The method of forward projections

Heinz H. Bauschke*and Dominikus Noll†

November 14, 2001

Abstract

The convex feasibility problem asks to find a point in the intersection of finitely many closed convex sets. It is of basic importance in various areas of mathematics and physical sciences, and it can be solved iteratively using the classical method of cyclic projections, which generates a sequence by projecting cyclically onto the sets. In his seminal 1967 paper, Bregman extended this method to non-orthogonal projections, using the notion of the Bregman distance induced by a convex function.

In this paper, we present a new algorithmic scheme which also extends the method of cyclic projections. Based on Bregman distances, we introduce a new type of non-orthogonal projection, the forward projection. The energy and the negative entropy allow forward projections — the former yields the classical orthogonal projection whereas the latter gives rise to a type of projection used implicitly in a manifestation of the Expectation-Maximization algorithm. We provide useful properties of forward projections, and a basic convergence result on the method of forward projections.

2000 Mathematical Reviews Subject Classification.
Primary 90C25; Secondary 52A41, 65K05.
Key words. Bregman distance, Bregman function, Bregman/Legendre function, Bregman projection, convex feasibility problem, convex set, forward projection, joint convexity, Legendre function, projection.

---

*Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Email: `bauschke@cecm.sfu.ca`. Research supported by NSERC.

†Université Paul Sabatier, Laboratoire MIP, 118 route de Narbonne, 31 062 Toulouse Cedex, France. Email: `noll@mip.ups-tlse.fr`.

1 Introduction

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

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

Here, the points in the intersection are thought of as the set of *solutions* for a given problem, and the sets $C_1, \ldots, C_N$ represent some *constraints*. Assuming it is possible to compute the *orthogonal projection* P_i onto each constraint C_i, the classical *method of cyclic (orthogonal) projections* generates, given a starting point y_0, a sequence (y_n) by projecting cyclically onto the constraints:

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

The method is known to work in Euclidean space (and even in Hilbert space): indeed, the sequence (y_n) converges to a point in C. This result is due to Bregman [7]; see also [1, 10, 12, 20] for recent pointers to numerous extensions.

In 1967, Bregman [8] generalized his result to non-orthogonal projections, which can be constructed as follows: for a sufficiently well-behaved convex function f, define the so-called *Bregman distance* between two points x and y by

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

D_f is *not* a distance in the sense of metric topology; for instance, D_f is generally *not* symmetric. However, if $f = \frac{1}{2}\| \cdot \|^2$, then $D_f(x, y) = \frac{1}{2}\|x - y\|^2$, and one essentially recovers the orthogonal case. By taking infima, distances between points induce naturally distances between points and sets. Further, these distances to sets give rise to corresponding Bregman projections. In the *method of cyclic Bregman projections*, each orthogonal projection is simply replaced by the corresponding Bregman projection. For classical convergence results, see [8, 9]. This result has also found various extension; however, here we will be concerned with a variant of the following result, taken from [2]: there is no need to visit the sets in a cyclic fashion when generating the sequence (y_n); it suffices to pick up each constraint infinitely often — the resulting algorithmic scheme is the *method of random Bregman projections*.

In this paper, we extend the method of orthogonal projections in a new direction. The approach is similar yet different to the framework of Bregman projections: the similarity is the construction of a non-orthogonal projection using the Bregman distance; the difference is that our new type of projection is truly distinct from a Bregman projection. To explain this further, let us fix a set S. The classical Bregman projection of a point z onto the set S is the set of minimizers of the problem $\min_{s \in S} D_f(s, z)$. In contrast, we propose the *forward projection* of z onto S, defined as the solution of the optimization problem $\min_{s \in S} D_f(z, s)$. Existence and uniqueness of forward projections requires suitable assumptions on the underlying function f. As before, orthogonal projections can be recovered by setting $f = \frac{1}{2}\|\cdot\|^2$. (In fact, this is essentially the only way to make D_f symmetric.)

The aim of this paper is to analyze forward projections, and to provide a basic convergence result for the method of random forward projections.

The paper is organized as follows. In Section 2, we formulate our assumptions on f, followed by a discussion on verifiability. Covered by these assumptions are the perhaps two single most important functions in convex analysis: the *energy* and the *(negative) entropy*. One particular assumption on f, namely joint convexity of D_f, will imply that the *Bregman distance of the Bregman distance* is nonnegative, a quantity crucial to our analysis. Forward projections are defined and studied in some detail in Section 3. Our main result — the sequence generated by the method of random forward projections is convergent to a solution — is proven in Section 4.

2 Assumptions and some facts

The standing assumptions

From now on, we will assume that

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

satisfies the following:

A1 f is a convex function of Legendre type;

A2 f'' exists and is continuous on $\operatorname{int} \operatorname{dom} f$;

A3 D_f is jointly convex;

A4 $D_f(x,\cdot)$ is strictly convex on $\operatorname{int}\operatorname{dom} f$, $\forall x \in \operatorname{int}\operatorname{dom} f$;

A5 $D_f(x,\cdot)$ is coercive, $\forall x \in \operatorname{int}\operatorname{dom} f$.

Remark 2.1. Some comments on the assumptions are in order.

A1 The notion of a Legendre function is due to Rockafellar; we refer the reader to [23, Section 26] and also to [2] for basic properties and examples.

A2 In practice, this condition almost always holds.

A3 The joint convexity of D_f is discussed in detail in [3]; see also Remark 2.11 below.

A4 This is needed to make (forward) projections uniquely defined. In the usual separable setting, this is essentially a consequence of **A3**! See Remark 2.15 below.

A5 Since f is Legendre (assumption **A1**), we can state an equivalent condition [2, Corollary 3.1]: $\operatorname{dom} f^*$ is open.

