AVERAGING STRINGS OF SEQUENTIAL ITERATIONS FOR CONVEX FEASIBILITY PROBLEMS

Y. Censor[a*], T. Elfving[b†] and G.T. Herman[c‡]

[a]Department of Mathematics, University of Haifa,
Mt. Carmel, Haifa 31905, Israel

[b] Department of Mathematics, Linköping University,
SE-581 83 Linköping, Sweden

[c] Department of Computer and Information Sciences,
Temple University, 1805 North Broad Street, Philadelphia, PA. 19122-6094, USA

An algorithmic scheme for the solution of convex feasibility problems is proposed in which the end-points of strings of sequential projections onto the constraints are averaged. The scheme, employing Bregman projections, is analyzed with the aid of an extended product space formalism. For the case of orthogonal projections we give also a relaxed version. Along with the well-known purely sequential and fully simultaneous cases, the new scheme includes many other inherently parallel algorithmic options depending on the choice of strings. Convergence in the consistent case is proven and an application to optimization over linear inequalities is given.

1. INTRODUCTION

In this paper we present and study a new algorithmic scheme for solving the *convex feasibility problem* of finding a point x^* in the nonempty intersection $C = \cap_{i=1}^{m} C_i$ of finitely many closed and convex sets C_i in the Euclidean space $\mathbf{R}^n$. Algorithmic schemes for this problem are, in general, either *sequential* or *simultaneous* or can also be *block-iterative* (see, e.g., Censor and Zenios [15, Section 1.3] for a classification of projection algorithms into such classes, and the review paper of Bauschke and Borwein [3] for a variety of specific algorithms of these kinds).

We now explain these terms in the framework of the algorithmic scheme proposed in this paper. For $t = 1, 2, \ldots, M$, let the *string* I_t be an ordered subset of $\{1, 2, \ldots, m\}$ of the form

$$I_t = (i_1^t, i_2^t, \ldots, i_{m(t)}^t), \qquad (1)$$

*Work supported by grants 293/97 and 592/00 of the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities, and by NIH grant HL-28438.

†Work supported by the Swedish Natural Science Research Council under Project M650-19981853/2000.

‡Work supported by NIH grant HL-28438.

with $m(t)$ the number of elements in I_t. We will assume that, for any t, the elements of I_t are distinct from each other; however, the extension of all that we say below to the case without this assumption is trivial (it only complicates the notation). Suppose that there is a set $S \subseteq \mathbf{R}^n$ such that there are operators $R_1, R_2, \ldots, R_m$ mapping S into S and an operator R which maps S^M into S.

Algorithmic Scheme

Initialization: $x^{(0)} \in S$ is arbitrary.

Iterative Step: given the current iterate $x^{(k)}$,
(i) calculate, for all $t = 1, 2, \ldots, M$,

$$T_t x^{(k)} = R_{i^t_{m(t)}} \ldots R_{i^t_2} R_{i^t_1} x^{(k)}, \tag{2}$$

(ii) and then calculate

$$x^{(k+1)} = R(T_1 x^{(k)}, T_2 x^{(k)}, \ldots, T_M x^{(k)}). \tag{3}$$

For every $t = 1, 2, \ldots, M$, this algorithmic scheme applies to $x^{(k)}$ successively the operators whose indices belong to the tth string. This can be done in parallel for all strings and then the operator R maps all end-points onto the next iterate $x^{(k+1)}$. This is indeed an algorithm provided that the operators $\{R_i\}_{i=1}^m$ and R all have algorithmic implementations. In this framework we get a *sequential* algorithm by the choice $M = 1$ and $I_1 = (1, 2, \ldots, m)$ and a *simultaneous* algorithm by the choice $M = m$ and $I_t = (t),\ t = 1, 2, \ldots, M$.

We demonstrate the underlying idea of our algorithmic scheme with the aid of Figure 1. For simplicity, we take the convex sets to be hyperplanes, denoted by H_1, H_2, H_3, H_4, H_5, and H_6, and assume all operators $\{R_i\}$ to be orthogonal projections onto the hyperplanes. The operator R is taken as a convex combination

$$R(x^1, x^2, \ldots, x^M) = \sum_{t=1}^{M} \omega_t x^t, \tag{4}$$

with $\omega_t > 0$, for all $t = 1, 2, \ldots, M$, and $\sum_{t=1}^{M} \omega_t = 1$.

Figure 1(a) depicts a purely sequential algorithm. This is the so-called POCS (Projections Onto Convex Sets) algorithm which coincides, for the case of hyperplanes, with the Kaczmarz algorithm, see, e.g., Algorithms 5.2.1 and 5.4.3, respectively, in [15] and Gubin, Polyak and Raik [26].

