Block-Iterative Algorithms with Underrelaxed Bregman Projections

Yair Censor
Department of Mathematics, University of Haifa,
Mt. Carmel, Haifa 31905, Israel (yair@math.haifa.ac.il).

Gabor T. Herman
Department of Computer Science, The Graduate Center,
City University of New York, 365 Fifth Avenue,
New York, NY 10016, USA (gherman@gc.cuny.edu).

May 16, 2001. Revised: November 25, 2001. Revised: February 14, 2002.

Abstract

The notion of relaxation is well understood for orthogonal projections onto convex sets. For general Bregman projections it was considered only for hyperplanes and the question of how to relax Bregman projections onto convex sets that are not linear (i.e., not hyperplanes or half-spaces) has remained open. A definition of underrelaxation of Bregman projections onto general convex sets is given here which includes as special cases the underrelaxed orthogonal projections and the underrelaxed Bregman projections onto linear sets as given by De Pierro and Iusem. With this new definition we construct a block-iterative projection algorithmic scheme and prove its convergence to a solution of the convex feasibility problem. The practical importance of relaxation parameters in the application of such projection algorithms to real-world problems is demonstrated on a problem of image reconstruction from projections.

Key Words. convex feasibility, projection algorithms, Bregman functions, block-iterative algorithms, underrelaxation.

1 Introduction

The *convex feasibility problem* of finding a point in the non-empty intersection $C := \cap_{i=1}^m C_i \neq \emptyset$ of a family of closed convex subsets $C_i \subseteq R^n$, $1 \leq i \leq m$, of the n-dimensional Euclidean space is fundamental in many areas of mathematics and the physical sciences, see, e.g., Stark and Yang [32], Combettes [15], [16], and references therein. It has been used to model significant real-world problems in image reconstruction from projections, see, e.g., Herman [21], in radiation therapy treatment planning, see Censor, Altschuler and Powlis [7], and in crystallography, see Marks, Sinkler and Landree [28], to name but a few, and has been used under additional names such as *set theoretic estimation* or the *feasible set approach.* A common approach to such problems is to use *projection algorithms*, see, e.g., Bauschke and Borwein [2], which employ *orthogonal projections* (i.e., nearest point mappings) onto the individual sets C_i.

Flexibility in actual use of such projection algorithms is often gained by using *relaxation parameters.* If $P_\Omega(z)$ is the orthogonal projection of a point $z \in R^n$ onto a closed convex set $\Omega \subseteq R^n$, i.e.,

$$P_\Omega(z) := \operatorname{argmin}\{\| z - x \|_2 \quad | \quad x \in \Omega\}, \tag{1}$$

where $\| \cdot \|_2$ is the Euclidean norm in R^n, and if λ is the so-called *relaxation parameter*, then

$$P_{\Omega,\lambda}(z) := (1-\lambda)z + \lambda P_\Omega(z) \tag{2}$$

is the *relaxed projection* of z onto Ω with relaxation λ. In this paper we restrict our attention to the case when $P_{\Omega,\lambda}(z)$ is a convex combination of z and $P_\Omega(z)$, i.e., when $\lambda \in [0,1]$. This is referred to as *underrelaxation.*

The well-known "Projections Onto Convex Sets" (POCS) algorithm for the convex feasibility problem allows such relaxation parameters, see Bregman [5], Gubin, Polyak and Raik [20], Youla [33] and the review papers by Combettes [15], [16]. Starting from an arbitrary initial point $x^0 \in R^n$, the POCS algorithm's iterative step is

$$x^{k+1} = x^k + \lambda_k (P_{C_{i(k)}}(x^k) - x^k), \tag{3}$$

where $\{\lambda_k\}_{k\geq 0}$ are relaxation parameters and $\{i(k)\}_{k\geq 0}$ is a *control sequence*, $1 \leq i(k) \leq m$, for all $k \geq 0$, which determines the set $C_{i(k)}$ onto which the current iterate x^k is projected. The effects of relaxation parameters have been studied theoretically, see, e.g., Censor, Eggermont and Gordon [9]. Their practical effect on early iterates of the POCS algorithm can be dramatic in some real-world situations, as we describe in Section 6 below.

Bregman projections onto closed convex sets were introduced and utilized by Censor and Lent [10], based on Bregman's seminal paper [6], and were subsequently used in a plethora of research works as a tool for building sequential and parallel feasibility and optimization algorithms, see, e.g., Censor and Elfving [8], Censor and Reich [11], Censor and Zenios [13], De Pierro and Iusem [17], Kiwiel [24], [25], Bauschke and Borwein [3] and references therein, to name but a few.

A *Bregman projection* of a point $z \in R^n$ onto a closed convex set $\Omega \subseteq R^n$ with respect to a, suitably defined (see, Definition 7 in Section 7), *Bregman function* f is denoted by $P^f_\Omega(z)$. It is formally defined as

$$P^f_\Omega(z) := \mathrm{argmin}\{D_f(x,z) \mid x \in \Omega \cap \mathrm{cl}S\}, \tag{4}$$

where $\mathrm{cl}S$ is the closure of the open convex set S, which is the *zone* of f, and $D_f(x,z)$ is the so-called *Bregman distance*, defined by

$$D_f(x,z) := f(x) - f(z) - \langle \nabla f(z), x - z \rangle, \tag{5}$$

for all pairs $(x,z) \in \mathrm{cl}S \times S$. If $\Omega \cap \mathrm{cl}S \neq \emptyset$, then (4) defines a unique $P^f_\Omega(z) \in \mathrm{cl}S$, for every $z \in S$, see [13, Lemma 2.1.2]. If, in addition, $P^f_\Omega(z) \in S$, for every $z \in S$, then f is called *zone consistent* with respect to Ω.

Orthogonal projections are a special case of Bregman projections, obtained from (4) by choosing $f(x) = (1/2)\|x\|^2$ and $S = R^n$ (see, e.g., [13, Example 2.1.1]). But in spite of this, relaxation of general (non-orthogonal) Bregman projections has not yet been defined – except for a special case when the set Ω is a half-space, which has been done by De Pierro and Iusem [17] in the following manner. Let, for some $a \in R^n$, $a \neq 0$, and $b \in R$,

$$L = \{x \in R^n \mid \langle a, x \rangle \leq b\} \tag{6}$$

be a half-space. For a $z \notin L$, De Pierro and Iusem [17] define the *underrelaxed Bregman projection* of z onto L, with respect to a Bregman function f and with relaxation parameter $\rho \in [0,1]$ by

$$P^f_{L,\rho}(z) := P^f_{\tilde{L}}(z), \tag{7}$$

where

$$\widetilde{L} = \{x \in R^n \mid \langle a, x\rangle \leq (1-\rho)\langle a, z\rangle + \rho b\}. \tag{8}$$

This means that the relaxed Bregman projection of z onto L is the unrelaxed Bregman projection of z onto a half-space $\widetilde{L}$ whose bounding hyperplane is parallel to that of L and lies between that of L and the point z.

