$D_{f \circ \lambda}(X, Y) = D_{f \circ \lambda}(X, \nabla(f \circ \lambda)\Delta \nabla(f \circ \lambda)(Y)) + (f \circ \lambda)(\nabla(f \circ \lambda)(Y)) - (f \circ \lambda)(\Delta \nabla(f \circ \lambda)(Y)).$$
(ii) 
\[ D_{f_\alpha}(\nabla(f^* \circ \lambda) \Delta \nabla(f \circ \lambda)(Y), Y) = (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) \\
- (f^* \circ \lambda)(\Delta \nabla(f \circ \lambda)(Y)) \\
= D_{f_\alpha}(\nabla(f \circ \lambda)(Y), \Delta \nabla(f \circ \lambda)(Y)). \]

(iii) 
\[ D_{f_\alpha}(X, Y) = D_{f_\alpha}(X, \nabla(f^* \circ \lambda) \Delta \nabla(f \circ \lambda)(Y)) \\
+ D_{f_\alpha}(\nabla(f^* \circ \lambda) \Delta \nabla(f \circ \lambda)(Y), Y). \]

Proof. \( Y \in \text{int}(\text{dom}(f \circ \lambda)) \) implies \( \nabla(f \circ \lambda)(Y) \in \text{int}(\text{dom}(f^* \circ \lambda)) = \lambda^{-1}(\text{int}(\text{dom} f^*)) \) by Fact 2.9, Fact 7.13, and Fact 7.14.(ii). Hence \( \Delta \nabla(f \circ \lambda)(Y) \in \lambda^{-1}(\text{int}(\text{dom} f^*)) \) = \( \text{int}(\text{dom}(f \circ \lambda)^*), \) by Corollary 7.23. Thus \( \nabla(f^* \circ \lambda) \Delta \nabla(f \circ \lambda)(Y) \in \text{int}(\text{dom}(f \circ \lambda)), \) 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):

\[ D_{f_\alpha}(X, Y) = (f \circ \lambda)(X) + (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - \langle \nabla(f \circ \lambda)(Y), X \rangle \\
= (f \circ \lambda)(X) + (f^* \circ \lambda)(\Delta \nabla(f \circ \lambda)(Y)) - \langle \Delta \nabla(f \circ \lambda)(Y), X \rangle \\
+ (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - (f^* \circ \lambda)(\Delta \nabla(f \circ \lambda)(Y)) \\
= D_{f_\alpha}(X, \nabla(f^* \circ \lambda) \Delta \nabla(f \circ \lambda)(Y)) \\
+ (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - (f^* \circ \lambda)(\Delta \nabla(f \circ \lambda)(Y)). \]

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 \( Y \in \text{int}(\text{dom}(f \circ \lambda)) \):

\[ D_{f_\alpha}(Y, \Delta Y) = (f \circ \lambda)(Y) - (f \circ \lambda)(\Delta Y) \geq 0; \]

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

We can thus interpret \( D_{f_\alpha}(Y, \Delta Y) \) 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:

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

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

**Proof.** (i): Consider $f(x) = \sum \frac{1}{p} |x_j|^p$ on $\text{dom} f = \mathbb{R}^J$.

(ii): Consider $f(x) = -\sum\frac{1}{p} x_j^p$ on $\text{dom} f = \{x \in \mathbb{R}^J : x_j \geq 0, \forall j\}$. Note that $\text{int}(\text{dom}(f \circ \lambda)) = \lambda^{-1}(\text{int}(\text{dom} f))$ is the positive definite matrices in $\mathcal{H}$. \hfill $\square$

**Example 7.28 (Hadamard's inequality; see, e.g., [24, Chapter 9 Section B])**

For all $Y \in \mathcal{H}$ that are positive definite:

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

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

**Proof.** Consider $f(x) = \sum \ln x_j$ on $\text{dom} f = \{x \in \mathbb{R}^J : x_j > 0 \forall j\}$. Then $(f \circ \lambda)(X) = -\ln \det X$ on the positive definite matrices in $\mathcal{H}$; see Lewis’ [19, Section 4]. \hfill $\square$

