Steered Sequential Projections for the Inconsistent Convex Feasibility Problem

Yair Censor[1], Alvaro R. De Pierro[2]
and Maroun Zaknoon[1*]

[1]Department of Mathematics,
University of Haifa, Mt. Carmel, Haifa 31905,
Israel ({zaknoon,yair}@math.haifa.ac.il).

[2]Department of Applied Mathematics,
State University of Campinas, CP6065, CEP 13081,
Campinas, SP, Brazil (alvaro@ime.unicamp.br).

November 19, 2001, Revised: May 2, 2004, Final Revision:
July 18, 2004

Abstract

We study a steered sequential gradient algorithm which minimizes the sum of convex functions by proceeding cyclically in the directions of the negative gradients of the functions and using steered step-sizes. This algorithm is applied to the convex feasibility problem by minimizing a proximity function which measures the sum of the Bregman distances to the members of the family of convex sets. The resulting algorithm is a new steered sequential Bregman projection method

*Current address: Department of Computer Science, Mar Elias Campus, Branch of the University of Indianapolis, P.O. Box 102, Ibillin 30012, Israel.

which generates sequences that converge if they are bounded, regardless of whether the convex feasibility problem is or is not consistent. For orthogonal projections and affine sets the boundedness condition is always fulfilled.

1 Introduction

Many problems in mathematics, in physical sciences and in real-world applications of various technological innovations can be modeled as a *convex feasibility problem.* This is the problem of finding a point $x^* \in Q := \cap_{i=0}^{m-1} Q_i$ in the intersection of finitely many closed convex sets $Q_i \subseteq R^n$ in the Euclidean space. A central role in the area of constructive solution of such problems is played by *projection algorithms.* These are iterative algorithms which use projections onto the individual sets Q_i, employed in a way dictated by the algorithmic recipe, to generate a sequence of iterates $\{x^k\}_{k\geq 0}$ which converges to an $x^* \in Q$. See, e.g., Bauschke and Borwein [6], Combettes [17] and Censor and Zenios [14, Chapter 5] for recent work in this field. If $Q \neq \emptyset$ the convex feasibility problem is called *consistent*, otherwise it is *inconsistent.* Fully simultaneous (parallel) algorithmic schemes are of the form

$$x^{k+1} = x^k + \lambda_k \left(\sum_{i=0}^{m-1} w_i^k P_{Q_i}(x^k) - x^k \right), \tag{1}$$

where $P_\Omega(x)$ stands for the orthogonal projection of a point x onto the closed convex set Ω, the parameters $\{w_i^k\}_{i=0}^{m-1}$ are, for every $k \geq 0$, a system of weights such that $w_i^k > 0$ for all $0 \leq i \leq m-1$, and $\sum_{i=0}^{m-1} w_i^k = 1$, and the parameters $\{\lambda_k\}_{k\geq 0}$, called *relaxation parameters,* are user-chosen and, in most convergence analyses, must remain in the interval $[\epsilon, 2-\epsilon]$, for an arbitrarily small $\epsilon > 0$, in order to guarantee convergence, see, e.g., Aharoni and Censor [1], where the behavior of such algorithms for the consistent case is studied. Such fully simultaneous algorithms generate iteration sequences $\{x^k\}_{k\geq 0}$ which always converge, even if the underlying convex feasibility problem is inconsistent, see, e.g., Combettes [16], Byrne and Censor [12]. At the other end (from the structural point of view) of the “spectrum” of projection algorithms one finds the well-known sequential method of successive orthogonal projections of Bregman [11], further studied by Gubin, Polyak and Raik [22] (see also [14, Algorithm 5.2.1] and Bauschke and Borwein [6]) – also known as the POCS (Projections Onto Convex Sets) method, see, e.g., Stark and Yang [32]. This is given by the algorithmic iterative step

$$x^{k+1} = x^k + \lambda_k \left(P_{Q_{i(k)}}(x^k) - x^k \right). \tag{2}$$

The progress of the algorithm is governed by a *cyclic control sequence* over the index set $\{0, 1, \cdots, m-1\}$, i.e.,

$$i(k) = k \bmod m, \text{ for } k \geq 0. \tag{3}$$

This sequential POCS method converges in the consistent case, when the intersection $\cap_{i=0}^{m-1} Q_i$ is nonempty, to a point in the intersection, see [22, Theorem 1]. However, in the inconsistent case it does not converge but rather demonstrates what is called *cyclic convergence*, i.e., convergence of the cyclic subsequences, see [22, Theorem 2].

The question of how projection algorithms behave in the inconsistent case when $\cap_{i=0}^{m-1} Q_i = \emptyset$ is significant in practical applications when it is not known a priori whether or not the problem is consistent. So, while simultaneous methods exhibit convergence to a minimum of a ceratin *proximity function*, there are to date no similar results available about convergence in the *inconsistent case* of *sequential* projection methods. The questions that arise are numerous: Are there sequential projection algorithms for the convex feasibility problem which converge in the inconsistent case? If so, under what conditions can their convergence in the inconsistent case be guaranteed? Can such results be formulated also for non-orthogonal projections, such as, for example, for the class of Bregman projections? Under what conditions?

These questions present a great challenge and we are able to offer here partial answers that hold only under some restrictive conditions. Whether or not our current restrictions can be relaxed or removed still remains to be seen. Our starting point is the minimization problem of a function $g(x) := \sum_{i=0}^{m-1} g_i(x)$ where $\{g_i\}_{i=0}^{m-1}$ is a family of convex functions on the Euclidean space R^n which have continuous derivatives everywhere. The minimization is achieved by a gradient method which uses the gradients of the functions g_i in a cyclic manner. This type of methods is known as *incremental gradient algorithm* and has already an extensive literature, in particular from the area of neural networks, where it is known as "back propagation", see, e.g., Bertsekas [9], Bertsekas and Tsitsiklis [10], Nedić and Bertsekas [27] and references therein or consult Bertsekas' book [8, Section 1.5.2]. Initializing the iterations at an arbitrary $x^0 \in R^n$, the following sequential cyclic gradient algorithm is proposed

$$x^{k+1} = x^k - \sigma_k \nabla g_{i(k)}(x^k),\ \ k \geq 0. \tag{4}$$

The sequence $\{\sigma_k\}_{k\geq 0}$ is a sequence of real positive numbers which satisfy the following conditions: (i) $\lim_{k\to\infty} \sigma_k = 0$, (ii) $\lim_{k\to\infty} \sigma_{km+j}/\sigma_{km} = 1$, for all $0 \leq j \leq m-1$, and (iii) $\sum_{k=0}^{\infty} \sigma_k = +\infty$. Because of these conditions, particularly condition (i) which "pushes" the sequence to zero, we call this parameters here *m-steering parameters*. This steering feature of the parameters has a profound effect on the behavior of the iterates $\{x^k\}_{k\geq 0}$. We prove that if the iterates $\{x^k\}_{k\geq 0}$, generated by (4), are *bounded* then they converge to a minimizer of the function $g(x)$. A similar assumption of boundedness appears in Gubin, Polyak and Raik's [22, Theorem 2] where cyclic convergence of the sequential POCS method, mentioned above, is proven for the inconsistent case and one of the, finitely many, convex sets has to be *bounded.* In spite of the large literature on incremental gradient (and subgradient) methods it seems that our theorem about the convergence of (4) is new with respect to the choice of the sequence $\{\sigma_k\}_{k\geq 0}$ of m-steering parameters. However, this is just a slight modification over previous results with steering parameters and is not our main result but rather a tool. The main thrust of our work is to derive results about convergence of sequential projection methods for the convex feasibility problem that will converge even in the inconsistent case, thus, to complement the only known result for this situation which is the theorem of Gubin, Polyak and Raik [22, Theorem 2].