The most important examples are the energy and the entropy — we state this now; however, we postpone the proof until a little later.

Example 2.2. The following functions satisfy **A1**–**A5**:

(i) $f(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}\sum_{j=1}^{J} |x_j|^2$ **(energy)**;

(ii) $f(x) = \sum_{j=1}^{J} x_j \ln(x_j) - x_j$ **(entropy)**.

The Bregman distance

Definition 2.3 (Bregman distance). Suppose $g : X \to \left]-\infty, +\infty\right]$ is differentiable on $\operatorname{int}\operatorname{dom} g \neq \emptyset$. Then the *Bregman "distance"* [8] is the map

$$D_g : X \times X : (x,y) \mapsto \begin{cases} g(x) - g(y) - \langle g'(y), x - y\rangle, & \text{if } y \in \operatorname{int}\operatorname{dom} g; \\ +\infty, & \text{otherwise.} \end{cases}$$

Example 2.4. Suppose x and y both belong to $\operatorname{int}\operatorname{dom} f$.

(i) If f is the **energy**, then $D_f(x,y) = \frac{1}{2}\|x-y\|^2$.

(ii) If f is the **entropy**, then $D_f(x,y) = \sum_{j=1}^J x_j \ln(x_j/y_j) - x_j + y_j$.

Remark 2.5. In the entropy case, the Bregman distance (restricted to the simplex) is widely known as the *Kullback-Leibler Information Divergence*; see, for instance, [18].

The next facts are basic and useful.

Fact 2.6. [23, Theorem 25.5] Suppose $g : X \to]-\infty, +\infty]$ is differentiable on $\operatorname{int}\operatorname{dom} g \neq \emptyset$. Then g' is continuous on $\operatorname{int}\operatorname{dom} g$. In particular, D_g is continuous on $(\operatorname{int}\operatorname{dom} g)^2$.

Fact 2.7 (Three-Point Identity). [11, Lemma 3.1] Suppose $x \in \operatorname{dom} f$ and $y, z \in \operatorname{int}\operatorname{dom} f$. Then $D_f(x,y) + D_f(y,z) - D_f(x,z) = \langle f'(y) - f'(z), y - x\rangle$.

Proposition 2.8. Suppose $g : X \to]-\infty, +\infty]$ is differentiable on $\operatorname{int}\operatorname{dom} g$. Suppose further that x, y are two points with $[x,y] \subseteq \operatorname{int}\operatorname{dom} g$. Then:

(i) $g|_{[x,y]}$ is convex $\Leftrightarrow D_g(x,y) \geq 0$.

(ii) $g|_{[x,y]}$ is affine $\Leftrightarrow D_g(x,y) = 0$.

Proof. This is essentially [22, Theorem 42.A]; see also [3, Section 2]. $\square$

The Bregman distance of the Bregman distance

Assumption **A3** on joint convexity of D_f makes the *Bregman distance of the Bregman distance* meaningful. The following two results are new and important to our analysis.

Lemma 2.9 (Key identity). The map D_{D_f} is nonnegative and continuous on $(\operatorname{int}\operatorname{dom} f)^4$. Given four points x, y, x_0, y_0 in $\operatorname{int}\operatorname{dom} f$, we have:

$$\begin{aligned} D_{D_f}\big((x,y),(x_0,y_0)\big) &= D_f(x,y) + D_f(x,x_0) - D_f(x,y_0) \\ &\quad + \langle f''(y_0).(x_0 - y_0), y - y_0\rangle. \end{aligned}$$

Proof. Because D_f is jointly convex, the nonnegativity of D_{D_f} on $(\operatorname{int}\operatorname{dom} f)^4$ is clear from Proposition 2.8.(i) (with $g = D_f$). Note that the derivative of D_f at (x_0, y_0) is $\big(f'(x_0) - f'(y_0), f''(y_0).(y_0 - x_0)\big)$; this and Fact 2.7 yield immediately the key identity. Now D_f is continuous on $(\operatorname{int}\operatorname{dom} f)^2$ (by Fact 2.6), and f'' is continuous on $\operatorname{int}\operatorname{dom} f$ (by assumption). The desired continuity of D_{D_f} thus follows from the key identity. $\square$

Lemma 2.10. Suppose c, x, y are in $\operatorname{int}\operatorname{dom} f$ with $D_{D_f}\big((c,c),(x,y)\big) = 0$. Then $x = y$.

Proof. By Proposition 2.8.(ii), D_f is affine on $\big[(c,c),(x,y)\big]$. This means

$$\begin{aligned} D_f\big((1-t)(c,c)+t(x,y)\big) &= (1-t)D_f(c,c)+tD_f(x,y) \\ &= tD_f(x,y), \end{aligned}$$

for every $t \in [0,1]$. By definition of D_f, we have

$$\begin{aligned} D_f\big((1-t)(c,c)+t(x,y)\big) &= D_f\big((1-t)c+tx,(1-t)c+ty\big) \\ &= f\big((1-t)c+tx\big) - f\big((1-t)c+ty\big) \\ &\quad - \big\langle f'\big((1-t)c+ty\big), t(x-y)\big\rangle. \end{aligned}$$

Altogether, we obtain

$$\begin{aligned} tD_f(x,y) &= tf(x) - tf(y) - t\langle f'(y), x-y\rangle \\ &= \Big[f\big(c+t(x-c)\big) - f(c)\Big] - \Big[f\big(c+t(y-c)\big) - f(c)\Big] \\ &\quad - t\big\langle f'\big(c+t(y-c)\big),(x-y)\big\rangle. \end{aligned}$$