The fully simultaneous algorithm appears in Figure 1(b). With orthogonal reflections instead of orthogonal projections it was first proposed, by Cimmino [16], for solving linear equations. Here the current iterate $x^{(k)}$ is projected on all sets simultaneously and the next iterate $x^{(k+1)}$ is a convex combination of the projected points.

In Figure 1(c) we show how a simple averaging of *successive* projections (as opposed to averaging of parallel projections in Figure 1(b)) works. In this case $M = m$ and $I_t = (1, 2, \ldots, t)$, for $t = 1, 2, \ldots, M$. This scheme, appearing in Bauschke and Borwein [3], inspired our proposed Algorithmic Scheme whose action is demonstrated in Figure

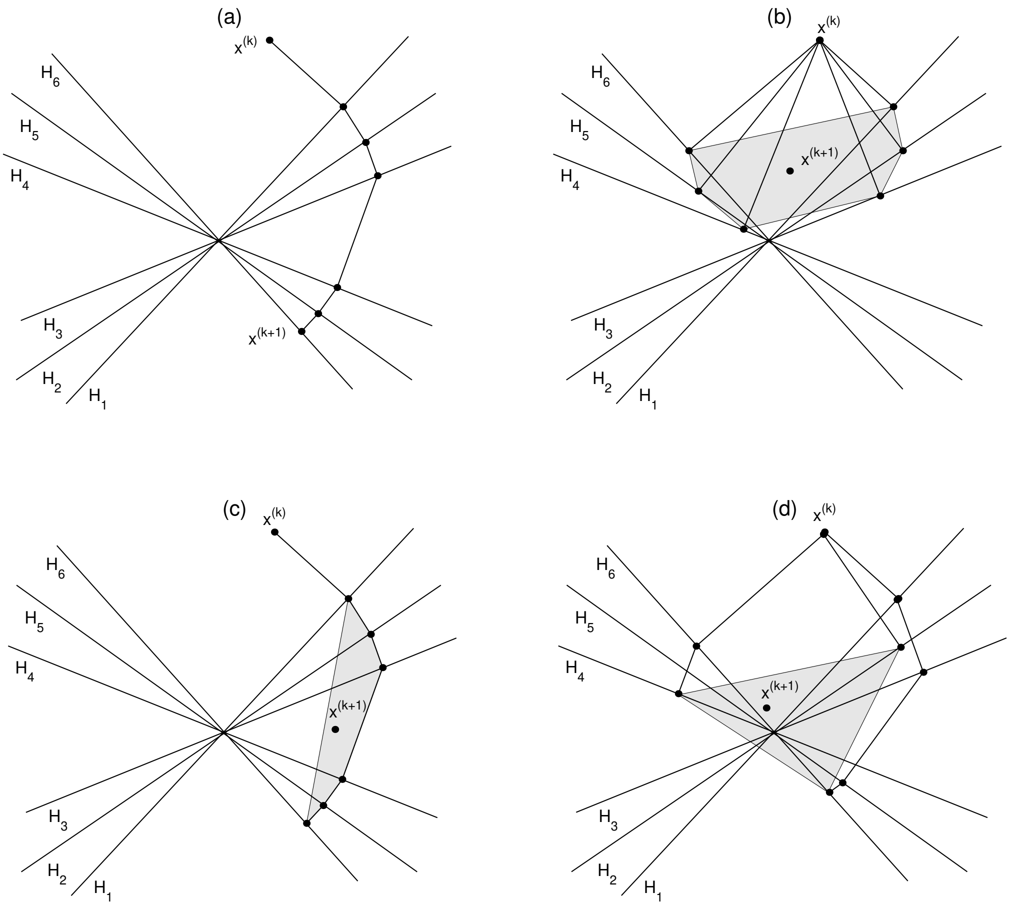

Figure 1. (a) Sequential projections. (b) Fully simultaneous projections. (c) Averaging sequential projections. (d) The new scheme: combining end-points of sequential strings.

1(d). It averages, via convex combinations, the end-points obtained from strings of sequential projections. This proposed scheme offers a variety of options for steering the iterates towards a solution of the convex feasibility problem. It is an inherently parallel scheme in that its mathematical formulation is parallel (like the fully simultaneous method mentioned above). We use this term to contrast such algorithms with others which are sequential in their mathematical formulation but can, sometimes, be implemented in a parallel fashion based on appropriate model decomposition (i.e., depending on the structure of the underlying problem). Being inherently parallel, our algorithmic scheme enables flexibility in the actual manner of implementation on a parallel machine.

We have been able to prove convergence of the Algorithmic Scheme for two special cases. In both cases it is assumed that (i) $C \cap \overline{S} \neq \emptyset$ (where $C = \cap_{i=1}^{m} C_i$ and $\overline{S}$ is the closure of S), (ii) every element of $\{1, 2, \ldots, m\}$ appears in at least one of the strings I_t, and (iii) all weights ω_t associated with the operator R are positive real numbers which sum up to one.