To do so we apply the algorithm of (4) to the minimization of a *proximity function* which measures the sum of the Bregman distances of its argument from a family of convex sets. In this way we arrive at convergence results for steered sequential projections for the inconsistent convex feasibility problem. More accurately, we minimize the function $d_{\mathcal{Q}}^f(x) := \sum_{i=0}^{m-1} d_{Q_i}^f(x)$, where $\mathcal{Q} := \{Q_i\}_{i=0}^{m-1}$ is a family of convex sets in R^n, f is a twice continuously differentiable Bregman function and $d_{Q_i}^f(x)$ is the *directed Bregman distance* from the point x to the set Q_i. By substituting $g_i(x) := d_{Q_i}^f(x)$ in the steered cyclic gradient method (4) we obtain an iterative step formula of the form

$$x^{k+1} = x^k + \sigma_k \nabla^2 f(x^k) \left(P_{Q_{i(k)}}^f(x^k) - x^k \right), \tag{5}$$

where ∇^2 is the Hessian matrix and $P_\Omega^f(z)$ stands for the Bregman projection, with respect to the Bregman function f, of a point z onto the closed convex set Ω. We show that bounded sequences of iterates generated by such sequential projection methods converge even in the inconsistent case of the convex feasibility problem, provided that the parameters σ_k are not just relaxation parameters but are rather m-steering parameters. Under such circumstances the iterates so generated converge to a minimizer of the proximity function $d^f_{\mathcal{Q}}(x)$ of the family $\mathcal{Q}$.

The paper is laid out as follows. In Section 2 we study the cyclic gradient method with steering parameters. Section 3 shows how to apply this to the proximity function minimization problem where proximity is measured by Bregman point-to-set distances. Special attention is given to the orthogonal case in Section 4 where the boundedness restriction can be removed for the special case when the sets Q_i are hyperplanes. Our results provide a simpler proof of the result of Censor, Eggermont and Gordon [13] by replacing the double-limit procedure employed there by a single limiting process. More insights, results and conjectures for and about the orthogonal case can be found in De Pierro [18]. In Section 5 we analyze the entropy case which is a different realization of Bregman functions and distances and show how the general theory covers this case and what the limitations to full coverage are.

2 The Steered Cyclic Gradient Method

In this section we present the prototypical cyclic gradient method and study its convergence. The steering parameters σ_k which appear in the algorithm must have some desirable properties which we list in the following definitions.

Definition 1 *A sequence $\{\sigma_k\}_{k\geq 0}$ of real positive numbers will be called a* `steering sequence` *if it satisfies the following conditions:*

$$\lim_{k\to\infty} \sigma_k = 0, \tag{6}$$

$$\lim_{k\to\infty} (\sigma_{k+1}/\sigma_k) = 1, \tag{7}$$

and

$$\sum_{k=0}^{\infty} \sigma_k = +\infty. \tag{8}$$

Different conditions of a similar nature have been used in the past, in conjunction with other algorithms, by several authors, see, e.g., Lions [25], Wittmann [33], Reich [29] and [30], Combettes [16], Bauschke [4], Deutsch and Yamada [19], and references therein, to name but a few.

Definition 2 *Let m be an integer. If, in Definition 1, the condition (7) is replaced by the following condition:*

$$\lim_{k\to\infty} \sigma_{km+j}/\sigma_{km} = 1, \text{ for all } 1 \le j \le m-1, \tag{9}$$

then the sequence is called an `m-steering sequence`.

Remark 3 *It can be shown without difficulty that condition (9) is weaker than (7).*

The next proposition gives a useful result about m-steering sequences.

Proposition 4 *If $\{\sigma_k\}_{k\ge0}$ is an $m-$steering sequence then $\sum_{k=0}^{\infty}\sigma_{mk} = +\infty$.*

Proof. From (9) we know that there exists an integer K such that for all $k > mK$ and all $1 \le j \le m-1$, it is true that $(\sigma_{mk+j}/\sigma_{mk}) < 2$. Let us assume by negation that $\sum_{k=0}^{\infty}\sigma_{mk} < \infty$, then we have

$$\begin{aligned}\sum_{k=mK}^{\infty}\sigma_k &= \sum_{k=K}^{\infty}\sigma_{mk} + \sum_{k=K}^{\infty}\sigma_{mk}\sum_{j=1}^{m-1}(\sigma_{mk+j}/\sigma_{mk}) \\ &< \sum_{k=K}^{\infty}\sigma_{mk} + \sum_{k=K}^{\infty}\sigma_{mk}(m-1)2 < \infty,\end{aligned} \tag{10}$$

which contradicts (8). ■

Let $\{g_i\}_{i=0}^{m-1}$ be a family of convex functions $g_i : R^n \to R$ which are continuously differentiable everywhere. The cyclic gradient method is as follows.

Algorithm 5 ***(The steered cyclic gradient method).***

Initialization: *$x^0 \in R^n$ is arbitrary.*

Iterative Step: *Given x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = x^k - \sigma_k \nabla g_{i(k)}(x^k). \tag{11}$$

Control Sequence: *$\{i(k)\}_{k\geq 0}$ is a cyclic control sequence, i.e., $i(k) = k \bmod m$.*

Steering Parameters: *The sequence $\{\sigma_k\}_{k\geq 0}$ is m-steering.*

Our convergence result for this algorithm is the following.

Theorem 6 *Let $\{g_i\}_{i=0}^{m-1}$ be a family of functions $g_i : R^n \to R$ which are convex and continuously differentiable everywhere, let $g(x) := \sum_{i=0}^{m-1} g_i(x)$ and assume that g has an unconstrained minimum. If $\left\{x^k\right\}_{k\geq 0}$ is a* **bounded** *sequence, generated by Algorithm 5, then the sequence $\left\{g\left(x^k\right)\right\}_{k\geq 0}$ converges to the minimum of g. If, in addition, we assume that g has a unique minimizer then the sequence $\left\{x^k\right\}_{k\geq 0}$ converges to this minimizer.*

This theorem's scope is obviously restricted because it proves convergence only if boundedness of the iterative sequence already exists. In order to prove this theorem we need the next four propositions. We use the following notations: $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ are the Euclidean norm and the standard inner product, respectively, $y^k := x^{mk}$, for all $k \geq 0$, the closure of the convex hull of the sequence $\left\{x^k\right\}_{k\geq 0}$ is denoted by C, and $\mu := \inf\{g(x) \mid x \in R^n\}$.

Proposition 7 *Under the assumptions of Theorem 6, for every $\nu > \mu$ there exists a $\delta > 0$ such that if $x \in C$ and $g(x) > \nu$ then $\|\nabla g(x)\| > \delta$.*

Proof. The set $\Psi := \{x \mid g(x) \geq \nu,\ x \in C\}$ is closed due to the continuity of $g(x)$. Since $\{g_i\}_{i=0}^{m-1}$ are convex functions, so is the function g. This implies that $\nabla g(x) \neq 0$ whenever x is not a minimizer of g. Thus, we obtain $\|\nabla g(x)\| > 0$ for all $x \in \Psi$. From this, from the closedness of Ψ and from the continuity of $\nabla g(x)$ we deduce that there is a number $\delta > 0$ such that $\|\nabla g(x)\| > \delta$ for all $x \in \Psi$. Hence the proposition follows. ■

Proposition 8 *If the assumptions of Theorem 6 hold then there exists a sequence $\{r^k\}_{k\geq 0}$ with $\lim_{k\to\infty} r^k = 0$, for which*

$$y^k - y^{k+1} = \sigma_{mk}(\nabla g(y^k) + r^k). \tag{12}$$

Proof. Since $\left\{x^k\right\}_{k\geq 0}$ is bounded and ∇g_i is continuous for every $0 \leq i \leq m-1$, the sequence $\left\{\nabla g_{i(k)}\left(x^k\right)\right\}_{k\geq 0}$ is bounded. The iterative step (11) implies that, for all $k \geq 0$,

$$\left\|x^{k+1} - x^k\right\| = |\sigma_k| \left\|\nabla g_{i(k)}\left(x^k\right)\right\|. \tag{13}$$

Therefore, for all $k \geq 0$ and for all $0 \leq j \leq m-1$,

$$\left\|x^{mk+j} - y^k\right\| = \left\|x^{mk+j} - x^{mk}\right\| \leq \sum_{t=mk}^{mk+j-1} \left\|x^{t+1} - x^t\right\|$$
$$= \sum_{t=mk}^{mk+j-1} |\sigma_t| \left\|\nabla g_{i(t)}\left(x^t\right)\right\|. \tag{14}$$