Rewriting (6) as

$$L = \{x \in R^n \mid g(x) \leq 0\}, \tag{9}$$

with $g(x) := \langle a, x\rangle - b$, we can view (7) as the unrelaxed Bregman projection onto the inflated set $\widetilde{L}$ of (8), which can be redefined as

$$\widetilde{L} = \{x \in R^n \mid g(x) \leq \varepsilon\}, \tag{10}$$

where $\varepsilon = (1-\rho)g(z)$. (Note that it is easy to show that if $\Omega = L$ and $f(x) = (1/2)\|x\|^2$, then $P_{\Omega,\rho}(z)$ as defined by (2) is the same as $P^f_{L,\lambda}(z)$ as defined by (7), for any $\lambda \in [0,1]$ and any $z \in R^n$.) This approach of projecting onto an inflated set would not necessarily work for a set Ω defined by a nonlinear function $g(x)$. This can be seen by taking a planar closed convex set Ω which is defined by an ellipse, and considering its inflated set $\widetilde{\Omega}$ to be a confocal ellipse lying between Ω and the point $z \notin \Omega$. Obviously, the orthogonal projection of some $z \notin \Omega$ onto $\widetilde{\Omega}$ in this case is not a relaxed orthogonal projection of z onto Ω because the two projections do not always lie along the same line.

Thus, we ask the following questions: (i) How should one define a relaxed Bregman projection onto a (not necessarily linear) closed convex set? The new definition should, of course, include as special cases at least both the orthogonal case for general convex sets and the underrelaxed Bregman projections onto half-spaces of De Pierro and Iusem [17] mentioned above. (ii) Can such relaxed Bregman projections be incorporated into a Bregman's projection algorithm for the convex feasibility problem? The Bregman's projection algorithm of [6], see also [13, Algorithm 5.8.1], allows only unrelaxed projections, i.e., its iterative step is of the form

$$x^{k+1} = P^f_{C_{i(k)}}(x^k), \quad \text{for all } k \geq 0. \tag{11}$$

(iii) Is it possible to construct a block-iterative Bregman's projection algorithm that will allow relaxed Bregman projections and variable blocks? Such an algorithm, with dynamically changing blocks, will naturally extend earlier block-iterative projection algorithms, such as Eggermont, Herman and Lent's [18] block-iterative ART (Kaczmarz) algorithm, and Aharoni and Censor's [1] block-iterative projections (BIP) method for the convex feasibility problem.

In this paper we constructively answer these three questions. We propose a definition for an underrelaxed Bregman projection onto a closed convex (not necessarily linear) set and prove convergence of a block-iterative projection algorithmic scheme with underrelaxed Bregman projections and dynamically varying blocks. This block-iterative scheme contains, as a new special case, the underrelaxed sequential Bregman's projection algorithm for the convex feasibility problem, generalizing the underrelaxed POCS method. The paper is organized as follows. In Section 2 we define underrelaxed Bregman projections and analyze some of their properties. In Section 3 we present the new block-iterative algorithmic scheme with underrelaxed Bregman projections and prove its convergence in Section 4. In Section 5 we demonstrate the new block-iterative algorithmic scheme by working out in detail the case with underrelaxed entropy projections. Computational experience with any algorithm that uses underrelaxed, non-orthogonal, Bregman projections is still missing, but we provide in Section 6 evidence of the advantages of using underrelaxation parameters when working with orthogonal projections in the real-world application of image reconstruction from projections. For the reader's convenience we attach an Appendix (Section 7) with a summary of definitions and results from the theory of Bregman functions which are used in this paper.

2 Underrelaxation of Bregman Projections

We consider that the underrelaxed Bregman projection with Bregman function f and relaxation parameter $\lambda \in [0,1]$ of a point z onto a closed convex set Ω, denoted by $P^f_{\Omega,\lambda}(z)$, should satisfy

$$\nabla f(P^f_{\Omega,\lambda}(z)) = (1-\lambda)\nabla f(z) + \lambda \nabla f(P^f_{\Omega}(z)). \tag{12}$$

The justification for this is that it makes $P^f_{\Omega,\lambda}(z)$ the appropriate (for λ) *convex combination with respect to the Bregman function* f, as defined by Censor and Reich [11, Definiton 4.1], of z and of $P^f_{\Omega}(z)$. In stating conditions under which (12) is a valid definition of $P^f_{\Omega,\lambda}(z)$ we make use of a result which appears in Bauschke and Borwein [3, Fact 2.9] and is based on Rockafellar's [30, Theorem 26.5], see Theorem 12 in the Appendix (Section 7) of this paper.

Proposition 1 *Let $f : R^n \to R$ be a Bregman function with zone $S = \mathrm{int}(\mathrm{dom}\, f)$ and let $\Omega \subseteq R^n$ be a closed convex set such that $\Omega \cap \mathrm{cl}\, S \neq \emptyset$. If f is a Legendre function, then for any $z \in S$ and any $\lambda \in [0, 1]$ there exists a unique $x \in S$ satisfying*

$$\nabla f(x) = (1 - \lambda)\nabla f(z) + \lambda \nabla f(P_\Omega^f(z)). \tag{13}$$

Proof. Since f is of the Legendre type we have, from Theorem 13 below, that it is zone consistent with respect to Ω. Moreover, [3, Fact 2.9], see also Theorem 12 in the Appendix below, guarantees that $\nabla f(S)$ is equal to the interior of the domain of the conjugate function f^*. Since dom f^* is a convex set (see, e.g., Luenberger [26, Proposition 1, Page 196]), so is its interior and, thus, the right hand side of (13) is in $\nabla f(S)$. Theorem 12 now ensures the existence and uniqueness of an x in S satisfying (13). ∎

In view of this, our definition of $P_{\Omega,\lambda}^f(z)$ is that it is the x whose existence and uniqueness is guaranteed by the proposition. The ability to invert the gradient operator is crucial for the applicability of Proposition 1, as well as for the applicability of the algorithmic formula, see (27) below, which describes our proposed block-iterative algorithmic scheme. Therefore, using functions which are both Bregman and Legendre [3, Remark 5.4], secures both the zone consistency and gradient invertibility. An anonymous referee made the conjecture that in Proposition 1 it may only be necessary to assume that f is of the Bregman/Legendre type, a less restrictive property. Examples of Bregman and Legendre functions are provided in Bauschke and Borwein [3].