Case I. Each R_i is the Bregman projection onto C_i with respect to a Bregman function f with zone S and the operator R of (3) is a generalized convex combination, with weights ω_t, to be defined in Section 2.1.

Case II. $S = \mathbf{R}^n$ and, for $i = 1, 2, \ldots, m$, $R_i x = x + \theta_i(P_{C_i} x - x)$, with $0 < \theta_i < 2$, where P_{C_i} is the orthogonal projection onto C_i and R is defined by (4).

A generalization of this operator R was used by Censor and Elfving [12] and Censor and Reich [14] in fully simultaneous algorithms which employ Bregman projections. Our proof of convergence for Case I is based on adopting a product space formalism which is motivated by, but is somewhat different from, the product space formalism of Pierra [31]. For the proof of Case II we use results of Elsner, Koltracht and Neumann [25] and Censor and Reich [14].

The details and proofs of convergence are given in Section 2. In Section 3 we describe an application to optimization of a Bregman function over linear equalities. We conclude with a discussion, including some open problems in Section 4. The Appendix in Section 5 describes the role of Bregman projections in convex feasibility problems.

2. PROOFS OF CONVERGENCE

We consider the convex feasibility problem of finding $x^* \in C = \cap_{i=1}^{m} C_i$ where, $C_i \subseteq \mathbf{R}^n$, for all $i = 1, 2, \ldots, m$, are closed convex sets and $C \neq \emptyset$. The two Cases I and II, mentioned in the introduction, are presented in detail and their convergence is proven. For both cases we make the following assumptions.

Assumption 1. $C \cap \overline{S} \neq \emptyset$ where $\overline{S}$ is the closure of S, the domain of the algorithmic operators $R_1, R_2, \ldots, R_m$.

Assumption 2. Every element of $\{1, 2, \ldots, m\}$ appears in at least one of the strings I_t, constructed as in (1).

Assumption 3. The weights $\{\omega_t\}_{t=1}^{M}$ associated with the operator R are positive real numbers and $\sum_{t=1}^{M} \omega_t = 1$.

2.1. Case I: An Algorithm for Bregman Projections

Let $\mathcal{B}(S)$ denote the family of Bregman functions with zone $S \subseteq \mathbf{R}^n$ (see, e.g., Censor and Elfving [12], Censor and Reich [14], or Censor and Zenios [15] for definitions, basic properties and relevant references). For a discussion of the role of Bregman projections in algorithms for convex feasibility problems we refer the reader to the Appendix at the end of the paper.

In Case I we define, for $i = 1, 2, \ldots, m$, the algorithmic operator $R_i x$ to be the Bregman projection, denoted by $P^f_{C_i} x$, of x onto the set C_i with respect to a Bregman function f. Recall that the generalized distance $D_f : \overline{S} \times S \subseteq \mathbf{R}^{2n} \to \mathbf{R}$ is

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

where $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbf{R}^n$. The Bregman projection $P^f_Q x$ onto a closed convex set Q is then defined by

$$P^f_Q x = \arg\min\{D_f(y, x) \mid y \in Q \cap \overline{S}\}. \tag{6}$$

Such a projection exists and is unique, if $Q \cap \overline{S} \neq \emptyset$, see [15, Lemma 2.1.2].

Following Censor and Reich [14] let us call an x which satisfies, for $(x^1, x^2, \ldots, x^M) \in S^M$,

$$\nabla f(x) = \sum_{t=1}^{M} \omega_t \nabla f(x^t), \tag{7}$$

a *generalized convex combination* of $(x^1, x^2, \ldots, x^M)$ with respect to f. We further assume:

Assumption 4. For any $\boldsymbol{x} = (x^1, x^2, \ldots, x^M) \in S^M$ and any set of weights $\{\omega_t\}_{t=1}^M$, as in Assumption 3, there is a unique x in S which satisfies (7).

The operator R is defined by letting $R\boldsymbol{x}$ be the x whose existence and uniqueness is guaranteed by Assumption 4. The applicability of the algorithm depends (similarly to the applicability of its predecessors in [12] and [14]) on the ability to invert the gradient ∇f explicitly. If the Bregman function f is essentially smooth, then ∇f is a one-to-one mapping with continuous inverse $(\nabla f)^{-1}$, see, e.g., Rockafellar [33, Corollary 26.3.1].

We now prove convergence of the Algorithmic Scheme in Case I.