The last inequality, the boundedness of $\left\{\nabla g_{i(k)}\left(x^k\right)\right\}_{k\geq 0}$ and (6) guarantee that, for all $0 \leq j \leq m-1$,

$$\lim_{k\to\infty} \left\|x^{mk+j} - y^k\right\| = 0. \tag{15}$$

Defining

$$\varphi_j^k := \nabla g_j\left(x^{mk+j}\right) - \nabla g_j(y^k), \tag{16}$$

(15) and the continuity of the gradients ∇g_i, $0 \leq i \leq m-1$, yield, for all $0 \leq j \leq m-1$,

$$\lim_{k\to\infty} \varphi_j^k = 0. \tag{17}$$

Denoting, for all $k \geq 0$ and for all $0 \leq j \leq m-1$,

$$\xi_j^k := (\sigma_{mk+j}/\sigma_{mk}) - 1, \tag{18}$$

(7) yields

$$\lim_{k\to\infty} \xi_j^k = 0. \tag{19}$$

By (16) and the iterative step (11), it is clear that

$$y^{k+1} - y^k = x^{m(k+1)} - x^{mk} = \sum_{j=0}^{m-1}(x^{mk+j+1} - x^{mk+j})$$
$$= \sum_{j=0}^{m-1}\left(-\sigma_{mk+j}\nabla g_{i(mk+j)}\left(x^{mk+j}\right)\right)$$
$$= -\sigma_{mk}\sum_{j=0}^{m-1}(1+\xi_j^k)\nabla g_j\left(x^{mk+j}\right)$$
$$= -\sigma_{mk}\sum_{j=0}^{m-1}(1+\xi_j^k)\left(\nabla g_j(y^k) + \varphi_j^k\right). \tag{20}$$

This result, the limits (17) and (19), the boundedness of $\nabla g_i(y^k)$, for all $0 \le i \le m-1$, and the definition of the function $g(x)$ guarantee the existence of the desired sequence

$$r^k = \sum_{j=0}^{m-1} \left(\varphi_j^k + \xi_j^k \left(\nabla g_j(y^k) + \varphi_j^k\right)\right), \tag{21}$$

thus, completing the proof. ■

Proposition 9 *Under the assumptions of Theorem 6, for every $\eta > 0$ and $\nu > \mu$ there exists an integer K such that for all $k > K$, if $g(y^k) > \nu$ then*

$$\left\|(y^k - y^{k+1}) - \sigma_{mk}\nabla g(y^k)\right\| < \eta\sigma_{mk}\left\|\nabla g(y^k)\right\|. \tag{22}$$

Proof. By Proposition 7 and by the assumption $g(y^k) > \nu$ we have that $\left\{\left\|\nabla g(y^k)\right\|\right\}_{k\ge 0}$ is bounded from below by $\delta > 0$. This, together with Proposition 8 and (6), completes the proof. ■

Proposition 10 *Under the assumptions of Theorem 6, for every $\nu > \mu$ there exist a real $\theta > 0$ and an integer K such that if $k > K$ and $g(y^k) > \nu$, then $g(y^k) - g(y^{k+1}) > \theta\sigma_{mk}$.*

Proof. Throughout this proof we consider only points y^k which fulfil $g(y^k) > \nu$. By Proposition 9, for $\eta = 1/2$ there exists an integer K_1 such that if $k > K_1$ and $g(y^k) > \nu$ then

$$\left\|(y^k - y^{k+1})/\sigma_{mk} - \nabla g(y^k)\right\| < (1/2)\left\|\nabla g(y^k)\right\|. \tag{23}$$

But this last inequality is equivalent to

$$\begin{aligned} &\left\|(y^k - y^{k+1})/\sigma_{mk}\right\|^2 - 2\left\langle (y^k - y^{k+1})/\sigma_{mk}, \nabla g(y^k)\right\rangle + \left\|\nabla g(y^k)\right\|^2 \\ &< \frac{1}{4}\left\|\nabla g(y^k)\right\|^2, \end{aligned} \tag{24}$$

which implies that

$$(3/8)\left\|\nabla g(y^k)\right\|^2 < \left\langle (y^k - y^{k+1})/\sigma_{mk}, \nabla g(y^k)\right\rangle. \tag{25}$$

By (23) it is easy to see that

$$\left\|(y^k - y^{k+1})/\sigma_{mk}\right\| < (3/2)\left\|\nabla g(y^k)\right\|. \tag{26}$$

From Proposition 7 there is a $\delta > 0$ such that, for sufficiently large k,

$$0 < (3/8)\,\delta^2 < (3/8)\left\|\nabla g(y^k)\right\|^2. \tag{27}$$

This, along with (25), ensures that $(y^k - y^{k+1})/\sigma_{mk} \neq 0$. Therefore, using (25) and (26), we obtain

$$(1/4)\left\|\nabla g(y^k)\right\| = \frac{(3/8)\left\|\nabla g(y^k)\right\|^2}{(3/2)\left\|\nabla g(y^k)\right\|} < \left\langle \frac{y^k - y^{k+1}}{\|y^k - y^{k+1}\|}, \nabla g(y^k) \right\rangle, \tag{28}$$

which means that the value of the directional derivative of g at the point y^k in the direction of the vector $y^k - y^{k+1}$ is at least $(1/4)\left\|\nabla g(y^k)\right\|$.

Denote the directional derivative of g at a point x in the direction d by $g'(x; d)$. From Proposition 7 and the inequality in (28), it follows that the directional derivative $g'(y^k; y^k - y^{k+1})$ is larger than the constant $\alpha = \delta/4 > 0$. By Proposition 8, the boundedness of $\nabla g(y^k)$ and the fact that $\lim_{k\to\infty} \sigma_{mk} = 0$, the difference between consecutive iterates $\left\|y^k - y^{k+1}\right\|$ tends to zero. So, using the continuity of the directional derivative $g'(x; d)$ in both x and d we can choose an integer $K > K_1$ such that for all $k > K$,

$$g'\left(y^k + \gamma(y^k - y^{k+1}); (y^k - y^{k+1})\right) > \alpha/2, \tag{29}$$

for any $\gamma \in [-1, 0]$, that is, the directional derivatives of g at any point in the segment connecting y^{k+1} with y^k, in the direction of that segment, is larger than $\alpha/2$. Lagrange's intermediate value theorem implies that there exists a value $\bar{\gamma} \in [-1, 0]$, such that,

$$g(y^k) - g(y^{k+1}) > g'\left(y^k + \bar{\gamma}(y^k - y^{k+1}); (y^k - y^{k+1})\right) \cdot \left\|y^k - y^{k+1}\right\| \tag{30}$$

$$> (\alpha/2)\left\|y^k - y^{k+1}\right\| > \theta\sigma_{mk}, \tag{31}$$

where the existence of $\theta > 0$ is guaranteed by Proposition 8 and the boundedness from below of the gradient's norm in Proposition 7. ■

Now we use Propositions 7–10 to deduce the proof of Theorem 6.

Proof. of Theorem 6. Assume first, that $\nu := \inf\{g(x) \mid x \in C\} > \mu$. Then, by Proposition 10, there exists a constant $\theta > 0$ such that, for some large enough integer K, for all $k > K$,

$$g(y^k) - g(y^{k+1}) > \theta\sigma_{mk}. \tag{32}$$

This, along with Proposition 4, implies that $\lim_{k\to\infty} g(y^k) = -\infty$, which is a contradiction to the boundedness of the sequence $\{y^k\}_{k\geq 0}$. Therefore, there is a subsequence of $\{g(y^k)\}_{k\geq 0}$ which converges to μ. Hence we may assume that g attains its minimum μ on C. Now we prove that $\lim_{k\to\infty} g(y^k) = \mu$. Assume, by negation, that an infinite number of elements of the sequence $\{g(y^k)\}_{k\geq 0}$ are larger than a constant number $\tilde{\nu} > \mu$, i.e., that there exists a subsequence that tends to $\tilde{\nu}$. By Proposition 10, there exist a constant $\tilde{\theta} > 0$ and an integer $\tilde{K}$ such that, for all $k > \tilde{K}$,

$$\text{if } g(y^k) > (\mu + \tilde{\nu})/2 \quad \text{then} \quad g(y^k) - g(y^{k+1}) > \tilde{\theta}\sigma_{mk}, \tag{33}$$

which yields that, for all $k > \tilde{K}$,

$$\text{if } g(y^k) > (\mu + \tilde{\nu})/2 \;\; \text{then } g(y^{k+1}) < g(y^k). \tag{34}$$