Now divide by $t \in \,]0,1[$ and let tend t to 0 from above. We deduce

$$D_f(x,y) = f'(c)(x-c) - f'(c)(y-c) - f'(c)(x-y) = 0.$$

Therefore, $x = y$, and the proof is complete. $\square$

Remark 2.11. The joint convexity of D_f is discussed in some detail in [3]. If f is (separable, or even) of the form $f(x) = \sum_{j=1}^{J} \varphi(x_j)$, then the problem is essentially one-dimensional. Assume that φ is four times differentiable. Then, by [3, Theorem 3.3.(i)],

$$\begin{aligned} & D_f \text{ is jointly convex} \\ \Leftrightarrow\quad & \varphi''\varphi'''' \geq 2(\varphi''')^2 \\ \Leftrightarrow\quad & \varphi''(s) + \varphi'''(s)(s-r) \geq \big(\varphi''(s)\big)^2/\varphi''(r), \text{ for all } r, s \text{ in } \operatorname{int}\operatorname{dom}\varphi. \end{aligned}$$

It is straight-forward to see that the energy and the entropy pass the condition involving φ''''; moreover, as pointed out in [3, Remark 3.6], these two functions are limiting cases in the class of convex functions with jointly convex Bregman distances in the sense that they make the inequality even an equality!

For the energy and the entropy, this can be made short and explicit:

Example 2.12. Suppose x, y, x_0, y_0 all belong to $\operatorname{int}\operatorname{dom} f$. Then:

(i) If f is the **energy**, then $D_{D_f}\big((x,y),(x_0,y_0)\big) = D_f\big(x, y + (x_0 - y_0)\big)$.

(ii) If f is the **entropy**, then $D_{D_f}\big((x,y),(x_0,y_0)\big) = D_f\big(x, y \cdot \frac{x_0}{y_0}\big)$.

Remark 2.13. We note that Example 2.12 yields — in conjunction with Proposition 2.8 — the joint convexity of the Bregman distance induced by the energy and the entropy with a proof distinct from the characterization mentioned in Remark 2.11. Furthermore, since both the energy and the entropy are strictly convex on the interior of their respective domains, Example 2.12 also results in a direct proof of Lemma 2.10.

Strict convexity of $D_f(x,\cdot)$

Proposition 2.14. Suppose the map $\operatorname{int}\operatorname{dom} f \to X : y \mapsto f''(y).(y-x)$ is strictly monotone, for all $x \in \operatorname{int}\operatorname{dom} f$. Then $D_f(x,\cdot)$ is strictly convex, for every $x \in \operatorname{int}\operatorname{dom} f$.

Proof. Fix $x \in \operatorname{int}\operatorname{dom} f$. The derivative of the map $y \mapsto D_f(x,y)$ is easily seen to be $y \mapsto f''(y).(y-x)$, which is strictly monotone by assumption. Strict convexity thus follows (from [22, Theorem 42B], for instance). ☐

Remark 2.15. It is worthwhile to visit the (separable) case when $f(x) = \sum_{j=1}^{J} \varphi(x_j)$, for some function φ defined on $\mathbb{R}$. Assume that φ is even three times differentiable. The strict monotonicity condition of Proposition 2.14 — and hence strict convexity of $D_f(x,\cdot)$ — then boils down to strict positivity of the derivative of the map $s \mapsto \varphi''(s)(s-r)$:

$$(\forall r \in \operatorname{int}\operatorname{dom}\varphi)(\forall s \in \operatorname{int}\operatorname{dom}\varphi) \quad \varphi'''(s)(s-r) + \varphi''(s) > 0.$$

Now we notice something immensely convenient: in this separable setting, the last strict inequality is always satisfied because of joint convexity of D_f (assumption **A3**) and Remark 2.11! In passing, we note that one can thus view **A4** not only as *less restrictive than joint convexity of* D_f, but also as *more restrictive than separate convexity of* D_f (by [3, Theorem 3.3.(ii)]). Moreover, these observations apply to the energy as well as the entropy, and they yield **A4** for these two functions.

More examples

Example 2.16. The following functions satisfy the assumptions **A1**–**A5**:

(i) the **energy** $f(x) = \sum_{j=1}^{J} x_j^2$;

(ii) the **entropy** $f(x) = \sum_{j=1}^{J} x_j \ln(x_j) - x_j$.

(iii) the **Fermi-Dirac entropy** $f(x) = \sum_{j=1}^{J} x_j \ln(x_j) + (1 - x_j) \ln(1 - x_j)$.

Proof. We start by noticing that every function is separable and infinitely differentiable on its domain — this will make life much simpler: **A2** clearly holds, for instance.

A1: Legendreness of (i), (ii), and (iii) is known, see [2, Section 6].

A3: Joint convexity of D_f was established in [3].

A4: Clear, since our setting is separable and so Remark 2.15 applies.

A5: In view of Remark 2.1, we need only to check that $\operatorname{dom} f^*$ is open. But this is clear for the three functions [2, Section 6]: the conjugate of f in (i), (ii), and (iii) respectively is (the separable extension of) (i) $r \mapsto \frac{1}{2}r^2$, (ii) exp, and (iii) $r \mapsto \ln(1 + \exp(r))$. So the domain of each conjugate is the entire space (which is open). ☐

We conclude this section by collecting some properties that will be useful later.

Fact 2.17. Suppose $x \in \operatorname{int} \operatorname{dom} f$ and (y_n) is a sequence in $\operatorname{int} \operatorname{dom} f$. If $\big(D_f(x, y_n)\big)$ is bounded, then the sequence (y_n) is bounded, and all its cluster points belong to $\operatorname{int} \operatorname{dom} f$.

Proof. [2, Theorem 3.7.(vi) and Theorem 3.8.(ii)]. ☐

Fact 2.18. Suppose $y \in \operatorname{int} \operatorname{dom} f$ and (y_n) is a sequence in $\operatorname{int} \operatorname{dom} f$. Then: $y_n \to y$ if and only if $D_f(y, y_n) \to 0$.