Remark 2 *If there exists an $x \in S$ which minimizes*

$$(1 - \lambda)D_f(x, z) + \lambda D_f(x, P_\Omega^f(z)) \tag{14}$$

over $\mathrm{cl}\, S$, then in fact that x satisfies (13), as follows by substituting for D_f using (5) and setting the gradient to zero. This provides additional indication of the reasonableness of our definition of underrelaxed Bregman projections.

For $f(x) = (1/2) \parallel x \parallel^2$ with $S = R^n$, our definition of an underrelaxed Bregman projection coincides with the notion of underrelaxation of orthogonal projections. The next proposition shows that, when Ω is the L of (6), our definition of underrelaxation of Bregman projections coincides with the one given by De Pierro and Iusem in [17], provided that their assumptions and the conditions of Proposition 1 are met. De Pierro and Iusem made the additional assumption that f is not only zone consistent with respect to the bounding hyperplane of L but also with respect to the bounding hyperplane of any half-space $\widetilde{L}$ as defined in (8) (this was termed in [10] *strong zone consistency of f with respect to the bounding hyperplane of L*, see, e.g., [13, Definition 2.2.1]).

Proposition 3 *Let f be a Bregman function with zone S and Ω be a half-space L as in (6) satisfying the conditions of Proposition 1. Assume also that f is strongly zone consistent with respect to the bounding hyperplane of L. Then, for every ρ $(0 \leq \rho \leq 1)$, there exists a λ $(0 \leq \lambda \leq 1)$ such that $P^f_{L,\rho}(z)$ of (7) fulfills (13) with $\Omega = L$, for every $z \in S$.*

Proof. If $z \in S \cap L$, then there is nothing to prove. Therefore, let $z \in S$ be outside the half-space L. From the well-known characterization of Bregman projections onto hyperplanes, see [13, Lemma 2.2.1], we know that the projection $P^f_H(z)$ onto the bounding hyperplane H of the half-space L is uniquely determined, along with the (real) *associated projection parameter* θ, by the system

$$\begin{gathered} \nabla f(P^f_H(z)) = \nabla f(z) + \theta a, \\ \left\langle P^f_H(z), a \right\rangle = b. \end{gathered} \tag{15}$$

We claim that $P^f_L(z) = P^f_H(z)$. To see this, first note that the θ of (15) is negative (because z is outside L, see [13, Lemma 2.2.2, Equation (2.27)]). Now consider any $x \in L$. Multiplying the inequality of (6) by the negative of the θ of (15) and then using the second line and, subsequently, the first line of (15) we get that

$$\left\langle \nabla f(z) - \nabla f(P^f_H(z)), x - P^f_H(z) \right\rangle \leq 0, \text{ for all } x \in L \cap \operatorname{cl} S. \tag{16}$$

This uniquely characterizes $P^f_H(z)$ as the projection $P^f_L(z)$, see [13, Theorem 2.4.2]. Similarly, by letting $\widetilde{H}$ be the bounding hyperplane of $\widetilde{L}$ and using strong zone consistency, we find that the Bregman projection $P^f_{\widetilde{L}}(z)$ of z

onto $\widetilde{L}$ of (8) is in fact $P^f_{\widetilde{H}}(z)$ and is uniquely determined, along with the associated projection parameter $\widetilde{\theta}$, by the system

$$\begin{aligned} \nabla f(P^f_{\widetilde{L}}(z)) &= \nabla f(z) + \widetilde{\theta} a, \\ \left\langle P^f_{\widetilde{L}}(z), a \right\rangle &= (1-\rho)\langle a, z\rangle + \rho b. \end{aligned} \tag{17}$$

Using (7) and the first lines of (15) and (17), we obtain that (recall that $\theta < 0$)

$$\nabla f(P^f_{L,\rho}(z)) = \nabla f(z) + \widetilde{\theta} a = \nabla f(z) + \frac{\widetilde{\theta}}{\theta}(\nabla f(P^f_L(z)) - \nabla f(z)) \tag{18}$$

$$= (1 - \frac{\widetilde{\theta}}{\theta})\nabla f(z) + \frac{\widetilde{\theta}}{\theta}\nabla f(P^f_L(z)). \tag{19}$$

Since θ is negative and by [13, Lemma 2.2.2] $\widetilde{\theta}$ is nonpositive, we have from [13, Lemma 2.2.4] that $\theta \leq \widetilde{\theta}$. These facts guarantee that if we define $\lambda = \widetilde{\theta}/\theta$ then $0 \leq \lambda \leq 1$, which completes the proof. ■

If $f : R^n \to R$ is a Bregman function with zone S and $\Omega \subseteq R^n$ is a closed convex set such that $\Omega \cap \operatorname{cl} S \neq \emptyset$ and if f is zone consistent with respect to Ω, then it is immediate from (4) that P^f_Ω is an *idempotent operator*, i.e.,

$$P^f_\Omega(P^f_\Omega(z)) = P^f_\Omega(z), \tag{20}$$

for any $z \in S$. For underrelaxed projections we have the result of the next proposition, which trivially holds for orthogonal projections.

Proposition 4 *Let f be a Bregman function with zone S and Ω be a closed convex set satisfying the conditions of Proposition 1. Then, for any $z \in S$, we have*

$$P^f_\Omega(P^f_{\Omega,\lambda}(z)) = P^f_\Omega(z), \tag{21}$$

for all $\lambda \in [0,1]$.