Since there is a subsequence of $\{g(y^k)\}_{k\geq 0}$ which tends to μ, and another one which tends to $\tilde{\nu}$, we can build two subsequences $\{y^{k_l}\}_{l\geq 0}$ and $\{y^{k_l+1}\}_{l\geq 0}$, in the following way. Take $g(y^{k_l})$ such that, for all $k_l > \tilde{K}$,

$$g(y^{k_l}) \leq (\mu + \tilde{\nu})/2. \tag{35}$$

and

$$g(y^{k_l+1}) > \tilde{\nu} \geq g(y^{k_l}). \tag{36}$$

The latter subsequence exists because both μ and $\tilde{\nu}$ are limit points of $\{g(y^k)\}_{k\geq 0}$. If such a sequence would not exist then the limit $\tilde{\nu}$ could never be attained, since in between $\tilde{\nu}$ and $(\mu + \tilde{\nu})/2$ the function g is decreasing. Subtracting (35) from the left-hand inequality of (36) yields, for every $k_l > \tilde{K}$,

$$g(y^{k_l+1}) - g(y^{k_l}) > (\tilde{\nu} - \mu)/2 > 0. \tag{37}$$

On the other hand, Proposition 8 and the facts that g is continuous and that $\{\nabla g(y^k)\}_{k\geq 0}$ is bounded yield that

$$\lim_{k\to\infty} \left(g(y^{k_l+1}) - g(y^{k_l})\right) = 0, \tag{38}$$

which is a contradiction. Therefore, $\lim_{k\to\infty} g(y^k) = \mu$. This along with the limit in (15) and the continuity of the function g give

$$\lim_{k\to\infty} g(x^k) = \mu, \tag{39}$$

which proves convergence of the function values to the minimum. In case g has a unique minimizer, say x^*, the limit

$$\lim_{k\to\infty} x^k = x^*, \tag{40}$$

must be satisfied, else there are an infinite subsequence $\left\{x^{k_l}\right\}_{l\geq 0}$ and a real number $\rho > 0$ such that $\left\|x^{k_l} - x^*\right\| > \rho$, for all $l = 0, 1, 2, \cdots$. This inequality, the uniqueness of the minimizer and the convexity and continuity of the function g yield that there is a real number $\mu_0 > \mu$ for which $g(x^{k_l}) \geq \mu_0$, for all $l = 0, 1, 2, \cdots$, which contradicts (39). ■

3 Cyclic Sequences of Steered Bregman Projections onto Convex Sets

In this section we apply the steered cyclic gradient method (Algorithm 5) with Bregman distance functions and derive from Theorem 6 a new convergence result for sequential Bregman projections onto convex sets in the inconsistent case. For background material on Bregman functions and Bregman distances and projections we refer the reader to the book by Censor and Zenios [14, Chapter 2], to Bauschke and Borwein [7], or to recent papers of Solodov and Svaiter [31], Eckstein [20, 21] and references therein. Let f be a Bregman function with *zone* S and denote by $D_f(x,y)$ the Bregman distance, between the points x and y, defined on $\mathrm{cl}\, S \times S$, where $\mathrm{cl}\, S$ denotes the closure of the set S, by

$$D_f(x,y) := f(x) - f(y) - \langle \nabla f(y), x - y \rangle. \tag{41}$$

If Ω is a closed convex subset of R^n such that $\Omega \cap \mathrm{cl}\, S \neq \emptyset$, we denote by $P_\Omega^f(x)$ the Bregman projection of x onto Ω with respect to f. This is defined by

$$P_\Omega^f(x) = \mathrm{argmin}\{D_f(z,x) \mid z \in \Omega \cap \mathrm{cl}\, S\}. \tag{42}$$

We define the *Bregman directed distance* from a point $x \in S$ to the set Ω as the value of the Bregman distance between the point and its Bregman projection onto the set

$$d_\Omega^f(x) := \min\{D_f(z,x) \mid z \in \Omega \cap \operatorname{cl} S\}. \tag{43}$$

From now on we always make the assumption that the Bregman functions f and the sets Ω in use are such that f is always *zone consistent with respect to* Ω. This means that $x \in S$ implies that $P_\Omega^f(x) \in S$. Bauschke and Borwein [7] identified the class of useful *Bregman/Legendre functions* for which this assumption always holds.

Obviously, by the definition of the Bregman projection of x onto Ω, see, e.g., [14, Definition 2.1.2], and the definition of the distance D_f,

$$d_\Omega^f(x) = D_f\left(P_\Omega^f(x), x\right) = f\left(P_\Omega^f(x)\right) - f(x) - \langle \nabla f(x), P_\Omega^f(x) - x\rangle. \tag{44}$$

The next proposition gives a geometric interpretation of the Bregman projection onto a set.

Proposition 11 *Let $\Omega \subseteq R^n$ be a closed convex set and let f be a Bregman function with zone S, such that $\Omega \cap \operatorname{cl} S \neq \emptyset$ and assume that f is zone consistent with respect to Ω. If $x \in S$, but $x \notin \Omega$, then the hyperplane*

$$H := \{u \in R^n \mid \left\langle u - P_\Omega^f(x), \nabla f(x) - \nabla f\left(P_\Omega^f(x)\right)\right\rangle = 0\} \tag{45}$$

separates x from $\Omega \cap \operatorname{cl} S$ and supports the set Ω at the point $P_\Omega^f(x) \in \Omega$.

Proof. From the characterization of a Bregman projection, see, e.g., [14, Theorem 2.4.2], it follows that

$$\Omega \cap \operatorname{cl} S \subseteq \left\{u \in R^n \mid \left\langle u - P_\Omega^f(x), \nabla f(x) - \nabla f\left(P_\Omega^f(x)\right)\right\rangle \leq 0\right\}. \tag{46}$$

On the other hand, using the nonnegativity of Bregman distances, see, e.g., [14, Lemma 2.1.1],

$$\begin{aligned}
&\left\langle x - P_\Omega^f(x), \nabla f(x) - \nabla f\left(P_\Omega^f(x)\right)\right\rangle \\
&= f(x) - f\left(P_\Omega^f(x)\right) - \left\langle x - P_\Omega^f(x), \nabla f\left(P_\Omega^f(x)\right)\right\rangle \\
&\quad + f\left(P_\Omega^f(x)\right) - f(x) - \left\langle P_\Omega^f(x) - x, \nabla f(x)\right\rangle \\
&= D_f\left(x, P_\Omega^f(x)\right) + D_f\left(P_\Omega^f(x), x\right) \geq 0,
\end{aligned} \tag{47}$$

which implies that x lies on the opposite side of H then Ω. ∎

Denoting the Hessian matrix by ∇^2, the following proposition gives a useful formula for the gradient of $d_\Omega^f(x)$.