Proof. "$\Rightarrow$": [2, Proposition 3.2.(ii)] or an easy direct verification.

"$\Leftarrow$": By Fact 2.17, the sequence (y_n) is bounded and has all its cluster points in $\operatorname{int} \operatorname{dom} f$. Now suppose to the contrary that (y_n) does not converge to y. After passing to a subsequence if necessary, we may and do assume that $y_n \to z \in \operatorname{int} \operatorname{dom} f$. Since D_f is continuous on $\operatorname{int} \operatorname{dom} f \times \operatorname{int} \operatorname{dom} f$, it follows that $0 \leftarrow D_f(y, y_n) \to D_f(y, z)$. Hence, by essential strict convexity of f, $y = z$ and the proof is complete. ☐

Fejér monotonicity with respect to D_f

The classical notion of a Fejér monotone sequence (taken with respect to the Euclidean metric) has been found tremendously useful in the analysis of algorithms [1, 10, 13, 14]. We will need the following variant, tailored for the Bregman distance:

Definition 2.19. Suppose S is a set with $S \cap \operatorname{dom} f \neq \emptyset$ and $(y_n)_{n\geq 0}$ is a sequence in $\operatorname{int}\operatorname{dom} f$. Then (y_n) is *Fejér monotone with respect to S* if

$$(\forall s \in S)(\forall n \geq 0) \qquad D_f(s, y_n) \geq D_f(s, y_{n+1}).$$

Lemma 2.20. Suppose S is a set with $S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and (y_n) is a sequence in $\operatorname{int}\operatorname{dom} f$ that is Fejér monotone with respect to $S \cap \operatorname{int}\operatorname{dom} f$. Then (y_n) converges to some point in $S \cap \operatorname{int}\operatorname{dom} f$ if and only if all cluster points of (y_n) lie in $S \cap \operatorname{int}\operatorname{dom} f$.

Proof. The "only if" part is clear. To prove the "if" part, observe first that (y_n) is bounded and all its cluster points belong to $\operatorname{int}\operatorname{dom} f$ (by Fact 2.17). Let s_1 and s_2 be two cluster points of (y_n), both belonging to $S \cap \operatorname{int}\operatorname{dom} f$. It suffices to show that $s_1 = s_2$. By Fejér monotonicity, the sequences $\big(D_f(s_1, y_n)\big)$ and $\big(D_f(s_2, y_n)\big)$ converge; hence, so does the sequence of differences, whose terms we can write (using Fact 2.7) as

$$D_f(s_1, y_n) - D_f(s_2, y_n) = \langle f'(s_1) - f'(y_n), s_1 - s_2 \rangle - D_f(s_2, s_1).$$

It follows that $\lambda := \lim_n \langle f'(s_1) - f'(y_n), s_1 - s_2 \rangle$ exists. Taking the limit along the subsequence converging to s_1 (resp. s_2) yields $\lambda = 0$ (resp. $\lambda = \langle f'(s_1) - f'(s_2), s_1 - s_2 \rangle$). (This step is justified by the continuity of f'; see Fact 2.6.) Altogether,

$$\langle f'(s_1) - f'(s_2), s_1 - s_2 \rangle = 0.$$

Consequently, by essential strict convexity of f, $s_1 = s_2$ as required. $\square$

Remark 2.21. Imposing that f be *Bregman* [9, 10] or *Bregman/Legendre* [2], we could extend Lemma 2.20 to the case where the sequence (y_n) converges to a point in $\operatorname{dom} f \setminus \operatorname{int}\operatorname{dom} f$.

3 Forward projections

Definition 3.1 (Forward projection). Suppose S is a closed convex set with $S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and $x \in \operatorname{dom} f$. Then the set

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

is called the *forward projection of* x *onto* S and denoted $\overrightarrow{P}_S(x)$. If $\overrightarrow{P}_S(x)$ is a singleton, say $\{y\}$, we will write $\overrightarrow{P}_S(x) = y$ in a slight (but convenient) abuse of notation.

Remark 3.2. The direction of the arrow points forward (to the right), this is meant to help the reader remember that we vary over the right-hand variable y in $D_f(x, y)$ when we compute the projection. The arrow notation is required to distinguish the new projection from its cousin, the *classical Bregman projection* [8]: given S with $S \cap \operatorname{dom} f \neq \emptyset$ and $y \in \operatorname{int}\operatorname{dom} f$, the set

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

is the *(backward) Bregman projection of* y *onto* S and denoted $P_S(y)$. See, e.g., [2] and [10] for further properties and examples.

Is the notion of a forward projection really a new one? The answer is affirmative, as the next result shows.

Proposition 3.3 (Reid). [21] Let f be the entropy on $\mathbb{R}$. Then there is no function $g : \mathbb{R} \to\,]-\infty, +\infty]$ such that $D_f(x, y) = D_g(y, x)$, for all $x, y > 0$.

Proof. We argue by contradiction and thus assume that there exists a function g with

$$(*) \qquad (\forall x > 0)(\forall y > 0) \quad g(y) - g(x) - g'(x)(y - x) = x \ln(x/y) - x + y.$$

Taking the derivative with respect to y yields $g'(y) - g'(x) = 1 - x/y$, for all $x, y > 0$. Assume $y \neq x$, and divide the last equality by $y - x$. Then $\big(g'(y) - g'(x)\big)/\big(y - x\big) = 1/y$. Fix $x > 0$ and take the limit as y tends to x. Then $g''(x) = 1/x$, $\forall x > 0$. On the other hand, consider again $(*)$:

$$(\forall x > 0)(\forall y > 0) \quad g(y) = g(x) + g'(x)(y - x) + x \ln(x/y) - x + y.$$

Take the derivative with respect to x yields $0 = g''(x)(y - x) + \ln(x/y)$. Now taking the derivative with respect to y results in $0 = g''(x)$. Altogether, we obtained the absurdity $0 = g''(x) = 1/x$. $\square$

Remark 3.4.