Proof. In case $\lambda = 1$, (21) follows from (20). We now assume that $\lambda \in [0,1)$. The projection $P^f_\Omega(z)$ can be characterized (Theorem 10 in Section 7) as the unique element of $\Omega \cap \operatorname{cl} S$ for which

$$\left\langle \nabla f(z) - \nabla f(P^f_\Omega(z)), x - P^f_\Omega(z) \right\rangle \leq 0, \text{ for all } x \in \Omega \cap \operatorname{cl} S. \tag{22}$$

Multiplying this by $(1-\lambda)$ and substituting for $(1-\lambda)\nabla f(z)$ using (12) yields that, for all $x \in \Omega \cap \operatorname{cl} S$,

$$\langle \nabla f(P^f_{\Omega,\lambda}(z)) - \nabla f(P^f_{\Omega}(z)), x - P^f_{\Omega}(z)\rangle \leq 0. \tag{23}$$

Using again the characterization of Theorem 10 in Section 7 we get (21). ■

3 A Block-Iterative Algorithmic Scheme With Underrelaxed Bregman Projections

In this section we propose a *block-iterative algorithmic scheme* with underrelaxed Bregman projections for the solution of the convex feasibility problem. By *block-iterative* we mean that, at the k-th iteration, the next iterate x^{k+1} is generated from the current iterate x^k by using a subset (called a block) of the family of sets $\{C_i\}_{i=1}^m$ of the convex feasibility problem, see, e.g., [13, Section 1.1.3]. We use the term *algorithmic scheme* to emphasize that different specific algorithms may be derived by different choices of Bregman functions, and by various block structures. For example, if all blocks consist of a single set, then our scheme gives rise to a sequential row-action type algorithm (consult [13, Definition 6.2.1] for this term). Taking the other extreme, if we let every block contain all sets, then we obtain a fully simultaneous algorithm. Such a block-iterative scheme for the convex feasibility problem was first proposed by Aharoni and Censor [1], using orthogonal projections onto convex sets. That block-iterative projections (BIP) method generalizes the sequential POCS method of Bregman [5], Gubin, Polyak and Raik [20] (see also Stark and Yang [32] and Censor and Zenios [13] for many more related references). Our proposed block-iterative scheme extends Aharoni and Censor's BIP method by employing underrelaxed Bregman projections which contain the underrelaxed orthogonal projections as a special case.

Appealing again to the definition of a convex combination with respect to a Bregman function f as defined by Censor and Reich [11, Definiton 4.1], the natural formula for a block-iterative step using underrelaxed Bregman projections is

$$\nabla f(x^{k+1}) = \sum_{i=1}^{m} v_i^k \nabla f(P^f_{C_i,\lambda_i^k}(x^k)), \tag{24}$$

where x^k is the k-th iterate, $\lambda_i^k \in [0,1]$ is the relaxation parameter used in the underrelaxed Bregman projection onto the set C_i during the k-th iterative step and the v_i^k are the weights of the convex combination for the k-th iterative step (i.e., $v_i^k \geq 0$ for $1 \leq i \leq m$ and $\sum_{i=1}^m v_i^k = 1$). Note that under the assumptions of Proposition 1, if $x^k \in S$ then x^{k+1} is uniquely defined by (24) and is also in S.

To simplify notation, from now on we use P_i^f to abbreviate $P_{C_i}^f$. Further, we observe that, according to (12),

$$\nabla f(x^{k+1}) = \sum_{i=1}^m v_i^k \left((1-\lambda_i^k)\nabla f(x^k) + \lambda_i^k \nabla f(P_i^f(x^k))\right). \tag{25}$$

Defining $w_i^k = v_i^k \lambda_i^k$, for $1 \leq i \leq m$, and introducing

$$w_{m+1}^k = 1 - \sum_{i=1}^m w_i^k \quad \text{and} \quad C_{m+1} = R^n, \tag{26}$$

we get that

$$\nabla f(x^{k+1}) = \sum_{i=1}^{m+1} w_i^k \nabla f(P_i^f(x^k)), \tag{27}$$

with $w_i^k \geq 0$ for $1 \leq i \leq m+1$ and $\sum_{i=1}^{m+1} w_i^k = 1$.

4 A Convergence Theorem

The following theorem establishes the convergence to a solution of the convex feasibility problem of a sequence generated by any block-iterative algorithm with underrelaxed Bregman projections. The method of proof is closely related to previous proofs of other results in this field, see, e.g., Bauschke and Borwein [3, Theorem 8.1], Censor and Reich [11, Theorem 3.1]. We will make use of a further condition on the w_i^k of (27).

Condition 5 *Let w_i^k be real numbers for $k \geq 0$ and $1 \leq i \leq m$ and for each k let*

$$I(k) := \{i \mid 1 \leq i \leq m,\ w_i^k > 0\}. \tag{28}$$

(i) There exists an $\varepsilon > 0$ such that $w_i^k \geq \varepsilon$ for all $k \geq 0$ and $i \in I(k)$.
(ii) Each i, $1 \leq i \leq m$, is included in infinitely many sets $I(k)$.

A practitioner might desire to rephrase Condition 5 in terms of the weights v_i^k and the relaxation parameters λ_i^k using (26) and the line above it. Condition 5(i) states that, for some positive ε, if v_i^k and λ_i^k are both positive, then they are both greater or equal ε. It should be noted, however, that Condition 5(i) is stronger then the condition used by Aharoni and Censor [1] regarding the weights they used in their block-iterative projections (BIP) method. The (weaker) condition that they use is that "for all $i = 1, 2, \ldots, m$, the series $\sum_{k=0}^{\infty} v_i^k = +\infty$". The purpose of our condition, as well as that of the condition of [1], is to guarantee that none of sets C_i is "gradually ignored" by ever diminishing weights. We do not know whether our convergence result, presented below, can be strengthened by using a condition similar to that of [1]. Notice also that if $\lambda_i^k = 1$ for all $i = 1, 2, \ldots, m$ and all $k \geq 0$, then no underrelaxation takes place and $w_{m+1}^k = 0$ for all $k \geq 0$ leaving only the weights v_i^k to affect the algorithm's progress. Finally, observe that the sequential algorithm is obtained from (27) by choosing, for every $k \geq 0$, the weights

$$v_i^k = \begin{cases} 1, & \text{if } \; i = i(k), \\ 0, & \text{otherwise,} \end{cases} \tag{29}$$

where $\{i(k)\}_{k \geq 0}$ is a control sequence such as, e.g., the *cyclic control* defined by $i(k) = k(\operatorname{mod} m) + 1$, for all $k \geq 0$.