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

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

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

then $B \in \text{dom}(f \circ \lambda) \setminus \text{int}(\text{dom}(f \circ \lambda))$. Suppose further $(\lambda_n)$, $(\mu_n)$, $(s_n)$, $(c_n)$ 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, $V_n \to I$, $B_n \in \text{int}(\text{dom}(f \circ \lambda)) = \text{the positive definite matrices in} \ \mathcal{H}$, and $B_n \to B$. Also, for $X \in \text{dom}(f \circ \lambda), Y \in \text{int}(\text{dom}(f \circ \lambda))$,

$$D_{f \circ \lambda} = (f \circ \lambda)(X) + (f^* \circ \lambda)(\nabla(f \circ \lambda)(Y)) - \langle \nabla(f \circ \lambda)(Y), X \rangle$$

$$= \sum_{j=1}^{2} \lambda_j(X) \ln \lambda_j(X) - \lambda_j(X) + \sum_{j=1}^{2} \lambda_j(Y) - \langle \nabla(f \circ \lambda)(Y), X \rangle.$$
Thus

\[ D_{f \circ \lambda}(0, B_n) = \sum_{j=1}^{n} \lambda_j(B_n) = \text{tr}(B_n) = \lambda_n + \mu_n \to 1, \]

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/\( \lambda \)” 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 \( 2 \times 2 \) 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 \( 2 \times 2 \) matrices via

\[ T x = T(x_1, x_2, x_3) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix}, \quad \forall x = (x_1, x_2, x_3) \in \mathbb{R}^3. \]

The matrix \( T x \) is positive semi-definite if and only if \( x_1 \geq 0 \) and \( x_1 x_3 - x_2^2 \geq 0 \). The eigenvalues of \( T x \) in decreasing order are

\[
\lambda(T x) = \frac{1}{2}(x_1 + x_3 + \sqrt{(x_1 - x_3)^2 + 4x_2^2}, x_1 + x_3 - \sqrt{(x_1 - x_3)^2 + 4x_2^2}).
\]

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.

39
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 \( \langle X, Y \rangle = \text{tr}(XY^T), \forall X, Y \in \mathcal{S} \) (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

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

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:

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

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

Let \( r \) be a random mapping for \( \{1, \ldots, N\} \), i.e. a surjective 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

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

For instance, we could consider the random function \( r(n) \equiv n \bmod N \) (where we let the modN 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 \( \text{int}(\text{dom } f) \) and is thus well-defined (the interiority/zone consistency condition!).

40
Theorem 8.1 Suppose assumption (A) and (at least) one of the following conditions hold:

(i) $f$ is a Bregman/Legendre function.
(ii) $\bigcap_i C_i \cap \text{int}(\text{dom} f) \neq \emptyset$ and $\text{dom} f^*$ is open.
(iii) $\text{dom} f$ and $\text{dom} f^*$ are open.

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

Proof. (i): Proposition 3.16 yields

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

Hence

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

Fix any $c \in \bigcap_i C_i \cap \text{dom} f$ 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 $(y_n)$ 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: $\bar{y} \in \text{dom} f$ and $D_f(\bar{y}, y_{k_n}) \to 0$.

If $\bar{y} \in \text{int}(\text{dom} f)$, then apply Proposition 3.2(ii). If $\bigcap_i C_i \cap \text{int}(\text{dom} f) \neq \emptyset$, then pick any $c$ in this intersection; Property 1 then follows from 2 and Theorem 3.8(ii). Otherwise, $\bar{y} \in \text{bd}(\text{dom} f)$ 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

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

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

So let us assume to the contrary that $I_{out} \neq \emptyset$. 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 $r(m_n) \in I_{in}$. This is possible, since $\rho \in I_{in}$ and $I_{out}$ is assumed to be nonempty.
Then, by definition of \( m_n \), for every \( k_n \leq \nu \leq m_n \), \( r(\nu) \in I_n \); 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 \( y_{m_n} \to \bar{z} \in \text{dom} f \) with \( D_f(\bar{z}, y_{m_n}) \to 0 \) (Property 1). If \( \{\bar{y}, \bar{z}\} \cap \text{int}(\text{dom} f) \neq \emptyset \), then 3 and Theorem 3.9.(iii) imply \( \bar{y} = \bar{z} \). Otherwise \( \bar{y}, \bar{z} \in \text{dom} f \setminus \text{int}(\text{dom} f) \) 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 \( i \in I_{\text{in}} \); and that \( y_{m_n + i} \to \bar{z} \in \text{dom} f \cap C_i \). Note that by 2,

4. \[ D_f(y_{m_n+i}, y_{m_n}) \to 0. \]

Claim 2: \( \bar{y} = \bar{z} \).
If \( \bar{y} \) or \( \bar{z} \) is in the interior of \( \text{dom} f \), 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 \( y_n \to \bar{z} \). By Properties 1 through 3 (for \( (y_n) \) and \( \bar{z} \)),

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

If \( \{\bar{y}, \bar{z}\} \cap \text{int}(\text{dom} f) \neq \emptyset \), then \( \bar{y} = \bar{z} \) by Theorem 3.9.(iii). Otherwise, \( \bar{y}, \bar{z} \in \text{dom} f \setminus \text{int}(\text{dom} f) \) and \( D_f(\bar{y}, y_n) \to 0 \) (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 \( \bar{y} \in \text{int}(\text{dom} f) \) is derived and that BL1 through BL3 are not needed.