(i) We discovered Proposition 3.3 first in Maple by arguing also by contradiction: if g exists so that $(*)$ holds true, then fix $y = 1$. The resulting differential equation has the solution

$$g(x) = g(1) - x\ln(x) + (1-x)\operatorname{dilog}(x) + c(1-x),$$

where c is a real constant, and $\operatorname{dilog}(x) := \int_1^x \frac{\ln(t)}{1-t}\,dt$. By construction, $D_g(1,x) = D_f(x,1)$, $\forall x > 0$. However, the difference $D_g(2,x) - D_f(x,2) = \operatorname{dilog}(x) + x\ln(2) + \pi^2/12 - \ln(4)$ is zero only when $x = 2$. Therefore, there is no function g with the desired properties.

(ii) Greg Reid's proof of Proposition 3.3 is much preferable to our previous reasoning; nonetheless, the function

$$g(x) = g(1) - x\ln(x) + (1-x)\operatorname{dilog}(x) + c(1-x),$$

just encountered has some quite intriguing properties (which we state here without proof as they are not needed elsewhere in the paper): D_g is actually jointly convex, but g is not Legendre. In contrast, the (alternating) derivatives $(-1)^n \cdot g^{(n)}$ appear to be Legendre, but without generating a jointly convex Bregman distance. The simplest explicit case is $g^{(2)}(x) = g''(x) = \ln(x)/(x-1)$.

(iii) Greg Reid pointed out to us that his proof can actually be reproduced entirely in Maple using Allan Wittkopf's package `rifsimp`, which detects that the corresponding PDE system is inconsistent — this output means that the argument is actually rigorous!

Lemma 3.5 (Existence and uniqueness). Suppose S is a closed convex set with $S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and $x \in \operatorname{int}\operatorname{dom} f$. Then $\overrightarrow{P}_S(x)$ is a singleton.

Proof. Pick a sequence (y_n) in $S \cap \operatorname{int}\operatorname{dom} f$ such that

$$D_f(x, y_n) \to \inf_{y \in S \cap \operatorname{int}\operatorname{dom} f} D_f(x,y).$$

Clearly, $\big(D_f(x,y_n)\big)$ is bounded. By Fact 2.17, the sequence (y_n) is bounded and all its cluster points belong to $S \cap \operatorname{int}\operatorname{dom} f$. It follows that $\emptyset \neq \overrightarrow{P}_S(x) \subseteq \operatorname{int}\operatorname{dom} f$. By joint convexity of D_f, $\overrightarrow{P}_S(x)$ is convex subset of $\operatorname{int}\operatorname{dom} f$. Furthermore, we assumed that the map $D_f(x,\cdot)$ is strictly convex on $\operatorname{int}\operatorname{dom} f$. Altogether, $\overrightarrow{P}_S(x)$ must be a singleton. $\square$

Lemma 3.6 (Characterization of forward projection). Suppose S is a closed convex set with $S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and let x_0, y_0 be in $\operatorname{int}\operatorname{dom} f$. Then:

$$y_0 = \overrightarrow{P}_S(x_0) \quad \Leftrightarrow \quad y_0 \in S \text{ and } \langle S - y_0, f''(y_0).(x_0 - y_0)\rangle \leq 0.$$

Proof. The derivative of the map $y \mapsto D_f(x_0, y)$ is $y \mapsto f''(y).(y - x_0)$. Hence the characterization is nothing but the optimality condition for a constrained optimization problem; see [23, Theorem 27.4] or [6, Section 2.1] □

The following result is crucial.

Corollary 3.7 (Key inequality and continuity). Suppose S is a closed convex set with $S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and let $s \in S$. Suppose further $x, \bar{x}$ both belong to $\operatorname{int}\operatorname{dom} f$. Then

$$D_f(x, s) + D_f(x, \bar{x}) \geq D_f(x, \overrightarrow{P}_S(\bar{x})) + D_{D_f}\big((x, s), (\bar{x}, \overrightarrow{P}_S(\bar{x}))\big),$$

and all terms in this inequality are nonnegative. Moreover, the map $\overrightarrow{P}_S$ is continuous on $\operatorname{int}\operatorname{dom} f$.

Proof. The inequality follows from Lemma 2.9 and Lemma 3.6. The nonnegativity of the D_f terms is clear (by convexity of f and Proposition 2.8.(i)). And the nonnegativity of the D_{D_f} term was already observed in Lemma 2.9. Now let (x_n) be a sequence in (without loss of generality) $\operatorname{int}\operatorname{dom} f$ converging to $\bar{x}$. For simplicity, set $\bar{s} := \overrightarrow{P}_S(\bar{x})$ and $s_n := \overrightarrow{P}_S(x_n)$, for all n. Fact 2.18 and another application of Lemma 2.9 and Lemma 3.6 yield

$$\begin{aligned} D_f(\bar{x}, \bar{s}) &\leftarrow D_f(\bar{x}, \bar{s}) + D_f(\bar{x}, x_n) \\ &\geq D_f(\bar{x}, s_n)) + D_{D_f}\big((\bar{x}, \bar{s}), (x_n, s_n)\big) \\ &\geq D_f(\bar{x}, s_n). \end{aligned}$$