Theorem 6 *Let $f : R^n \to R$ be a Bregman function and let $S = \operatorname{int}(\operatorname{dom} f)$ be its zone. Let $C_i \subseteq R^n$ be closed convex sets such that $\cap_{i=1}^m C_i \cap \operatorname{cl} S \neq \emptyset$. Assume that f is also a Legendre function. For $k \geq 0$, let w_i^k be nonnegative for $1 \leq i \leq m+1$ such that $\sum_{i=1}^{m+1} w_i^k = 1$ and Condition 5 is satisfied. Then the sequence $\{x^k\}_{k \geq 0}$ generated by (27) from any $x^0 \in S$ converges to a point $x^* \in \cap_{i=1}^m C_i \cap \operatorname{cl} S$.*

Proof. The well-definedness of the algorithm described by (27) can be shown by a straightforward generalization of the proof of Proposition 1 (in which (27) is replaced by (13)). Legendreness of the function f also ensures, by Theorem 13 below, zone consistency of f with respect to each set C_i – a fact which is repeatedly used in this proof. Using Equations (5) and (27) we have, for every $k \geq 0$ and for any $x \in \operatorname{cl} S$,

$$D_f(x, x^{k+1}) = \sum_{i=1}^{m+1} w_i^k \left(f(x) - f(x^{k+1}) - \langle \nabla f(P_i^f(x^k)), x - x^{k+1} \rangle \right). \tag{30}$$

By repeated application of (5) to the expression inside the parentheses on the right–hand side of (30) we obtain

$$D_f(x, x^{k+1}) = \sum_{i=1}^{m+1} w_i^k \left(D_f(x, P_i^f(x^k)) - D_f(x^{k+1}, P_i^f(x^k)) \right). \tag{31}$$

Therefore,

$$\begin{aligned} D_f(x, x^k) - D_f(x, x^{k+1}) &= \sum_{i=1}^{m+1} w_i^k D_f(x^{k+1}, P_i^f(x^k)) \\ &+ \sum_{i=1}^{m+1} w_i^k \left(D_f(x, x^k) - D_f(x, P_i^f(x^k)) \right). \end{aligned} \tag{32}$$

For any point $x \in C_i \cap \operatorname{cl} S$, the difference under the sum in the last line fulfills

$$D_f(x, x^k) - D_f(x, P_i^f(x^k)) \geq D_f(P_i^f(x^k), x^k) \geq 0. \tag{33}$$

This follows from well-known inequalities in the theory of Bregman distances. The left-hand inequality in (33) follows by replacing z, y and Ω in Theorem 9 of Section 7 by x, x^k and C_i, respectively, and the nonnegativity in (33) follows from [13, Lemma 2.1.1]. Since all quantities on the right-hand side of (32) are nonnegative, we conclude from (32) that, for any point $x \in \cap_{i=1}^m C_i \cap \operatorname{cl} S$,

$$D_f(x, x^{k+1}) \leq D_f(x, x^k), \quad \text{for all} \quad k \geq 0, \tag{34}$$

which means that the sequence $\{x^k\}_{k\geq 0}$ is D_f-*Fejér-monotone* with respect to $\cap_{i=1}^m C_i$ and implies that $\{x^k\}_{k\geq 0}$ is bounded, see [13, p. 108]. Therefore, to conclude the proof we will show the following: (i) if there exist a cluster point x^* in $C = \cap_{i=1}^m C_i$, then it is the limit of the sequence, and (ii) every cluster point must belong to C.

We first make the observation that, for any $x \in C \cap \operatorname{cl} S$, (34) and the nonnegativity of $\{D_f(x, x^k)\}_{k\geq 0}$ guarantee the existence of the limit

$$\lim_{k\to\infty} D_f(x, x^k) = \theta. \tag{35}$$

To prove (i) let $x^* \in C$ be a cluster point of $\{x^k\}_{k\geq 0}$ and assume that x^{**} is another cluster point, i.e.,

$$\lim_{\substack{k\to\infty \\ k\in K_1}} x^k = x^* \quad \text{and} \quad \lim_{\substack{k\to\infty \\ k\in K_2}} x^k = x^{**}, \tag{36}$$

with infinite $K_1 \subseteq N$ and infinite $K_2 \subseteq N$ and $N := \{0, 1, 2, \cdots\}$. Since $x^k \in S$, for all $k \geq 0$, $x^* \in \operatorname{cl} S$ and so (35) holds for $x = x^*$ (for some θ). Applying the property of Bregman functions given by Definition 7(iv) in Section 7 to the subsequence defined by $k \in K_1$, we get that in fact

$$\lim_{k\to\infty} D_f(x^*, x^k) = 0, \tag{37}$$

which is true, in particular, for the subsequence defined by $k \in K_2$. Then, using another property of Bregman functions, given by Definition 7(v) in Section 7, $x^* = x^{**}$ follows.

To prove (ii) assume, by way of negation, that

$$\lim_{l\to\infty} x^{k_l} = x^* \quad \text{and} \quad x^* \notin C. \tag{38}$$

Define

$$I_{in} := \{i \mid 1 \leq i \leq m, \quad x^* \in C_i\}, \tag{39}$$

$$I_{out} := \{i \mid 1 \leq i \leq m, \quad x^* \notin C_i\}. \tag{40}$$

Because of Condition 5(ii), we may assume, without loss of generality (passing to a subsequence if necessary) that, for every $l = 1, 2, \cdots$,

$$I(k_l) \cup I(k_l + 1) \cup \cdots \cup I(k_{l+1} - 1) = \{i \mid 1 \leq i \leq m\}. \tag{41}$$

For every $l = 1, 2, \cdots$, let μ_l be the smallest element in the set

$$\{k_l, k_l + 1, k_l + 2, \cdots, k_{l+1} - 1\} \tag{42}$$

such that

$$I(\mu_l) \cap I_{out} \neq \emptyset. \tag{43}$$

Such an element exists by (41) and since, by (38) and (40), $I_{out} \neq \emptyset$.

We want to show now that the sequence $\{x^{\mu_l}\}_{l\geq 0}$ also converges to x^*. By definition, $k_l \leq \mu_l$ for all $l = 1, 2, \cdots$. If $\nu \in [k_l, \mu_l)$, then

$$I(\nu) \subseteq I_{in} \tag{44}$$

and so, from (32), for any $x \in \operatorname{cl} S$,

$$\begin{aligned} D_f(x, x^\nu) - D_f(x, x^{\nu+1}) = &\sum_{i \in I_{in}} w_i^\nu D_f(x^{\nu+1}, P_i^f(x^\nu)) \\ &+ \sum_{i \in I_{in}} w_i^\nu \left(D_f(x, x^\nu) - D_f(x, P_i^f(x^\nu)) \right). \end{aligned} \tag{45}$$

For any point $x \in \cap_{i \in I_{in}} C_i \cap \operatorname{cl} S$, it follows from (33) and (45) that, for $\nu \in [k_l, \mu_l)$,

$$D_f(x, x^{\nu+1}) \leq D_f(x, x^\nu). \tag{46}$$

In other words, with x replaced by x^*, we have for all $l = 1, 2, \cdots$,

$$\begin{aligned} 0 &\leq D_f(x^*, x^{\mu_l}) \leq D_f(x^*, x^{\mu_l - 1}) \\ &\leq \cdots \leq D_f(x^*, x^{k_l + 1}) \leq D_f(x^*, x^{k_l}). \end{aligned} \tag{47}$$