Theorem 2.1 *Let $f \in \mathcal{B}(S)$ be a Bregman function and let $C_i \subseteq \mathbf{R}^n$ be given closed convex sets, for $i = 1, 2, \ldots, m$, and define $C = \cap_{i=1}^m C_i$. If $P^f_{C_i} x \in S$ for any $x \in S$ and Assumptions 1–4 hold, then any sequence $\{x^{(k)}\}_{k \geq 0}$, generated by the Algorithmic Scheme for Case I, converges to a point $x^* \in C \cap \overline{S}$.*

Proof. Let $V = \mathbf{R}^n$ and consider the product space $\boldsymbol{V} = V^M = V \times V \times \cdots \times V$ in which, for any $\boldsymbol{x} \in \boldsymbol{V}$, $\boldsymbol{x} = (x^1, x^2, \ldots, x^M)$ with $x^t \in V$, for $t = 1, 2, \ldots, M$. The scalar product in $\boldsymbol{V}$ is denoted and defined by

$$\langle\langle \boldsymbol{x}, \boldsymbol{y} \rangle\rangle = \sum_{t=1}^{M} \langle x^t, y^t \rangle, \tag{8}$$

and we define in $\boldsymbol{V}$, for $j = 1, 2, \ldots, m$, the product sets

$$\boldsymbol{C}_j = \prod_{t=1}^{M} C_{j,t}, \tag{9}$$

with $C_{j,t}$ depending on the strings I_t as follows:

$$C_{j,t} = \begin{cases} C_{i_j^t}, & \text{if } j = 1, 2, \ldots, m(t), \\ V, & \text{if } j = m(t)+1, m(t)+2, \ldots, m. \end{cases} \tag{10}$$

Let

$$\boldsymbol{\Delta} = \{\boldsymbol{x} \mid \boldsymbol{x} = (x, x, \ldots, x),\ x \in V\}, \tag{11}$$

and

$$\delta : V \to \boldsymbol{\Delta}, \quad \delta(x) = (x, x, \ldots, x). \tag{12}$$

The set $\boldsymbol{\Delta}$ is called the *diagonal set* and the mapping δ is the *diagonal mapping*. In view of Assumption 2, the following equivalence between the convex feasibility problems in V and $\boldsymbol{V}$ is obvious:

$$x^* \in C \text{ if and only if } \delta(x^*) \in (\cap_{j=1}^m \boldsymbol{C}_j) \cap \boldsymbol{\Delta}. \tag{13}$$

The proof is based on examining Bregman's sequential projections algorithm (see Bregman [7, Theorem 1] or Censor and Zenios [15, Algorithm 5.8.1]) applied to the convex feasibility problem on the right-hand side of (13) in the product space $\boldsymbol{V}$. This is done as follows. With weights $\{\omega_t\}_{t=1}^M$, satisfying Assumption 3, we construct the function

$$\boldsymbol{F}(\boldsymbol{x}) = \sum_{t=1}^{M} \omega_t f(x^t). \tag{14}$$

By [12, Lemma 3.1], $\boldsymbol{F}$ is a Bregman function with zone $\boldsymbol{S}$ in the product space, i.e., $\boldsymbol{F} \in \mathcal{B}(\boldsymbol{S})$, where $\boldsymbol{S} = S^M$. Further, denoting by $\boldsymbol{P}_{\boldsymbol{Q}}^{\boldsymbol{F}} \boldsymbol{x}$ the Bregman projection of a point $\boldsymbol{x} \in \boldsymbol{V}$ onto a closed convex set $\boldsymbol{Q} = Q_1 \times Q_2 \times \cdots \times Q_M \subseteq \boldsymbol{V}$, with respect to $\boldsymbol{F}$, we can express it, by [12, Lemma 4.1], as

$$\boldsymbol{P}_{\boldsymbol{Q}}^{\boldsymbol{F}} \boldsymbol{x} = (P_{Q_1}^f x^1, P_{Q_2}^f x^2, \cdots, P_{Q_M}^f x^M). \tag{15}$$

From (2), (9), (10) and (15) we obtain

$$\boldsymbol{P}_{\boldsymbol{C}_m}^{\boldsymbol{F}} \cdots \boldsymbol{P}_{\boldsymbol{C}_2}^{\boldsymbol{F}} \boldsymbol{P}_{\boldsymbol{C}_1}^{\boldsymbol{F}} \boldsymbol{x} = (T_1 x^1, T_2 x^2, \cdots, T_M x^M). \tag{16}$$

Next we show that, for every $\boldsymbol{x} \in \boldsymbol{V}$,

$$\boldsymbol{P}_{\boldsymbol{\Delta}}^{\boldsymbol{F}} \boldsymbol{x} = \delta(x), \tag{17}$$

with $x = R(\boldsymbol{x})$. By (6), (11) and (12), the x which satisfies (17) is

$$x = \arg\min\{\boldsymbol{D}_{\boldsymbol{F}}(\delta(y), \boldsymbol{x}) \mid \delta(y) \in \overline{\boldsymbol{S}}\}, \tag{18}$$

where $\boldsymbol{D}_{\boldsymbol{F}}(\delta(y),\boldsymbol{x})$ is the Bregman distance in $\boldsymbol{V}$ with respect to $\boldsymbol{F}$. Noting that

$$\nabla \boldsymbol{F}(\boldsymbol{x}) = (\omega_1 \nabla f(x^1), \omega_2 \nabla f(x^2), \ldots, \omega_M \nabla f(x^M)), \tag{19}$$

we have, by (5), (8) and (14), that

$$\boldsymbol{D}_{\boldsymbol{F}}(\delta(y),\boldsymbol{x}) = \sum_{t=1}^{M} \omega_t (f(y) - f(x^t) - \langle \nabla f(x^t), y - x^t \rangle). \tag{20}$$