Proposition 12 *Let $\Omega \subseteq R^n$ be a closed convex set and let f be a Bregman function with zone S, such that $\Omega \cap \operatorname{cl} S \neq \varnothing$ and assume that f is zone consistent with respect to Ω. If f is twice continuously differentiable, if $d_\Omega^f(x)$ is a convex function and if for every hyperplane H the derivatives*

$$\frac{\partial}{\partial x_j}\left(P_H^f(x)\right)_i, \quad 1 \leq i,\ j \leq n, \tag{48}$$

exist and are continuous for all $x \in S$, then $d_\Omega^f(x)$ is a continuously differentiable function, and

$$\nabla\left(d_\Omega^f(x)\right) = \nabla^2 f(x)\left(x - P_\Omega^f(x)\right). \tag{49}$$

Proof. For any given point $x \in S$, let E be the singleton $E := \left\{P_\Omega^f(x)\right\}$. Writing out $d_E^f(y)$ for any $y \in S$, according to (44), and recalling that

$$P_E^f(y) = P_\Omega^f(x) \tag{50}$$

because E is a singleton, leads to

$$d_E^f(y) = f\left(P_\Omega^f(x)\right) - f(y) - \langle \nabla f\,(y)\,, P_\Omega^f(x) - y\rangle. \tag{51}$$

Direct differentiation then shows that, for any $y \in S$,

$$\nabla\left(d_E^f(y)\right) = \nabla^2 f(y)\left(y - P_\Omega^f(x)\right). \tag{52}$$

By Proposition 11, it is easy to see that

$$\Omega \cap \operatorname{cl} S \subseteq H^f := \{u \in R^n \mid \langle u, a\rangle \leq b\}, \tag{53}$$

where

$$a := \nabla f\,(x) - \nabla f\left(P_\Omega^f(x)\right) \tag{54}$$

and

$$b := \left\langle P_\Omega^f(x), \nabla f(x) - \nabla f\left(P_\Omega^f(x)\right)\right\rangle. \tag{55}$$

Let us calculate $\nabla(d_{H^f}^f(y))$, for any $y \in S$. It is clear that

$$\nabla\left(d_{H^f}^f(y)\right) = 0 = \nabla^2 f(y)\left(y - P_{H^f}^f(y)\right), \text{ for all } y \in S \cap \operatorname{int} H^f. \tag{56}$$

It remains to calculate $\nabla(d_{H^f}^f(y))$ for $y \in S \diagdown \operatorname{int} H^f$. Therefore, from here till a new statement we assume that $y \in S \diagdown \operatorname{int} H^f$. We want to call the reader's attention that if $y \in S \diagdown \operatorname{int} H^f$ then $P_{H^f}^f(y) = P_{\operatorname{bd} H^f}^f(y)$. This equation will be used several times in this proof. By using this equation with (44) we write

$$d_{H^f}^f(y) = f\left(P_{\operatorname{bd} H^f}^f(y)\right) - f(y) - \sum_{t=1}^n \frac{\partial f}{\partial y_t}(y)\left(\left(P_{\operatorname{bd} H^f}^f(y)\right)_t - y_t\right), \tag{57}$$

careful differentiation of which gives, for $1 \le j \le n$,

$$\begin{aligned}\left(\nabla\left(d_{H^f}^f(y)\right)\right)_j = {} & \left(\nabla^2 f(y)\left(y - P_{\operatorname{bd} H^f}^f(y)\right)\right)_j \\ & + \left\langle \pi^j(y), \nabla f(P_{\operatorname{bd} H^f}^f(y)) - \nabla f(y)\right\rangle,\end{aligned} \tag{58}$$

where

$$\pi^j(y) := \left(\frac{\partial}{\partial y_j}\left(P_{\operatorname{bd} H^f}^f(y)\right)_1, \frac{\partial}{\partial y_j}\left(P_{\operatorname{bd} H^f}^f(y)\right)_2, \cdots, \frac{\partial}{\partial y_j}\left(P_{\operatorname{bd} H^f}^f(y)\right)_n)\right) \tag{59}$$

For every $y \in S \diagdown \operatorname{int} H^f$. Now we use the characterization of the Bregman projections onto hyperplanes (or half-spaces), see, e.g., [14, Lemma 2.2.1], which yields that the projection of a point $y \in S \diagdown \operatorname{int} H^f$ onto the half-space H^f (i.e., onto the hyperplane $\operatorname{bd} H^f$) is the unique solution of the system

$$\nabla f(P_{\operatorname{bd} H^f}^f(y)) = \nabla f(y) + \alpha(y) a, \tag{60}$$

$$\left\langle P_{\operatorname{bd} H^f}^f(y), a\right\rangle = b, \tag{61}$$

where the real number $\alpha(y)$ is the *Bregman parameter* associated with this Bregman projection. Differentiating both sides of (61) with respect to y_j yields

$$\left\langle \pi^j(y), a\right\rangle = 0, \text{ for every } y \in S \diagdown \operatorname{int} H^f. \tag{62}$$

Combining (62) and (60) shows that the second summand of (58) is always zero. It follows that, $\nabla(d_{H^f}^f(y)) = \nabla^2 f(y)\left(y - P_{H^f}^f(y)\right)$ for all $y \in S \setminus \operatorname{int} H^f$. This result with (56) yield

$$\nabla(d_{H^f}^f(y)) = \nabla^2 f(y)\left(y - P_{H^f}^f(y)\right) \text{ for all } y \in S. \tag{63}$$

Substituting $y = x$ into (52) and into (63) shows that

$$\nabla\left(d_{H^f}^f(x)\right) = \nabla\left(d_E^f(x)\right) = \nabla^2 f(x)\left(x - P_\Omega^f(x)\right). \tag{64}$$

Here we have used the fact that

$$P_{H^f}^f(x) = P_\Omega^f(x) \tag{65}$$

which follows from (45), (54), (55) and the characterization of the Bregman projections onto hyperplanes, see, e.g., [14, Lemma 2.2.1].

It is clear from the definitions and from the relative positions of the sets (consult (45)) that, for every $y \in S$, the Bregman directed distances to the sets H^f, Ω and E are ordered as follows

$$d_{H^f}^f(y) \leq d_\Omega^f(y) \leq d_E^f(y). \tag{66}$$

For $y = x$ there are equalities in (66). Thus, by (64), their gradients at x must all be equal, hence (49) follows. Since the function $P_\Omega^f(x)$ is continuous, it follow that $\nabla\left(d_\Omega^f(x)\right)$ is also continuous. ■

Remark 13 *Proposition 12 generalizes to Bregman projections a well-known result for orthogonal projections, see, e.g., Aubin and Cellina [3, Page 24, Proposition 1]. The classical result is obtained by rewriting (49) for the Bregman function* $f(x) := (1/2)\|x\|^2$ *with* $S = R^n$, $\nabla f(x) = x$ *and* $\nabla^2 f(x) = I$ *the unit matrix.*

Now we are able to formulate the new cyclic Bregman projection method. Let f be a Bregman function with zone S as above, let $\mathcal{Q} = \{Q_i\}_{i=0}^{m-1}$ be a family of closed convex sets such that $Q_i \cap \operatorname{cl} S \neq \emptyset$, for all $0 \leq i \leq m-1$. We denote and define the *proximity function* of the family $\mathcal{Q}$ by

$$d_{\mathcal{Q}}^f(x) := \sum_{i=0}^{m-1} d_{Q_i}^f(x). \tag{67}$$

If the individual Bregman distances to the sets $d^f_{Q_i}(x)$, $0 \le i \le m-1$, are convex functions then we can implement the results of the last section and obtain from Algorithm 5, by employing $g_i(x) = d^f_{Q_i}(x)$, for all $0 \le i \le m-1$, and by using Proposition 12 the following algorithm.

Algorithm 14 ***(The steered cyclic Bregman projections method).***
Initialization: *$x^0 \in S$ is arbitrary.*
Iterative Step: *Given x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = x^k + \sigma_k \nabla^2 f(x^k) \left(P^f_{Q_{i(k)}}(x^k) - x^k \right). \tag{68}$$

Control Sequence: *$\{i(k)\}_{k \ge 0}$ is a cyclic control sequence, i.e., $i(k) = k \bmod m$.*
Steering Parameters: *The sequence $\{\sigma_k\}_{k \ge 0}$ is an m-steering sequence.*

The convergence of this algorithm follows, according to the next theorem, directly from Theorem 6.