Letting $l \to \infty$ in (47) yields, by (38) and Definition 7(iv) in Section 7,

$$\lim_{l \to \infty} D_f(x^*, x^{\mu_l}) = 0. \tag{48}$$

As a subsequence of the whole sequence $\{x^k\}_{k\geq 0}$ which is bounded, $\{x^{\mu_l}\}_{l\geq 0}$ is bounded, thus has a cluster point. Combining Definition 7(v) in Section 7 with (48) shows that any convergent subsequence of $\{x^{\mu_l}\}_{l\geq 0}$ must converge to x^*, hence

$$\lim_{l \to \infty} x^{\mu_l} = x^*. \tag{49}$$

From (43) it follows that there exists an index $\hat{\imath} \in I_{out}$ such that $\hat{\imath} \in I(\mu_l)$ for infinitely many indices l. Removing from the sequence $\{\mu_l\}_{l\geq 0}$ all elements μ_l for which $\hat{\imath} \notin I(\mu_l)$, we end up with a new infinite sequence $\{\mu_l\}_{l\geq 0}$, such that $\hat{\imath} \in I(\mu_l) \cap I_{out}$, for $l = 1, 2, \cdots$. Taking an arbitrary $x \in C \cap \operatorname{cl} S$, consider the limits of both sides of (32) for the new sequence $\{\mu_l\}_{l\geq 0}$. Due to

(35), the left-hand side converges to zero and, therefore, so must the right-hand side. Since all quantities on the right-hand side are nonnegative and $w_{\hat{\imath}}^{\mu_l} \geq \varepsilon > 0$ (for all $l = 1, 2, \cdots$) by Condition 5(i), we obtain that

$$\lim_{l\to\infty}(D_f(x, x^{\mu_l}) - D_f(x, P_{\hat{\imath}}^f(x^{\mu_l}))) = 0. \tag{50}$$

From (33) we obtain

$$\lim_{l\to\infty} D_f(P_{\hat{\imath}}^f(x^{\mu_l}), x^{\mu_l}) = 0. \tag{51}$$

If we could show that $\{P_{\hat{\imath}}^f(x^{\mu_l})\}_{l\geq 0}$ is bounded then (49) and (51) would imply, by using again Definition 7(v) in Section 7, that

$$\lim_{l\to\infty} P_{\hat{\imath}}^f(x^{\mu_l}) = x^*, \tag{52}$$

which means that $x^* \in C_{\hat{\imath}}$, yielding the sought-after contradiction with the choice of $\hat{\imath}$ made above.

Therefore, we conclude the proof by showing that $\{P_{\hat{\imath}}^f(x^{\mu_l})\}_{l\geq 0}$ is bounded. Indeed, (33) for $i = \hat{\imath}$, with $k = \mu_l$ and for $x \in C \cap \operatorname{cl} S$, shows that

$$D_f(x, P_{\hat{\imath}}^f(x^{\mu_l})) \leq D_f(x, x^{\mu_l}) - D_f(P_{\hat{\imath}}^f(x^{\mu_l}), x^{\mu_l}), \quad \text{for every} \quad l \geq 0. \tag{53}$$

Applying (34) and (51) to (53) shows that $\{D_f(x, P_{\hat{\imath}}^f(x^{\mu_l}))\}_{l\geq 0}$ is bounded, which, by Definition 7(iii) in Section 7, implies that $\{P_{\hat{\imath}}^f(x^{\mu_l})\}_{l\geq 0}$ is bounded, and this concludes the proof. ■

5 An Example: Block-Iterative Underrelaxed Entropy Projections

A well-known Bregman function is the negative "$x \log x$" entropy (also called: Shannon's entropy) function, see [13, Example 2.1.2] and the many references given to the literature on this topic in that book or consult the book by Fang, Rajasekera and Tsao [19] and its references. The "$x \log x$" entropy has been used in numerous applications in science and engineering up to and including recent work in the field of computational machine learning, see, e.g., Collins, Shapire and Singer [14]. It is denoted by $\operatorname{ent} x$ and maps the nonnegative orthant R_+^n into R according to

$$\operatorname{ent} x := -\sum_{j=1}^{n} x_j \log x_j, \tag{54}$$

where "log" denotes the natural logarithmic function and, by definition, $0 \log 0 = 0$. Its negative, $f(x) = \sum_{j=1}^{n} x_j \log x_j$, is a Bregman function with zone $S = \operatorname{int} R_+^n$, see [13, Lemma 2.1.3], the j-th component of whose gradient is $\partial f / \partial x_j = 1 + \log x_j$.

In order to derive a block-iterative algorithm with underrelaxed Bregman entropy projections, for the iterative solution of a linear system of equations $Ax = b$, we consider the sets

$$C_i = \{x \mid \langle a^i, x \rangle = b_i\}, \text{ for } i = 1, 2, \cdots, m, \tag{55}$$

where $a^i \in R^n$ is the i-th column of the transposed matrix A^T and $b_i \in R$ is the i-th component of $b \in R^m$. The iterative step (27) takes the form

$$\log x_j^{k+1} = \sum_{i=1}^{m+1} w_i^k \log(P_i^f(x^k))_j, \text{ for } j = 1, 2, \cdots, n. \tag{56}$$

Using the first line of (15) (with H, z, a and θ replaced by C_i, x^k, a^i and θ_i^k, respectively), substituting into (56) and taking exponents we obtain

$$x_j^{k+1} = x_j^k \prod_{i=1}^{m} \exp(w_i^k \theta_i^k a_j^i) \text{ for } j = 1, 2, \cdots, n, \tag{57}$$

where θ_i^k is the Bregman parameter associated with the "entropy projection" of x^k onto the i-th hyperplane C_i. If one replaces in the iterative step (57) the θ_i^k's with the quantities

$$d_i^k := \log \frac{b_i}{\langle a^i, x^k \rangle}, \tag{58}$$

for all i and all k, then the resulting formula resembles the iterative step formula of the block-iterative MART algorithm of Censor and Segman [12], see also [13, Algorithm 6.7.1, Equation (6.124)], the difference being the lack of underrelaxation parameters and of variable block structure and composition in the latter.

6 On the Practical Usefulness of Underrelaxation Parameters

In this section we demonstrate the importance of underrelaxation parameters in the field of image reconstruction from projections. Projection algorithms have been used to solve the fully discretized model in this field and experimental work has shown again and again that there are great advantages in using underrelaxation of the projections. For a recent example in the area of Positron Emission Tomography (PET) see Obi *et al.* [29] and in the area of Electron Microscopy see Marabini, Herman and Carazo [27]. Since in such practical applications the data are physically collected and so the feasibility condition in Theorem 6 cannot be guaranteed, we report here on an experiment which illustrates the usefulness of underrelaxation when all conditions of Theorem 6 are satisfied.