Since a Bregman distance is convex with respect to its first (vector) variable (see, e.g., [15, Chapter 2]), at the point x where (20) achieves its minimum, the gradient (with respect to y) must be zero. Thus, differentiating the right-hand side of (20), we get that this x must satisfy (7) and, therefore, by Assumption 4, it is in fact $R(\boldsymbol{x})$.

The convergence ([7, Theorem 1] or [15, Algorithm 5.8.1]) of Bregman's sequential algorithm guarantees, by taking $\boldsymbol{x}^{(0)} = \delta(x^{(0)})$ with $x^{(0)} \in S$ and, for $k \geq 0$, iterating

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{P}^{\boldsymbol{F}}_{\boldsymbol{\Delta}} \boldsymbol{P}^{\boldsymbol{F}}_{\boldsymbol{C}_m} \cdots \boldsymbol{P}^{\boldsymbol{F}}_{\boldsymbol{C}_2} \boldsymbol{P}^{\boldsymbol{F}}_{\boldsymbol{C}_1} \boldsymbol{x}^{(k)}, \tag{21}$$

that $\lim_{k\to 0} \boldsymbol{x}^{(k)} = \boldsymbol{x}^* \in (\cap_{j=1}^m \boldsymbol{C}_j) \cap \boldsymbol{\Delta}$. Observing (3), (16), and the fact that the x of (17) is $R(\boldsymbol{x})$, we get by induction that, for all $k \geq 0$, $\boldsymbol{x}^{(k)} = \delta(x^{(k)})$. By (13), this implies that $\lim_{k\to 0} x^{(k)} = x^* \in C$. ■

2.2. Case II: An Algorithm for Relaxed Orthogonal Projections

The framework and method of proof used in the previous subsection do not let us introduce relaxation parameters into the algorithm. However, drawing on findings of Elsner, Koltracht and Neumann [25] and of Censor and Reich [14] we do so for the special case of orthogonal projections.

In Case II we define, for $i = 1, 2, \ldots, m$, the algorithmic operators

$$R_i x = x + \theta_i (P_{C_i} x - x), \tag{22}$$

where $P_{C_i} x$ is the orthogonal projection of x onto the set C_i and θ_i are *periodic relaxation parameters*. By this we mean that the θ_i are fixed for each set C_i as in Eggermont, Herman and Lent [23, Theorem 1.2]. The algorithmic operator R is defined by (4) with weights ω_t as in Assumption 3. Equation (4) can be obtained from (7) by choosing the Bregman function $f(x) = ||x||_2^2$ with zone $S = \mathbf{R}^n$. In this case $P^f_{C_i} = P_{C_i}$ is the orthogonal projection and the Bregman distance is $D_f(y,x) = ||y-x||_2^2$, see, e.g., [15, Example 2.1.1]. The convergence theorem for the Algorithmic Scheme in Case II now follows.

Theorem 2.2 *If Assumptions 1–3 hold and if, for all* $i = 1, 2, \ldots, m$, *we have* $0 < \theta_i < 2$, *then any sequence* $\{x^{(k)}\}_{k\geq 0}$, *generated by the Algorithmic Scheme for Case II, converges to a point* $x^* \in C$.

Proof. By [25, Example 2] a relaxed projection operator of the form (22) is *strictly nonexpansive* with respect to the Euclidean norm, for any $0 < \theta_i < 2$. By this we mean that [25, Definition 2], for any pair $x, y \in \mathbf{R}^n$,

$$\text{either} \quad ||R_i x - R_i y||_2 < ||x - y||_2 \quad \text{or} \quad R_i x - R_i y = x - y. \tag{23}$$