Remark 15 *For this algorithm to be well-defined zone consistency alone of f with respect to all sets Q_i is not sufficient because of the form of the iteration formula (68). We must impose on the iterates, generated by Algorithm 14, the condition that $x^k \in S$ for all $k \ge 0$. This holds for the orthogonal case discussed in the next section. In any other specific application of the algorithm, resulting from some choice of the Bregman function f, unless the zone S is the whole space, the condition has to be verified separately.*

Theorem 16 *Let f be a twice continuously differentiable Bregman function with zone S. Let $\mathcal{Q} = \{Q_i\}_{i=0}^{m-1}$ be a family of closed convex sets for which $Q_i \cap \operatorname{cl} S \ne \emptyset$, for all $0 \le i \le m-1$. Assume that $x^k \in S$, for all $k \ge 0$, holds for all iterates of Algorithm 14. Assume further that the distance functions $d^f_{Q_i}(x)$, $0 \le i \le m-1$, are convex, and that for every hyperplane H the derivatives $\frac{\partial}{\partial x_j}\left(P^f_H(x) \right)_i$, $1 \le i, j \le n$, exist and are continuous for all $x \in S$. Then any bounded sequence $\left\{x^k\right\}_{k \ge 0}$, generated by Algorithm 14, converges to a minimizer of $d^f_{\mathcal{Q}}(x)$.*

4 The Orthogonal Case

In this section we deal with the special Bregman function $f(x) := (1/2)\|x\|^2$ with $S = R^n$, in which case Bregman projections become the least-Euclidean distance (orthogonal) projections. Even for this special case, of the theory discussed earlier, our result is new. It is clear that $f(x) = (1/2)\|x\|^2$ satisfies the conditions of Theorem 16. For a family of closed convex sets $\mathcal{Q} = \{Q_i\}_{i=0}^{m-1}$ the distances

$$d_{Q_i}(x) := d_{Q_i}^f(x) = \min\left\{\frac{1}{2}\|x-y\|^2 \mid y \in Q_i\right\}, \tag{69}$$

for all $1 \le i \le m-1$, are the (half-squared) Euclidean distances from the point x to the sets Q_i and it is clear that $d_{Q_i}(x)$, $1 \le i \le m-1$, are convex functions. Furthermore, $\nabla^2 f$ is the identity matrix of order n. Therefore, we obtain as a direct consequence from Algorithm 14 the next algorithm which minimizes the proximity function $d_{\mathcal{Q}}(x) = \sum_{i=0}^{m-1} d_{Q_i}(x)$ by using the orthogonal projections $P_{Q_i} \equiv P_i$, $1 \le i \le m-1$, in a cyclic manner. Its convergence theorem, given below, follows from Theorem 16. Observe the similarity between the POCS method of (2) and (70) below. The only difference is that (70) has steering parameters instead of the relaxation parameters in POCS.

Algorithm 17 ***(The steered cyclic orthogonal projections method).***
Initialization: *$x^0 \in R^n$ is arbitrary.*
Iterative Step: *Given x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = x^k + \sigma_k\left(P_{i(k)}(x^k) - x^k\right). \tag{70}$$

Control Sequence: *$\{i(k)\}_{k\ge0}$ is a cyclic control sequence, i.e., $i(k) = k \bmod m$.*

Steering Parameters: *The sequence $\{\sigma_k\}_{k\ge0}$ is an m-steering sequence.*

Theorem 18 *Let $\mathcal{Q} = \{Q_i\}_{i=0}^{m-1}$ be a family of closed convex sets in R^n and let $d_{\mathcal{Q}}(x) := \frac{1}{2}\sum_{i=0}^{m-1}\|x - P_i(x)\|^2$, the proximity function attached to $\mathcal{Q}$. If $\left\{x^k\right\}_{k\ge0}$ is a bounded sequence, generated by Algorithm 17, then the sequence $\left\{d_{\mathcal{Q}}(x^k)\right\}_{k\ge0}$ converges to the minimum of $d_{\mathcal{Q}}(x)$. If in addition we assume that $d_{\mathcal{Q}}(x)$ has a unique minimizer then the sequence $\left\{x^k\right\}_{k\ge0}$ converges to this minimizer.*

The requirement that $\{x^k\}_{k\geq 0}$ is bounded in Theorem 18 is obviously a strong assumption and we do not know if and how it might be relaxed. In [18] it is conjectured that, in this case, the existence of the least-squares solution is sufficient to guarantee boundedness. However, the next theorems assure that in the important special case of affine sets Q_i this always holds. An *affine set* (sometimes called a *flat* or *affine manifold*) is a set that has the property that with every two points of the set the whole line through these points is also in the set. Every affine set is a translation of a proper subspace in R^n, and hyperplanes are affine sets of co-dimension one in R^n. Let $\mathcal{H}$ be a *finite family of affine sets* in R^n. Then we define a *sequence of orthogonal projections on* $\mathcal{H}$ to be a sequence $\{x^k\}_{k\geq 0}$ of points such that $x^0 \in R^n$ is arbitrary and, for all $k \geq 0$,

$$x^{k+1} = P_{H_k}(x^k), \tag{71}$$

where $H_k \in \mathcal{H}$ is an arbitrary member of $\mathcal{H}$. Aharoni, Duchet and Wajnryb [2, Theorem 2] proved, and Meshulam [26] simplified the proof, of the following result about boundedness of orthogonal projections onto affine sets.

Theorem 19 *Let $\mathcal{H}$ be a finite family of affine sets in R^n. Then there exists an $r \geq 0$ such that for any sequence $\{x^k\}_{k\geq 0}$ of orthogonal projections on $\mathcal{H}$ there holds, for every k,*

$$\|x^k\| \leqslant \|x^0\| + r. \tag{72}$$

In [2, Theorem 2] this theorem is stated and proved for a more general *quasi-finite* family $\mathcal{H}$. For our purposes we shall need the following variation of this theorem.

Theorem 20 *Let $\mathcal{H}$ be a finite family of hyperplanes in R^n. Then for every point $x \in R^n$ there exists a bounded convex set $C \subseteq R^n$, containing x, which has the property that every orthogonal projection of any point $z \in C$ onto any member of $\mathcal{H}$ belongs to C.*

Proof. Let D be the union of all sequences of orthogonal projections on $\mathcal{H}$ initialized at x. From Theorem 19 it follows that D is bounded, thus, its convex hull $C := \operatorname{conv} D$ is also bounded. Letting $z \in C$, taking any $H \in \mathcal{H}$ and letting $y := P_H(z)$ we need to show that $y \in C$. Since $z \in C$, we have $z = \sum_{i\in I} \alpha_i d^i$, where I is a finite set of indices, $d^i \in D$ and $\alpha_i \geq 0$, for all

$i \in I$, and $\sum_{i\in I} \alpha_i = 1$. Since the orthogonal projection onto a hyperplane is a linear operator, it is not difficult to verify that

$$P_H\left(\sum_{i\in I} \alpha_i d^i\right) = \sum_{i\in I} \alpha_i P_H\left(d^i\right). \tag{73}$$

This, together with the fact that $P_H(d^i) \in D$, for all $i \in I$, yields

$$y = \sum_{i\in I} \alpha_i P_H(d^i) \in \operatorname{conv} D = C, \tag{74}$$

which completes the proof. ■

The next corollary guarantees that if $\mathcal{Q}$ is a family of hyperplanes, then every sequence $\left\{x^k\right\}_{k\geq 0}$, generated by Algorithm 17, is bounded. If Ω is a closed convex set and P_Ω is the orthogonal projection operator onto Ω and if $0 < \sigma < 1$ is a real number, then the *underrelaxed orthogonal projection* operator onto Ω, denoted by $P_{\Omega,\sigma}$, is defined by $P_{\Omega,\sigma}(x) := (1-\sigma)x + \sigma P_\Omega(x)$.

Corollary 21 *Any sequence of underrelaxed orthogonal projections on a finite family $\mathcal{H}$ of hyperplanes is bounded.*

Proof. Let x^0 be the initialization point of the sequence, and let C be the bounded convex set, which contains x^0, whose existence is guaranteed by Theorem 20. Since C is convex and invariant under projections onto $\mathcal{H}$, it is closed also under underrelaxed projections on members of $\mathcal{H}$. Therefore, the sequence is contained in the bounded set C. ■

5 The Entropy Case

In this section we apply our *steered cyclic Bregman projections method*, i.e., Algorithm 14, to the Bregman function $f(x) = -\operatorname{ent} x$, where $\operatorname{ent} x$ is Shannon's entropy function which maps the nonnegative orthant R_+^n into R according to

$$\operatorname{ent} x := -\sum_{j=1}^{n} x_j \log x_j. \tag{75}$$

Here "log" denotes the natural logarithms and, by definition, $0 \log 0 = 0$. The reader is referred to Censor and Zenios [14, Example 2.1.2 and Lemma 2.13] for a verification that $f(x)$ is a Bregman function with zone

$$S_e := \{x \in R^n \mid x_j > 0, \text{ for all } 1 \le j \le n\} \tag{76}$$

and that

$$D_f(x, y) = \sum_{j=1}^{n} x_j (\log (x_j / y_j) - 1) + \sum_{j=1}^{n} y_j. \tag{77}$$

Before applying Algorithm 14 we prove the following three lemmas.