Hence $\big(D_f(\bar{x}, s_n)\big)$ is a bounded sequence. By Fact 2.17, the sequence (s_n) is bounded, and all its cluster points lie in $\operatorname{int}\operatorname{dom} f$. We must show that $s_n \to \bar{s}$. We argue by contradiction: after passing to a subsequence if necessary, we assume that $s_n \to \hat{s} \in S \cap \operatorname{int}\operatorname{dom} f$ with $\hat{s} \neq \bar{s}$. Clearly, $D_f(\bar{x}, \cdot)$ is continuous on $\operatorname{int}\operatorname{dom} f$. Thus the displayed inequalities yield, after taking limits, the inequality $D_f(\bar{x}, \bar{s}) \geq D_f(\bar{x}, \hat{s})$. By uniqueness of $\overrightarrow{P}_S(\bar{x}) = \bar{s}$, we conclude $\hat{s} = \bar{s}$, which is absurd. The proof is complete. □

Remark 3.8 (forward vs backward projection). Fix $x \in \operatorname{int}\operatorname{dom} f$ and $s \in S \cap \operatorname{int}\operatorname{dom} f$. If y denotes the forward projection of x onto S, then (by Corollary 3.7)

$$D_f(s,x) \geq D_f(s,y) + D_{D_f}\big((s,s),(x,y)\big).$$

In contrast, if $\tilde{y}$ denotes the backward projection of x onto S, then (see, for instance, [2, Proposition 3.16]):

$$D_f(s,x) \geq D_f(s,\tilde{y}) + D_f(\tilde{y},x).$$

Thus both types of projections have similar inequalities for their respective projections.

Remark 3.9. It is instructive to view the question on continuity of $\overrightarrow{P}_S$ in the context of *variational inequalities* or *well-posed optimization problems* — the results from these two areas yield continuity of $\overrightarrow{P}_S$ under additional assumptions such as compactness of S. For further information, we refer the reader to [16, Section IX.1] and [17, Section 5].

Remark 3.10. Corollary 3.7 is not only interesting in its own right, but also useful in other contexts involving forward projections. For example, the *Expectation-Maximization algorithm* for a particular Poisson model can be viewed as an alternating backward/forward projection algorithm (see [15] and also [19]); Corollary 3.7 then yields new asymptotic results [4].

Forward projections onto hyperplanes

Assume now $x \in \operatorname{int}\operatorname{dom} f$, $H = \{z \in X : \langle a, z\rangle = \beta\}$ is a hyperplane, where $a \in X$ and $\beta \in \mathbb{R}$, and assume further that $H \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$.

How do we find $\overrightarrow{P}_H(x)$? Using Lemma 3.6, it is easy to see that $y = \overrightarrow{P}_H(x)$ precisely when

$$\text{(H)} \qquad f''(y).(x-y) = ra \quad \text{and} \quad \langle y, a\rangle = \beta,$$

for some $r \in \mathbb{R}$.

Solving (H) in general appears to be difficult. However, for the energy and the entropy, much more can be said.

Example 3.11. Suppose $x \in \operatorname{int}\operatorname{dom} f$ and $H = \{z \in X : \langle a, z\rangle = \beta\}$ is a hyperplane with $H \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$.

(i) If f is the **energy**, then

$$\overrightarrow{P}_H(x) = x - \frac{\langle a, x\rangle - b}{\|a\|^2} a.$$

(ii) If f is the **entropy**, then $\overrightarrow{P}_H(x)$ can be found in two steps:

1. Solve $\sum_j a_j x_j/(1 + ra_j) = \beta$ for $r \in \mathbb{R}$.
2. Set $y_j = x_j/(1 + ra_j)$, $\forall j$.

Proof. Let $y = \overrightarrow{P}_H(x)$.

(i): for the energy, we have $f'' \equiv I$. By (H), $y = x - ra$. Take the inner product of this equality with a, use $\langle a, y\rangle = \beta$, and solve for r to obtain the desired formula.

(ii): for the entropy, $f''(y)$ is the diagonal matrix with entries $1/y_j$. Hence (H) yields

$$\frac{1}{y_j}(x_j - y_j) = ra_j; \quad \text{equivalently, } y_j = \frac{x_j}{1 + ra_j},$$

for every j. Since $y \in H$, we have $\langle a, y\rangle = \beta$ or

$$\sum_{j=1}^{J} \frac{a_j x_j}{1 + ra_j} = \beta.$$

The result follows. □

Remark 3.12 (More on the entropy projection onto a hyperplane).

(i) Simple calculus shows that the function $S(r) := \sum_j a_j x_j/(1 + ra_j)$ is strictly decreasing. Also, from Step 2 of Example 3.11.(ii), the sought-after r is bounded by

$$\sup_{j:a_j>0} \frac{-1}{a_j} < r < \inf_{j:a_j<0} \frac{-1}{a_j},$$

where we use the usual convention $\inf \emptyset = +\infty$ and $\sup \emptyset = -\infty$. Hence even a simple bisection algorithm can be used to determine r efficiently.

(ii) While there is in general no closed form for $\overrightarrow{P}_H(x)$, the "inverse projection" is explicit: fix $y \in H$. Then, for every $x \in \operatorname{int}\operatorname{dom} f$, we have (from Step 2 in Example 3.11.(ii)) the characterization $\overrightarrow{P}_H(x) = y$ if and only if $x_j = (1+ra_j)y_j > 0$, $\forall j$. Thus the inverse projection results in lines. (In contrast, the "inverse backward projection" is generally a nonlinear curve.)

(iii) It is not hard to see that $\overrightarrow{P}_H(x)$ can be computed *explicitly* in two cases: when $a \in \mathbb{R}\mathbf{1}$; or when $J = 2$, i.e., we work in the Euclidean plane (so that Step 1 of Example 3.11.(ii) becomes solving a quadratic equation).