Further, since every finite composition of strictly nonexpansive operators is a strictly non-expansive operator [25, p. 307], any finite composition of relaxed projections operators of the form (2) is strictly nonexpansive. Consequently, each such T_t is also a *paracontracting* operator in the sense of [25, Definition 1]), namely, $T_t : \mathbf{R}^n \to \mathbf{R}^n$ is continuous and for any *fixed point* $y \in \mathbf{R}^n$ of T_t, i.e., $T_t y = y$, and any $x \in \mathbf{R}^n$

$$||T_t x - y||_2 < ||x - y||_2 \quad \text{or} \quad T_t x = x. \tag{24}$$

From Censor and Reich [14, Section 4] we then conclude the convergence of any sequence generated by

$$x^{(k+1)} = \sum_{t=1}^{M} \omega_t T_t x^{(k)}, \tag{25}$$

to a common fixed point x^* of the family $\{T_t\}_{t=1}^{M}$, which in our case means convergence to a feasible point in C. This is so because, for each $t = 1, 2, \ldots, M,\ \ T_t$ is a product of the paracontractions R_i, given by (2.18), for all $i \in I_t$, and [25, Corollary 1] then implies that x^* is a fixed point of each R_i, thus of each P_{C_i}. The periodic relaxation and the fixed strings guarantee the finite number of paracontractions, thus enabling the use of the convergence results of [14]. ■

3. APPLICATION TO OPTIMIZATION OVER LINEAR EQUALITIES

In this application, we use the fact that the Algorithmic Scheme for Case I solves the convex feasibility problem to prove its nature as an optimization problem solver. Let f be a Bregman function with zone $S \subseteq \mathbf{R}^n$, let A be a matrix and let $d \in \mathcal{R}(A)$ be a vector in the range of A. Consider the following optimization problem

$$\min\{f(x) \mid x \in \overline{S},\ Ax = d\}. \tag{26}$$

We will show that the Algorithmic Scheme for Case I can be used to solve this problem.

Let the matrix A be of order $\overline{m} \times n$, with $\overline{m} = \sum_{i=1}^{m} \nu_i$, and partition it into m row-blocks of sizes ν_i as follows,

$$A^T = (A_1^T, A_2^T, \ldots, A_m^T), \tag{27}$$

where we denote vector and matrix transposition by T, and let

$$C_i = \{x \in \mathbf{R}^n \mid A_i x = d_i,\ d_i \in \mathbf{R}^{\nu_i}\}, \quad i = 1, 2, \ldots, m, \tag{28}$$

where $d^T = (d_1^T, d_2^T, \ldots, d_m^T)$. Partitioning a system of linear equations in this way has been shown to be useful in real-world problems, particularly for very large and sparse matrices, see, e.g., Eggermont, Herman and Lent [23].

We prove the following optimization result.

Theorem 3.1 *If $f \in \mathcal{B}(S)$ satisfies the assumptions in Theorem 2.1 and*

$$\nabla f(x^{(0)}) \in \mathcal{R}(A^T), \tag{29}$$

then the Algorithmic Scheme for Case I, applied to the sets C_i of (28), generates a sequence which converges to a solution x^* of (26).

Proof. Applying the Algorithmic Scheme for Case I to the convex feasibility problem with the sets (28), convergence towards a point $x^* \in C \cap \overline{S}$ is guaranteed by Theorem 2.1. Defining

$$Z = \{x \in S \mid \exists z \in \mathbf{R}^{\overline{m}} \text{ such that } \nabla f(x) = A^T z\}, \tag{30}$$

and

$$U = \{x \in \mathbf{R}^n \mid Ax = d,\ x \in \overline{S}\}, \tag{31}$$

we will use the result in [7, Lemma 3], which says that if $x^* \in U \cap \overline{Z}$ then x^* is a solution of (26). Therefore, we show now that $x^{(k)} \in Z$, for all $k \geq 0$, from which $x^* \in \overline{Z}$ follows.

For any $f \in \mathcal{B}(S)$ and $C = \{x \in \mathbf{R}^n \mid Ax = d\}$ such that $P_C^f x$ belongs to S, for any $x \in S$, it is the case that $\nabla f(P_C^f x) - \nabla f(x)$ is in the range of A^T. This follows from [12, Lemma 6.1] (which extends [15, Lemma 2.2.1]). Using this and the fact that, for all $j = 1, 2, \ldots, m(t)$,

$$\mathcal{R}(A^T_{i^t_j}) \subseteq \mathcal{R}(A^T), \tag{32}$$

we deduce that

$$\nabla f(T_t x^{(k)}) - \nabla f(x^{(k)}) \tag{33}$$

is in the range of A^T. Multiplying (33) by ω_t and summing over t we obtain, using (7) and $\sum_{t=1}^{M} \omega_t = 1$, that

$$\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) \in \mathcal{R}(A^T). \tag{34}$$

Using the initialization (29), we do induction on k with (34) and obtain that $x^{(k)} \in Z$, for all $k \geq 0$. ■

4. DISCUSSION AND SOME OPEN PROBLEMS

All algorithms and results presented here apply, in particular to orthogonal unrelaxed projections, because those are a special case of Bregman projections (see the comments made before Theorem 2.2) as well as of the operators in (22). Thus our Algorithmic Scheme generalizes the method described by Bauschke and Borwein [3, Examples 2.14 and 2.20] where they define an operator $T = \frac{1}{m}(P_1 + P_2 P_1 + \cdots + P_m \cdots P_2 P_1)$, with P_i orthogonal projections onto given sets, for $i = 1, 2, \ldots, m$, and show weak convergence in Hilbert space of $\{T^k x^{(0)}\}_{k \geq 0}$ to some fixed point of T, for every $x^{(0)}$.