Lemma 22 *Let $f(x) := -\operatorname{ent} x$. Then the Hessian of $D_f(x, y)$, $\nabla^2 D_f(x, y)$, is positive semi-definite on $S_e \times S_e$.*

Proof. To verify that the matrix $\nabla^2 D_f(x, y)$ is positive semi-definite on $S_e \times S_e$ it is sufficient to show that its eigenvalues are nonnegative on $S_e \times S_e$. The eigenvalues are the roots of the equation $|\nabla^2 D_f(x, y) - \lambda I| = 0$ for the variable λ, where I is the $2n \times 2n$ identity matrix and $|\cdot|$ stands for the determinant. From (77) we see that $|\nabla^2 D_f(x, y) - \lambda I| = 0$ becomes

$$\begin{vmatrix} \frac{x_1}{y_1^2} - \lambda & 0 & \cdots & 0 & | & -\frac{1}{y_1} & 0 & \cdots & 0 \\ 0 & \frac{x_2}{y_2^2} - \lambda & \cdots & 0 & | & 0 & -\frac{1}{y_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots & | & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{x_n}{y_n^2} - \lambda & | & 0 & 0 & \cdots & -\frac{1}{y_n} \\ -- & -- & -- & -- & -- & -- & -- & -- & -- \\ -\frac{1}{y_1} & 0 & \cdots & 0 & | & \frac{1}{x_1} - \lambda & 0 & \cdots & 0 \\ 0 & -\frac{1}{y_2} & \cdots & 0 & | & 0 & \frac{1}{x_2} - \lambda & \cdots & 0 \\ \vdots & \vdots & & \vdots & | & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -\frac{1}{y_n} & | & 0 & 0 & \cdots & \frac{1}{x_n} - \lambda \end{vmatrix} = 0. \tag{78}$$

Performing the following elementary operations on (78): add (x_i/y_i) times the $(n+i)$-th column to the i-th column and add $(-x_i/y_i)$ times the i-th row to $(n+i)$-th row, for all $1 \le i \le n$, we obtain a determinant of a $2n \times 2n$ triangular matrix whose main diagonal elements are

$$-\lambda, \ -\lambda, \ \cdots, \ -\lambda, \ \frac{x_1^2+y_1^2}{y_1^2 x_1} - \lambda, \ \frac{x_2^2+y_2^2}{y_2^2 x_2} - \lambda, \ \cdots, \ \frac{x_n^2+y_n^2}{y_n^2 x_n} - \lambda. \quad (79)$$

Since the resulting matrix is triangular, its determinant is the product of its main diagonal elements. Therefore, the eigenvalues of the matrix $\nabla^2 D_f(x,y)$ are

$$0, \ \frac{x_1^2+y_1^2}{y_1^2 x_1}, \ \frac{x_2^2+y_2^2}{y_2^2 x_2}, \ \cdots, \ \frac{x_n^2+y_n^2}{y_n^2 x_n}, \quad (80)$$

which are all nonnegative on $S_e \times S_e$ and this complete the proof. ■

The next lemma shows that the Bregman distance with respect to the entropy function is jointly convex. For a general treatment of convexity of Bregman distance see Bauschke and Borwein [5].

Lemma 23 *Let $f(x) := -\operatorname{ent} x$. For every closed convex set Ω for which $\Omega \cap \operatorname{cl} S_e \neq \emptyset$ the function $d_\Omega^f(x)$ is convex on S_e.*

Proof. The function $d_\Omega^f(x)$ is in this case

$$d_\Omega^f(x) := \min \{ D_f(z,x) \mid z \in \Omega \cap \operatorname{cl} S_e \}. \quad (81)$$

By Lemma 22 the Hessian of the function $D_f(z,x)$ is positive semi-definite on $S_e \times S_e$, hence, by [28, Theorem 4.5], the function $D_f(z,x)$ is jointly convex on $S_e \times S_e$, i.e.,

$$\begin{aligned} & D_f\left(\lambda z^1 + (1-\lambda) z^2, \ \lambda x^1 + (1-\lambda) x^2\right) \\ & \le \lambda D_f\left(z^1, \ x^1\right) + (1-\lambda) D_f\left(z^2, \ x^2\right), \end{aligned} \quad (82)$$

for all $z^1, x^1, z^2, x^2 \in S_e$ and for all $\lambda \in [0,1]$. Now, for any $u^1, u^2 \in S_e$ and any $\alpha \in [0,1]$ we have

$$\begin{aligned} & d_\Omega^f\left(\alpha u^1 + (1-\alpha) u^2\right) \\ & = D_f\left(P_\Omega^f\left(\alpha u^1 + (1-\alpha) u^2\right), \ \alpha u^1 + (1-\alpha) u^2\right) && (83) \\ & \le D_f\left(\alpha P_\Omega^f\left(u^1\right) + (1-\alpha) P_\Omega^f\left(u^2\right), \ \alpha u^1 + (1-\alpha) u^2\right) && (84) \\ & \le \alpha D_f\left(P_\Omega^f\left(u^1\right), \ u^1\right) + (1-\alpha) D_f\left(P_\Omega^f\left(u^2\right), \ u^2\right) && (85) \\ & = \alpha d_\Omega^f\left(u^1\right) + (1-\alpha) d_\Omega^f\left(u^2\right). && (86) \end{aligned}$$

In the above formula, (83) follows from (44); (84) is true because $\alpha P_{\Omega}^{f}(u^1) + (1-\alpha) P_{\Omega}^{f}(u^2)$ belongs to Ω and is not necessarily equal to the projection $P_{\Omega}^{f}(\alpha u^1 + (1-\alpha) u^2)$; (85) follows from the joint convexity (82); and, finally, (86) uses again (44), and this completes the proof. ■

Note that the argument of Lemma 23 applies in a broader setting and every Bregman function whose D_f is jointly convex has the function $d_{\Omega}^{f}(x)$ convex on its zone.

Lemma 24 *Let $f(x) := -\operatorname{ent} x$. If H is any hyperplane in R^n for which $H \cap S_e \neq \emptyset$ then the derivatives $\frac{\partial}{\partial x_j}\left(P_H^f(x)\right)_t$, $1 \leq t, j \leq n$, exist and are continuous on S_e.*

Proof. Let $H = \{x| \ \langle a, \ x\rangle = b\}$ be a given hyperplane such that $H \cap \operatorname{cl} S_e \neq \emptyset$. By [14, Lemma 2.2.1], for every $x \in S_e$ there is a unique real number λ such that

$$\nabla f\left(P_H^f(x)\right) = \nabla f(x) + \lambda a, \tag{87}$$

$$\left\langle a, \ P_H^f(x)\right\rangle = b. \tag{88}$$

Since $(\nabla f(x))_t = \log x_t + 1$, (87) becomes $\log\left(P_H^f(x)\right)_t = \log x_t + \lambda a_t$, for all $1 \leq t \leq n$. Therefore, the system (87)–(88) is

$$\left(P_H^f(x)\right)_t = x_t \exp(\lambda a_t), \text{ for all } 1 \leq t \leq n, \tag{89}$$

$$\sum_{t=1}^{n} a_t \left(P_H^f(x)\right)_t = b. \tag{90}$$

By combining the last two equations we get

$$\sum_{t=1}^{n} a_t x_t \exp(\lambda a_t) - b = 0. \tag{91}$$

Differentiating with respect to λ gives

$$\frac{\partial}{\partial \lambda}\left(\sum_{t=1}^{n} a_t x_t \exp(\lambda a_t) - b\right) = \sum_{t=1}^{n} a_t^2 x_t \exp(\lambda a_t). \tag{92}$$

Equation (91) defines λ as an implicit function of $(x_1, x_2, \cdots, x_n)$ and, since $a_t \neq 0$ for at least one t and $x \in S_e$, the right-hand side of (92) is never zero. This together with the Implicit Function Theorem, see, e.g., Cheney [15, Page 137], yield that the function λ is continuously differentiable with respect to the variables $x_1, x_2, \cdots, x_n$. This differentiability together with equation (89) yield that the derivatives $\frac{\partial}{\partial x_j}\left(P_H^f(x)\right)_t$, $1 \leq t, j \leq n$, exist and are continuous. ■

Calculating the Hessian of f,

$$\nabla^2 f(x) = \nabla^2 \left(\sum_{j=1}^{n} x_j \log x_j\right) = \begin{pmatrix} \frac{1}{x_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_n} \end{pmatrix}, \tag{93}$$

we obtain from all the above, as a direct consequence of Algorithm 14, the following algorithm which minimizes the proximity function $d_{\mathcal{Q}}^f(x) = \sum_{i=0}^{m-1} d_{Q_i}^f(x)$, where $f(x) := -\operatorname{ent} x$ and $\mathcal{Q} = \{Q_i\}_{i=0}^{m-1}$, is a family of closed convex sets for which $Q_i \cap \operatorname{cl} S_e \neq \emptyset$. This sequential algorithm uses the entropy projections $P_{Q_i}^f \equiv P_i^f$, $0 \leq i \leq m-1$, in a cyclic manner. Its convergence theorem, given below, follows from Theorem 16 and Lemmas 23 and 24.