The experiment has been performed with the algebraic reconstruction technique (ART) described in Herman, Matej and Carvalho [23, Equation (6)] for the purpose of image reconstruction from X-ray data obtained by a scanner utilizing a helical cone-beam data collection geometry. In terms of our block-iterative step formula (24), $f(x) = (1/2)\|x\|^2$, the C_i are hyperplanes and, for every $k \geq 0$, $v_i^k = 1$ for exactly one $i = i(k)$ and is zero otherwise, and the relaxation parameters $\lambda_i^k = \lambda$ are constant. We will be comparing the values $\lambda = 1$, that is no relaxation, with $\lambda = 0.01$, which amounts to quite strong underrelaxation. The number of hyperplanes is $m = 4,915,000$ and the dimensionality of the image vector x is $n = 965,887$. To insure feasibility, we used the object reconstructed in [23] from the not necessarily feasible data used in that paper. (In other words, we replaced the system of equations $Ax = b$ that was treated in [23] by the system $Ax = As$, where s is the output of the algorithm reported in [23].) This reconstructed object is to be interpreted as values within a rectangular region of the three-dimensional space, a graphical representation of a single slice through this region is shown on the left of Figure 1. In this graphical representation all values less than or equal to 1.00 are shown as black, all values grater than equal to 1.04 are shown as white, with greyness levels representing the intermediate values. For our purposes, this previously reconstructed object is the "phantom" (test image) which is the object (vector) in the intersection of $4,915,000$ hyperplanes whose descriptions are known to our program.

Under the conditions of this special case of (24), it is known that the algorithm (provided that it is started with the same vector) should, in the limit, converge to the same vector irrespective of which of the two investigated values of the relaxation parameter is chosen (assuming perfect computer accuracy), this follows, e.g., from Herman, Lent and Lutz [22, Corollary 1] or from Bauschke *et al.* [4, Fact 2.2]. However, for such large problems, the algorithm is computationally intensive and so it is important that one should get to a reasonable solution in relatively fewer steps. For those who have not had experience with such projection algorithms it may come as a surprise that underrelaxation is actually useful for this purpose. We illustrate this in Figure 1 in which the central and right images show our new reconstructions, using no relaxation, i.e., $\lambda = 1$, and underrelaxation, i.e., $\lambda = 0.01$, respectively. The iteration index k at which the algorithm was stopped is the same in both cases $k = 16m$, i.e., we have cycled through the data 16 times. The same slice through the three-dimensional region is shown in all three images, represented in the same way. The quality of the underrelaxed reconstruction is so good that it is practically indistinguishable from the phantom, this is certainly not the case for the reconstruction with no relaxation. This is also reflected by numerical calculations: considering only those locations in space (not only in the slice shown in Figure 1) for which the values of the phantom are in the range $[1.00, 1.04]$, the Euclidean distance between the phantom and reconstruction is 2.2 in the underrelaxed case and it is 6.9 in the no relaxation case.

Thus this experiment, satisfying the conditions of Theorem 6, confirms the previously reported results in applications: a small relaxation parameter is likely to be very useful in allowing us to get to a high quality reconstruction faster than it is possible with no relaxation.

Figure 1. Slices of the phantom (left), reconstruction with no relaxation (center) and reconstruction with underrelaxation (right). See the text for details.

7 Appendix: Some Definitions and Results from Bregman Function Theory

In this Appendix we review some definitions and results from the theory of Bregman functions used in this paper.

Definition 7 *Let S be a nonempty open convex set in R^n with closure* $\operatorname{cl} S$*. Let $f : \operatorname{cl} S \to R$ be a differentiable function and define $D_f(x,z) : \operatorname{cl} S \times S \to R$ by*

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

We say that f is a Bregman function with zone S and that D_f is the Bregman distance associated with it if the following conditions are satisfied:

(i) f is continuous and strictly convex on $\operatorname{cl} S$*;*

(ii) f is continuously differentiable on S;

(iii) for any $x \in \operatorname{cl} S$ the level sets $\{y \in S \mid D_f(x,y) \le \alpha\}$ are bounded.

(iv) if $y^k \in S$ and $\lim_{k\to\infty} y^k = y^*$, then $\lim_{k\to\infty} D_f(y^*, y^k) = 0$;

(v) if $x^k \in \operatorname{cl} S$ and $y^k \in S$, with $\{x^k\}$ bounded, $\lim_{k\to\infty} y^k = y^*$ and $\lim_{k\to\infty} D_f(x^k, y^k) = 0$, then $\lim_{k\to\infty} x^k = y^*$.

Remark 8 *(i) It can be shown that, if the Bregman function f is separable, then the condition that $\{x^k\}$ be bounded in Definition 7(v) is redundant.*
(ii) As noted by Bauschke and Borwein [3], conditions (i)–(v) of Definition 7 imply that for any $y \in S$ the level sets $\{x \in \operatorname{cl} S \mid D_f(x,y) \le \alpha\}$ are also bounded.
(iii) Solodov and Svaiter [31] showed recently that Condition (v) of Definition 7 is redundant (i.e., it follows from the remaining conditions).

Let Ω be a closed convex set in R^n and $z \in S$ a given point. The *Bregman projection* of z onto Ω is the point $P_\Omega^f(z) \in \Omega$ which minimizes $D_f(x,z)$ over all $x \in \Omega \cap \operatorname{cl} S$. Bregman projections exist and are unique provided that the set Ω is closed and convex and that $\Omega \cap \operatorname{cl} S$ is nonempty (see, e.g., [13, Lemma 2.1.2]). Furthermore, we assume that $P_\Omega^f(z) \in S$ whenever $z \in S$ (this is commonly called *zone consistency*). The useful inequality expressed in the next theorem then holds, see, e.g., [13, Theorem 2.4.1].

Theorem 9 *Let f be a Bregman function with zone S and let $\Omega \subseteq R^n$ be a closed convex set such that $\Omega \cap \operatorname{cl} S \neq \emptyset$. Assume that f is zone consistent* *with respect to* Ω *and let* $z \in \Omega \cap \operatorname{cl} S$ *be given. Then for any* $y \in S$ *the inequality*

$$D_f(z,y) - D_f(z, P_\Omega^f(y)) \geq D_f(P_\Omega^f(y), y) \tag{59}$$

holds.