Earlier work concerning the convergence of (random) products of averaged mappings is due to Reich and coworkers; see, e.g., Dye and Reich [21], Dye and Reich [20, Theorem 5] and Dye et al. [22, Theorem 5]. In the infinite-dimensional case they require some conditions on the fixed point sets of the mappings which are not needed in the finite-dimensional case. The above-mentioned method of Bauschke and Borwein can also be understood by using the results of Baillon, Bruck and Reich [6, Theorems 1.2 and 2.1], Bruck and Reich [9, Corollary 1.3], and Reich [32, Proposition 2.4]. A more recent study is Bauschke [2].

At the extremes of the "spectrum of algorithms," derivable from our Algorithmic Scheme, are the generically sequential method, which uses one set at a time, and the fully simultaneous algorithm, which employs all sets at each iteration. The "block-iterative projections" (BIP) scheme of Aharoni and Censor [1] (see also Butnariu and Censor [10], Bauschke and Borwein [3], Bauschke, Borwein and Lewis [5] and Elfving [24]) also has the sequential and the fully simultaneous methods as its extremes in terms of block structures. The question whether there are any other relationships between the BIP scheme of [1] and the Algorithmic Scheme of this paper is of theoretical interest. However, the current lack of an answer to it does not diminish the value of the proposed Algorithmic Scheme, because its new algorithmic structure gives users a tool to design algorithms that will average sequential strings of projections.

We have not as yet investigated the behavior of the Algorithmic Scheme, or special instances of it, in the inconsistent case when the intersection $C = \cap_{i=1}^{m} C_i$ is empty. For results on the behavior of the fully simultaneous algorithm with orthogonal projections in the inconsistent case see, e.g., Combettes [18] or Iusem and De Pierro [27]. Another way to treat possible inconsistencies is to reformulate the constraints as $c \leq Ax \leq d$ or $||Ax - d||_2 \leq \epsilon$, see e.g. [15]. Also, variable iteration-dependent relaxation parameters and variable iteration-dependent string constructions could be interesting future extensions.

The practical performance of specific algorithms derived from the Algorithmic Scheme needs still to be evaluated in applications and on parallel machines.

5. APPENDIX: THE ROLE OF BREGMAN PROJECTIONS

Bregman generalized distances and generalized projections are instrumental in several areas of mathematical optimization theory. Their introduction by Bregman [7] was initially followed by the works of Censor and Lent [13] and De Pierro and Iusem [19] and, subsequently, lead to their use in special-purpose minimization methods, in the proximal point minimization method, and for stochastic feasibility problems. These generalized distances and projections were also defined in non-Hilbertian Banach spaces, where, in the absence of orthogonal projections, they can lead to simpler formulas for projections.

In the Euclidean space, where our present results are formulated, Bregman's method for minimizing a convex function (with certain properties) subject to linear inequality constraints employs Bregman projections onto the half-spaces represented by the constraints, see, e.g., [13,19]. Recently the extension of this minimization method to nonlinear convex constraints has been identified with the Han-Dykstra projection algorithm for finding the projection of a point onto an intersection of closed convex sets, see Bregman, Censor and Reich [8].

It looks as if there might be no point in using non-orthogonal projections for solving the convex feasibility problem in $\mathbf{R}^n$ since they are generally not easier to compute. But this is not always the case. In [29,30] Shamir and co-workers have used the multiprojection method of Censor and Elfving [12] to solve filter design problems in image restoration and image recovery posed as convex feasibility problems. They took advantage of that algorithm's flexibility to employ Bregman projections with respect to *different* Bregman functions within the same algorithmic run.

Another example is the seminal paper by Csiszár and Tusnády [17], where the central procedure uses alternating entropy projections onto convex sets. In their "alternating minimization procedure," they alternate between minimizing over the first and second argument of the Bregman distance (Kullback-Leibler divergence, in fact). These divergences are nothing but the generalized Bregman distances obtained by using the negative of Shannon's entropy as the underlying Bregman function.

Recent studies about Bregman projections (Kiwiel [28]), Bregman/Legendre projections (Bauschke and Borwein [4]), and averaged entropic projections (Butnariu, Censor and Reich [11]) – and their uses for convex feasibility problems in $\mathbf{R}^n$ discussed therein – attest to the continued (theoretical and practical) interest in employing Bregman projections in projection methods for convex feasibility problems. This is why we formulated and studied Case I of our Algorithmic Scheme within the framework of such projections.