Algorithm 25 ***(The steered cyclic entropy projections method).***

Initialization: *$x^0 \in S_e$ is arbitrary.*

Iterative Step: *Given x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = x^k + \sigma_k \begin{pmatrix} \frac{1}{x_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_n} \end{pmatrix} \left(P_{i(k)}^f(x^k) - x^k\right). \tag{94}$$

Control Sequence: *$\{i(k)\}_{k\geq 0}$ is a cyclic control sequence, i.e., $i(k) = k \bmod m$.*

Steering Parameters: *The sequence $\{\sigma_k\}_{k\geq 0}$ is an m-steering sequence.*

Theorem 26 *Let $f(x) := -\operatorname{ent} x$, and Let $\mathcal{Q} = \{Q_i\}_{i=0}^{m-1}$ be a family of closed convex sets for which $Q_i \cap \operatorname{cl} S_e \neq \emptyset$, for all $0 \leq i \leq m-1$. Assume that*

$x^k \in S_e$, for all $k \geq 0$, holds for all iterates of Algorithm 25. If $\{x^k\}_{k\geq 0}$ is a bounded sequence, generated by Algorithm 25, then the sequence $\left\{d_Q^f(x^k)\right\}_{k\geq 0}$ converges to the minimum of the function $d_Q^f(x)$. If, in addition, we assume that $d_Q^f(x)$ has a unique minimizer then $\{x^k\}_{k\geq 0}$ converges to this minimizer.

At this time we do not know though whether a boundedness result for entropy projections on a finite family $\mathcal{H}$ of hyperplanes, similar to the result described above for the orthogonal case, holds or not.

Acknowledgments. We are grateful to Professor Ron Aharoni from the Department of Mathematics at the Technion – Israel Institute of Technology in Haifa for his collaboration and insightful contributions throughout this research. We thank a referee whose illuminating comments helped to improve this paper. The work of M. Zaknoon on this research is part of his Ph.D. thesis [34]. The research of Y. Censor on the topic of this paper is partially supported by grant No. 592/00 of the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities and by NIH grant No. HL70472. The work of A.R. De Pierro was supported by CNPq grant No. 300969/2003–1 and FAPESP grant No. 2002/07153–2.

References

[1] R. Aharoni and Y. Censor, Block-iterative projection methods for parallel computation of solutions to convex feasibility problems, *Linear Algebra and Its Applications*, **120** (1989), 165–175.

[2] R. Aharoni, P. Duchet and B. Wajnryb, Successive projections on hyperplanes, *Journal of Mathematical Analysis and Applications*, **103** (1984), 134–138.

[3] J.-P. Aubin and A. Cellina, *Differential Inclusions*, Springer-Verlag, Berlin, Germany, 1984.

[4] H.H. Bauschke, The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space, *Journal of Mathematical Analysis and Applications*, **202** (1996), 150–159.

[5] H.H. Bauschke and J.M. Borwein, Joint and Separate Convexity of Bregman Distance, In D. Butnariu, Y. Censor, and S. Reich *Inherently Parallel Algorithms in Feasibility and optimization and their Applications*, (Haifa 2000), Pages 23-36. Elsevier Science, North-Holland, 2001.

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

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

[8] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, USA, 1995.

[9] D.P. Bertsekas, A new class of incremental gradient methods for least squares problems, *SIAM Journal on Optimization*, **7** (1997), 913–926.

[10] D.P. Bertsekas and J.N. Tsitsiklis, Gradient convergence in gradient methods, *SIAM Journal on Optimization*, **36** (2000), 627–642.

[11] L.M. Bregman, The method of successive projections for finding a common point of convex sets, *Soviet Mathematics Doklady*, **6** (1965), 688–692.

[12] C. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Annals of Operations Research*, **105** (2001), 77–98.

[13] Y. Censor, P.P.B. Eggermont and D. Gordon, Strong underrelaxation in Kaczmarz's method for inconsistent systems, *Numerische Mathematik*, **41** (1983), 83–92.

[14] Y. Censor and S.A. Zenios, *Parallel Algorithms: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.

[15] W. Cheney, *Analysis for Applied Mathematics*, Springer-Verlag, New York, USA, 2002.

[16] P.L. Combettes, Inconsistent signal feasibility: Least-squares solutions in a product space, *IEEE Transactions on Signal Processing*, **SP-42** (1994), 2955–2966.

[17] P.L. Combettes, The convex feasibility problem in image recovery, in: *Advances in Imaging and Electron Physics*, P.W. Hawkes, ed., pp. 155–270, Academic Press, New York, NY, USA, 1996.

[18] A.R. De Pierro, From parallel to sequential projection methods and vice versa in convex feasibility: results and conjectures, in: *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, D. Butnariu, Y. Censor, and S. Reich, eds., pp. 187–201, Elsevier Science Publishers, Amsterdam, The Netherlands, 2001.

[19] F. Deutsch and I. Yamada, Minimizing certain convex functions over the intersection of the fixed point sets of nonexpansive mappings, *Numerical Functional Analysis and Optimization*, **19** (1998), 33–56.

[20] J. Eckstein, Approximate iterations in Bregman-function-based proximal algorithms, *Mathematical Programming*, **83** (1998), 113–123.

[21] J. Eckstein, A practical general approximation criterion for methods of multipliers based on Bregman distances, *Mathematical Programming*, Series A, **96** (2003), 61–86.

[22] L. Gubin, B. Polyak and E. Raik, The method of projections for finding the common point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, **7** (1967), 1–24.

[23] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer-Verlag, Berlin, Heidelberg, Germany, 2001.

[24] S. Kaczmarz, Angenäherte Auflösung von Systemen linearer Gleichungen, *Bulletin de l'Académie Polonaise des Sciences at Lettres,* **A35** (1937), 355–357.

[25] P.-L. Lions, Approximation de points fixes de contractions, *C. R. Acad. Sci. Paris Sér. A-B*, **284** (1977), A1357–A1359.

[26] R. Meshulam, On products of projections, *Discrete Mathematics*, **154** (1996), 307–310.

[27] A. Nedić and D.P. Bertsekas, Incremental subgradient methods for non-differentiable optimization, *SIAM Journal on Optimization*, **12** (2001), 109–138.

[28] T. Rockaffellar, *Convex analysis*, Princeton University Press, Princeton, NJ, USA, 1972.

[29] S. Reich, Approximating fixed points of holomorphic mappings, *Mathematica Japonica*, **37** (1992), 457–459.

[30] S. Reich, Approximating fixed points of nonexpansive mappings, *Panamerican Mathematical Journal*, **4** (1994), 23–28.

[31] M.V. Solodov and B.F. Svaiter, An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Mathematics of Operations Research*, **25** (2000), 214–230.

[32] H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics,* John Wiley & Sons, New York, NY, USA, 1998.

[33] R. Wittmann, Approximation of fixed points of nonexpansive mappings, *Archiv der Mathematik*, **58** (1992), 486–491.

[34] M. Zaknoon, *Algorithmic Developments for the Convex Feasibility Problem*, Ph.D. Thesis, University of Haifa, Department of Mathematics, 73 pp., April 2003.