Remark 3.13 (Forward projection of a smooth set). Suppose S is a closed convex set with $\operatorname{int} S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$. Recall S is *smooth* at a point $y \in \operatorname{bdry} S$, if the tangent cone of S at y is a halfspace. Two equivalent conditions are: the normal cone of S at y is a ray, or there exists a unique supporting hyperplane to S at y. Now suppose S is *smooth*, i.e., smooth at every boundary point. Let x and y both belong to $\operatorname{int}\operatorname{dom} f$, where $x \notin S$. Using Lemma 3.6, it is not hard to show that

$$y = \overrightarrow{P}_S(x) \Leftrightarrow y = \overrightarrow{P}_H(x), \text{ and } H \text{ supports } S \text{ at } y.$$

(An analogous result holds for the backward Bregman projection.)

The combination of Remark 3.13 and Remark 3.12 results in the following:

Remark 3.14 (Inverse entropy projection of a smooth set). Suppose f is the **entropy**, S is a closed convex smooth set with $\operatorname{int} S \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and $y \in \operatorname{bdry} S \cap \operatorname{int}\operatorname{dom} f$. Then the inverse projection of y can be found as follows:

(i) Find the (unique) supporting hyperplane H to S at y.

(ii) Let a be the (unique) unit vector that is normal to H and points outwards S.

(iii) Then $x \in \operatorname{int}\operatorname{dom} f$ satisfies $\overrightarrow{P}_S(x) = y$ if and only if

$$x_j = y_j(1 + ra_j), \quad \forall j,$$

for some $r \geq 0$.

In passing, we mention the analogous result for the inverse (backward) projection: the *backward projection* of a point $x \in \operatorname{int}\operatorname{dom} f$ is equal to y if and only if

$$x_j = y_j \exp(r a_j), \quad \forall j,$$

for some $r \geq 0$. Thus the inverse projection is a curve, nonlinear in general.

4 Main result

From now on, we assume that $C_1, \ldots, C_N$ are finitely many closed convex sets such that

$$\boxed{C := \bigcap_{i=1}^{N} C_i \quad \text{and} \quad C \cap \operatorname{int}\operatorname{dom} f \neq \emptyset.}$$

Let $r : \mathbb{N} \to \{1, \ldots, N\}$ be a *random map*: r is onto and it assume each value in $\{1, \ldots, N\}$ infinitely often.

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

$$\boxed{y_0 \in \operatorname{int}\operatorname{dom} f, \qquad \text{and } y_{n+1} := \overrightarrow{P}_{C_{r(n+1)}}(y_n), \quad \forall n \geq 0.}$$

Theorem 4.1. For an arbitrary starting point y_0, the sequence (y_n) generated by the method of random forward projections converges to some point $\bar{y} \in C \cap \operatorname{int}\operatorname{dom} f$.

Proof. We proceed in several steps.

`Claim 1`: $D_f(c, y_n) \geq D_f(c, y_{n+1}) + D_{D_f}\big((c, c), (y_n, y_{n+1})\big)$,
$\forall n \geq 0$, $\forall c \in C_{r(n+1)} \cap \operatorname{int}\operatorname{dom} f$.
This follows from Corollary 3.7 (with $S = C_{r(n+1)}$, $x := s := c$, and $x_0 := y_n$).

`Claim 2`: (y_n) is Fejér monotone with respect to $C \cap \operatorname{int}\operatorname{dom} f$.
This is immediate from `Claim 1`.

`Claim 3`: (y_n) is a bounded, and its cluster points belong to $\operatorname{int}\operatorname{dom} f$.
Pick $c \in C \cap \operatorname{int}\operatorname{dom} f$. By `Claim 2`, the sequence $\big(D_f(c, y_n)\big)$ is decreasing and thus bounded. `Claim 3` thus follows from Fact 2.17.

Next, suppose that

$$\bar{y} \text{ is a cluster point of } (y_n), \text{ say } y_{k_n} \to \bar{y}.$$

`Claim 4`: $\bar{y} \in \operatorname{int}\operatorname{dom} f$, and $D_f(\bar{y}, y_{k_n}) \to 0$.
Clear from `Claim 3` and Fact 2.18.

After passing to a subsequence if necessary, we may assume that $r(k_n) \equiv \rho \in \{1, \dots, N\}$, and thus $\bar{y} \in C_\rho$.

We now define

$$I_{\text{in}} := \{i : \bar{y} \in C_i\} \text{ and } I_{\text{out}} := \{i : \bar{y} \notin C_i\}.$$

`Claim 5`: $I_{\text{out}} = \emptyset$.

We prove this by contradiction and thus assume that $I_{\text{out}} \neq \emptyset$. After passing to another subsequence if necessary, we may further assume that $\{r(k_n), r(k_n+1), \dots, r(k_{n+1}-1)\} = \{1, \dots, N\}$ — this is possible, because r is a random map. Since $\rho \in I_{\text{in}}$, we can pick m_n in $\{k_n, k_n+1, \dots, k_{n+1}-1\}$ maximal with $r(m_n) \in I_{\text{in}}$. By repeated use of `Claim 1`, we obtain

$$D_f(\bar{y}, y_{m_n}) \leq D_f(\bar{y}, y_{k_n}), \quad \forall n.$$

This, together with `Claim 4`, yields $D_f(\bar{y}, y_{m_n}) \to 0$. Hence, using Fact 2.18, $y_{m_n} \to \bar{y}$. After passing to a final subsequence if necessary, we assume that $r(m_n+1) \equiv \sigma \in I_{\text{out}}$ and that $y_{m_n+1} \to \bar{z} \in C_\sigma \cap \operatorname{int} \operatorname{dom} f$ (using `Claim 3` and closedness of C_σ). Fix $c \in C \cap \operatorname{int} \operatorname{dom} f$. By `Claim 1`,

$$D_{D_f}\big((c,c),(y_{m_n}, y_{m_n+1})\big) \to 0.$$

In Lemma 2.9, we observed the continuity of D_{D_f} on $(\operatorname{int} \operatorname{dom} f)^4$. Thus

$$D_{D_f}\big((c,c),(\bar{y}, \bar{z})\big) = 0.$$

Consequently, by Lemma 2.10, $\bar{y} = \bar{z}$. But this implies the absurdity $\sigma \in I_{\text{in}}$. `Claim 5` is thus verified.