The next result is a characterization of Bregman projections onto convex sets, given in [13, Theorem 2.4.2].

Theorem 10 *Under the assumptions of Theorem 9, for any* $y \in S$*, the point* $P_\Omega^f(y)$ *is the Bregman projection of* y *onto* Ω *with respect to* f *if and only if*

$$\left\langle \nabla f(y) - \nabla f(P_\Omega^f(y)), x - P_\Omega^f(y) \right\rangle \leq 0, \text{ for all } x \in \Omega \cap \operatorname{cl} S. \tag{60}$$

We make use in this paper of Legendre functions and some of their basic properties. Therefore, we give here a definition from Bauschke and Borwein [3, Definition 2.8], see also Rockafellar [30, Section 26].

Definition 11 *Suppose that* f *is a closed convex proper function on* R^n*. Then* f *is a Legendre function if it is both essentially smooth and essentially strictly convex, i.e.,* f *satisfies the following properties:*
(i) int(dom f*)*$\neq \emptyset$,
(ii) f *is differentiable on int(dom* f*),*
(iii) for every $x \in$*bd(dom* f*) and every* $y \in$*int(dom* f*)*

$$\lim_{t \to 0^+} \langle \nabla f(x + t(y-x)), y - x \rangle = -\infty, \tag{61}$$

(iv) f *is strictly convex on int(dom* f*).*

The next result, characterizing and describing Legendre functions, is quoted from [3, Fact 2.9] and based on [30, Theorem 26.5].

Theorem 12 *A convex function* f *is a Legendre function if and only if its conjugate* f^* *is. In this case, the gradient mapping*

$$\nabla f : int(dom f) \to int(dom f^*) \tag{62}$$

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

Finally we quote from Bauschke and Borwein [3, Theorem 3.14] the following important fact.

Theorem 13 *If f is a Legendre function and $S = \text{int}(\text{dom}\, f)$, then f is zone consistent with respect to any closed convex set Ω such that $\Omega \cap clS \neq \emptyset$.*

Acknowledgments. Initial work on this paper was in collaboration with Charles Byrne from the Department of Mathematical Sciences at the University of Massachusetts Lowell and we gratefully acknowledge that his contributions greatly helped in formulating the material that appears in this final version. We thank Bruno Motta de Carvalho for his help with creating the images. Part of the work of Y. Censor was done during a visit to the Department of Mathematics of the University of Linköping, Linköping, Sweden. The support and hospitality of Tommy Elfving and Åke Björck, head of the Numerical Analysis Group there, are gratefully acknowledged. We thank the referees for their constructive and helpful comments on earlier versions of this paper.

The research of Y. Censor was supported by NIH grant HL-28438 and by grants 293/97 and 592/00 of the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities. The research of G.T. Herman was supported by NIH grants HL-28438 and HL-70472.

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*, 120:165–175, 1989.

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

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

[4] H.H. Bauschke, F. Deutsch, H. Hundal and S.-H. Park, Fejér monotonicity and weak convergence of an accelerated method of projections, in:

Constructive, Experimental, and Nonlinear Analysis (Limoges, 1999), American Mathematical Society, Providence, RI, USA, 2000, pp. 1–6.

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

[6] 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, *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[7] Y. Censor, M.D. Altschuler and W.D. Powlis, On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning, *Inverse Problems*, 4 (1988), 607–623.

[8] Y. Censor and T. Elfving, A multiprojections algorithm using Bregman projections in a product space, *Numerical Algorithms*, 8:221–239, 1994.

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

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

[11] Y. Censor and S. Reich, Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization, *Optimization*, 37:323–339, 1996.

[12] Y. Censor and J. Segman, On block-iterative entropy maximization, *Journal of Information and Optimization Sciences*, 8:275–291, 1987.

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

[14] M. Collins, R.E. Shapire and Y. Singer, Logistic regression, AdaBoost and Bregman distances, *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, Stanford University, Stanford, CA, USA, June 28–July 1, 2000, pp. 158–169.

[15] P.L. Combettes, The foundations of set-theoretic estimation, *Proceedings of the IEEE*, 81:182–208, 1993.

[16] P.L. Combettes, The convex feasibility problem in image recovery, *Advances in Imaging and Electron Physics*, 95:155–270, 1996.

[17] A.R. De Pierro and A.N. Iusem, A relaxed version of Bregman's method for convex programming, *Journal of Optimization Theory and Applications*, 51:421–440, 1986.

[18] P.P.B. Eggermont, G.T. Herman and A. Lent, Iterative algorithms for large partitioned linear systems, with applications to image reconstruction, *Linear Algebra and Its Applications*, 40:37–67, 1981.

[19] S.-C. Fang, R.J. Rajasekera and H.-S.J. Tsao, *Entropy Optimization and Mathematical Programming*, Kluwer Academic Publishers, Boston, MA, USA, 1997.

[20] 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*, 7:1–24, 1967.

[21] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, USA, 1980.

[22] G.T. Herman, A. Lent and P.H. Lutz, Relaxation methods for image reconstruction, *Communications of the ACM*, 21:152–158, 1978.

[23] G.T. Herman, S. Matej and B.M. Carvalho, Algebraic reconstruction techniques using smooth basis functions for helical cone-beam tomography, in: D. Butnariu, Y. Censor and S. Reich (Editors), *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, Elsevier Science Publishers, Amsterdam, The Netherlands, 2001, pp. 307–324.

[24] K. Kiwiel, Generalized Bregman projections in convex feasibility problems, *Journal of Optimization Theory and Applications*, 96:139–157, 1998.

[25] K. Kiwiel, Free–steering relaxation methods for problems with strictly convex costs and linear constraints, *Mathematics of Operations Research*, 22:326–349, 1997.

[26] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, NY, USA, 1969.

[27] R. Marabini, G.T. Herman and J.M. Carazo, 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs), *Ultramicroscopy* 72:53-65, 1998.

[28] L.D. Marks, W. Sinkler and E. Landree, A feasible set approach to the crystallographic phase problem, *Acta Crystallographica*, A55:601–612, 1999.

[29] T. Obi, S. Matej, R.M. Lewitt and G.T. Herman, 2.5D simultaneous multislice reconstruction by series expansion methods from Fourier rebinned PET data, *IEEE Transactions on Medical Imaging,* 19:474–484, 2000.

[30] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, USA, 1970.

[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*, 25:214–230, 2000.

[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] D.C. Youla, Mathematical theory of image restoration by the method of convex projections, in: H. Stark (Editor), *Image Recovery: Theory and Applications*, Academic Press, Orlando, Florida, USA, 1987, pp. 29–77.