(iii): is then clear, since (iii) implies (ii). \( \square \)

**Remarks 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 $\bigcap_i C_i \cap \text{dom} f \neq \emptyset$ holds and $f$ is just Legendre with $\text{dom} f^*$ open. Here, the conditions on $f$ are readily verifiable whereas the constraint qualification requires some a priori knowledge.

Remarks 8.3

• The proof of Theorem 8.1 is an extension of the proof of the authors’ [3, Theorem 3.10]. We want to remark that the latter proof in turn relies on an idea developed almost simultaneously by Flam and Zowe [14], by Tseng [30], and by Elsner et al. [13] around the beginning of the decade; this idea is also present in Censor and Reich’s analysis [10].

• Though similar, the assumptions in Censor and Reich’s work [10] 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 [10] 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 straightforward; 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) $\bigcap_i C_i \cap \text{int}(\text{dom} f) \neq \emptyset$ and $\text{dom} f^*$ is open.
(iii) $\text{dom} f$ and $\text{dom} f^*$ are open.
Suppose further the set \( \{ z \in E : \nabla f(z) \in \text{span}(a_1, \ldots, a_N) \} \) 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

\[
\arg\min_{x \in \text{dom}\ f \cap \cap_i C_i} f(x).
\]

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

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

\[
\begin{align*}
f(y) - f(x) &= f(y) - f(x) - \langle \nabla f(y_n), y - x \rangle \\
&= D_f(y, y_n) - D_f(x, y_n) \\
&\leq D_f(y, y_n) \\
&\to 0.
\end{align*}
\]

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

**Remarks 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 \( \arg\min_{x \in \text{dom} f} f(x) \) is actually contained in \( \text{int}(\text{dom} f) \). 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 \( \nabla f(y_0) = 0 \in \text{span}(a_1, \ldots, a_N) \). 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.
<table>
<thead>
<tr>
<th>( f(x) )</th>
<th>( \text{argmin}_{x \in \text{dom} f} f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{2}</td>
<td>x</td>
</tr>
<tr>
<td>( x \ln x - x )</td>
<td>1</td>
</tr>
<tr>
<td>( x \ln x + (1 - x) \ln(1 - x) )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>( \frac{1}{p}</td>
<td>x</td>
</tr>
<tr>
<td>( -\sqrt{1 - x^2} )</td>
<td>0</td>
</tr>
</tbody>
</table>

- Related though somewhat different to Theorem 8.4 are the following:
  - Censor and Reich’s [10] (more general operators but the function that induces the Bregman distance has to be Bregman; see the Remarks in Section 4).

Acknowledgments

It is our pleasure to thank Adrian Lewis for extremely helpful discussions on the subsection on Hermitian matrices in Section 7, for the proof of Proposition 7.22, and for sending us [19, 23, 21, 22]. We also wish to thank Regina Burachik, Dan Butnariu, Yair Censor, and Alfredo Iusem for providing us with [5], [1], [6], and [17].

References


convex feasibility problems. Accepted for publication in SIAM Review. Also available
as Technical Report 93-12, Simon Fraser University, Burnaby, Canada.

[4] L.M. BREGMAN. The relaxation method of finding the common point of con-
veX sets and its application to the solution of problems in convex programming,
U.S.S.R. Computational Mathematics and Mathematical Physics, 7(3):200–217,
1967.

[5] R.S. BURACHIK. Generalized Proximal Point Methods of the Variational In-


July 1981.

[8] Y. CENSOR and A. LENT. Optimization of “log x” entropy over linear equality
1987.

IUSEM. On iterative methods for linearly constrained entropy maximization.
In A. Wàkulicz, editor, Numerical Analysis and Mathematical Modelling, pages
Volume 24.

[10] Y. CENSOR and S. REICH. Iterations of paracontractions and firmly nonexp-
pansive operators with applications to feasibility and optimization. Submitted,
1994.

nimization algorithm using Bregman functions. SIAM Journal on Optimization,

[12] J. ECKSTEIN. Nonlinear proximal point algorithms using Bregman functions,
with applications to convex programming. Mathematics of Operations Research,

and asynchronous nonlinear paracontractions. Numerische Mathematik, 62:305–

[14] S.D. FLÀM and J. ZOWE. Relaxed outer projections, weighted averages and

imization Algorithms II, volume 306 of Grundlehren der mathematischen Wiss-


manuscript.