`Conclusion`: We have shown that (y_n) is Fejér monotone with respect to $C \cap \operatorname{int} \operatorname{dom} f$ (`Claim 2`), and that all its cluster points lie in $C \cap \operatorname{int} \operatorname{dom} f$ (`Claim 5`). Therefore, by Lemma 2.20, the entire sequence (y_n) converges to some point in $C \cap \operatorname{int} \operatorname{dom} f$. ☐

Remark 4.2.

(i) The proof is modeled after the proof of [2, Theorem 8.1], which in turn is standard for random methods; see [2, Remark 8.3]. The new approach is Corollary 3.7, which relies upon the joint convexity of D_f and which resulted in (see `Claim 1` in the proof of Theorem 4.1) the crucial inequality

$$D_f(c, y_n) \geq D_f(c, y_{n+1}) + D_{D_f}\big((c,c),(y_n, y_{n+1})\big).$$

It is instructive to compare this to the key inequality for the sequence $(\tilde{y}_n)$ generated by the method of backward projections (see [2, Proof of Theorem 8.1]):

$$D_f(c, \tilde{y}_n) \geq D_f(c, \tilde{y}_{n+1}) + D_f(\tilde{y}_{n+1}, \tilde{y}_n).$$

See also Remark 3.8.

(ii) For applications, it would be desirable to have a convergence result allowing for $C \cap \operatorname{int}\operatorname{dom} = \emptyset$. In the standard (backward) Bregman projections setting, this case can be accommodated by imposing further properties on f, for instance, that f be *Bregman* [9, 10] or *Bregman/Legendre* [2]. Unfortunately, this technique is not applicable here: the stumbling block in the proof appears to be `Claim 5`, which relies on Lemma 2.10 to obtain uniqueness of the cluster points. However, Lemma 2.10 in turn requires the points to lie in $\operatorname{int}\operatorname{dom} f$. In essence, this is closely related to the previous item (i): for backward projections, the crucial term $D_f(\tilde{y}_{n+1}, \tilde{y}_n)$ is *independent* of c; however, this is not true for the corresponding term $D_{D_f}\big((c,c),(y_n,y_{n+1})\big)$ in the forward projections context.

(iii) Preliminary numerical experiments (using two lines intersecting in the positive orthant, using forward entropy projections; see Section 3) indicate that the qualitative behavior of forward entropic projections is similar to orthogonal (energy) projections. A more detailed numerical study would be an interesting topic for further research.

Acknowledgment

We wish to thank Greg Reid, Asen Dontchev, and Jong-Shi Pang for helpful comments which resulted in Proposition 3.3, Remark 3.4, and Remark 3.9. The initial part of this work was carried out while HHB visited Toulouse in Summer 2000. HHB wishes to thank the Department of Mathematics at University Paul Sabatier for their hospitality and support.

References

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

[2] H.H. Bauschke and J.M. Borwein. Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis* 4(1):27–67, 1997.

[3] H.H. Bauschke and J.M. Borwein. Joint and separate convexity of the Bregman distance. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently parallel algorithms in feasibility and optimization and their applications*, volume 8 of *Studies in computational mathematics*, pages 23–36. Elsevier, 2001.

[4] H.H. Bauschke and D. Noll. On the asymptotic behaviour of alternating backward/forward projections. In preparation.

[5] H.H. Bauschke, D. Noll, A. Celler, and J.M. Borwein. An EM-algorithm for dynamic SPECT tomography. *IEEE Transactions on Medical Imaging*, 18(3):252–261, 1999.

[6] J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization.* Springer-Verlag, 2000.

[7] L.M. Bregman. The method of successive projection for finding a common point of convex sets. *Soviet Mathematics Doklady* 6:688–692, 1965.

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

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

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

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

[12] P.L. Combettes. Hilbertian convex feasibility problem: convergence of projection methods. *Applied Mathematics and Optimization* 35(3):311–330, 1997.

[13] P.L. Combettes. Fejér-monotonicity in convex optimization. In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization.* Kluwer, 2000.

[14] P.L. Combettes. Quasi-Fejérian analysis of some optimization algorithms. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently parallel algorithms in feasibility and optimization and their applications*, volume 8 of *Studies in computational mathematics*, pages 115–152. Elsevier, 2001.

[15] I. Csiszár and G. Tusnády. Information geometry and alternating minimization procedures. *Statistics and Decisions* (Supplement 1), 205–237, 1984.

[16] A.L. Dontchev and T. Zolezzi. *Well-Posed Optimization Problems.* Lecture Notes in Mathematics 1543. Springer-Verlag, 1993.

[17] P.T. Harker and J.-S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming* 48:161–220, 1990.

[18] A.N. Iusem. Convergence analysis for a multiplicatively relaxed EM algorithm. *Mathematical Methods in the Applied Sciences* 14(8):573–593, 1991.

[19] A.N. Iusem. A short convergence proof of the EM algorithm for a specific Poisson model. *Revista Brasileira de Probabilidade e Estatistica* 6:57–67, 1992.

[20] K.C. Kiwiel and B. Łopuch. Surrogate projection methods for finding fixed points of firmly nonexpansive mappings. SIAM Journal on Optimization 7(4):1084–1102, 1997.

[21] G. Reid. Personal communication, 2000.

[22] A.W. Roberts and D.E. Varberg. *Convex Functions.* Academic Press, 1973.

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