Acknowledgements. We are grateful to Charles Byrne for pointing out an error in an earlier draft and to the anonymous referees for their constructive comments which helped to improve the paper. We thank Fredrik Berntsson for help with drawing the figures. Part of the work was done during visits of Yair Censor at the Department of Mathematics of the University of Linköping. The support and hospitality of Professor Åke Björck, head of the Numerical Analysis Group there, are gratefully acknowledged.

REFERENCES

1. R. Aharoni and Y. Censor, Block-iterative methods for parallel computation of solutions to convex feasibility problems, *Linear Algebra and Its Applications* **120** (1989) 165–175.
2. H.H. Bauschke, A norm convergence result on random products of relaxed projections in Hilbert space, *Transactions of the American Mathematical Society* **347** (1995) 1365–1373.
3. H.H. Bauschke and J.M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* **38** (1996) 367–426.
4. H.H. Bauschke and J.M. Borwein, Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis* **4** (1997) 27–67.
5. H.H. Bauschke, J.M. Borwein and A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space *Contemporary Mathematics* **204** (1997) 1–38.
6. J.M. Baillon, R.E. Bruck and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston Journal of Mathematics* **4** (1978) 1–9.
7. 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 Com-*

putational Mathematics and Mathematical Physics **7** (1967) 200–217.
8. L.M. Bregman, Y. Censor and S. Reich, Dykstra's algorithm as the nonlinear extension of Bregman's optimization method, *Journal of Convex Analysis* **6** (1999) 319–334.
9. R.E. Bruck and S. Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston Journal of Mathematics* **3** (1977) 459–470.
10. D. Butnariu and Y. Censor, Strong convergence of almost simultaneous block-iterative projection methods in Hilbert spaces, *Journal of Computational and Applied Mathematics* **53** (1994) 33–42.
11. D. Butnariu, Y. Censor and S. Reich, Iterative averaging of entropic projection for solving stochastic convex feasibility problems, *Computational Optimization and Applications* **8** (1997) 21–39.
12. Y. Censor and T. Elfving, A multiprojection algorithm using Bregman projections in a product space, *Numerical Algorithms* **8** (1994) 221–239.
13. Y. Censor and A. Lent, An iterative row-action method for interval convex programming, *Journal of Optimization Theory and Applications* **34** (1981) 321–353.
14. Y. Censor and S. Reich, Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization, *Optimization* **37** (1996) 323–339.
15. Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications,* (Oxford University Press, New York, 1997).
16. G. Cimmino, Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari, *La Ricerca Scientifica* **XVI** Series II, Anno IX, 1 (1938), 326–333.
17. I. Csiszár and G. Tusnády, Information geometry and alternating minimization procedures, *Statistics and Decisions* Supplement Issue No. 1 (1984) 205–237.
18. P.L. Combettes, Inconsistent signal feasibility problems: least-squares solutions in a product space, *IEEE Transactions on Signal Processing* **SP-42** (1994), 2955–2966.
19. 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** (1986) 421–440.
20. J.M. Dye and S. Reich, Unrestricted iteration of projections in Hilbert space, *Journal of Mathematical Analysis and Applications* **156** (1991) 101–119.
21. J.M. Dye and S. Reich, On the unrestricted iterations of nonexpansive mappings in Hilbert space, *Nonlinear Analysis, Theory, Methods and Applications* **18** (1992) 199–207.
22. J.M. Dye, T. Kuczumow, P.-K. Lin and S. Reich, Convergence of unrestricted products of nonexpansive mappings in spaces with the Opial property, *Nonlinear Analysis, Theory, Methods and Applications* **26** (1996) 767–773.
23. 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** (1981) 37–67.
24. T. Elfving, Block-iterative methods for consistent and inconsistent linear equations, *Numerische Mathematik* **35** (1980) 1–12.
25. L. Elsner, I. Koltracht and M. Neumann, Convergence of sequential and asynchronous nonlinear paracontractions, *Numerische Mathematik* **62** (1992) 305–319.
26. L.G. Gubin, B.T. Polyak and E.V. Raik, The method of projections for finding the common point of convex sets, *USSR Computational Mathematics and Mathematical Physics* **7** (1967) 1–24.

27. A.N. Iusem and A.R. De Pierro, Convergence results for an accelerated nonlinear Cimmino algorithm, *Numerische Mathematik* **49** (1986) 367–378.
28. K.C. Kiwiel, Generalized Bregman projections in convex feasibility problems, *Journal of Optimization Theory and Applications* **96** (1998) 139–157.
29. T. Kotzer, N. Cohen and J. Shamir, A projection-based algorithm for consistent and inconsistent constraints, *SIAM Journal on Optimization* **7** (1997) 527–546.
30. D. Lyszyk and J. Shamir, Signal processing under uncertain conditions by parallel projections onto fuzzy sets, *Journal of the Optical Society of America A* **16** (1999) 1602–1611.
31. G. Pierra, Decomposition through formalization in a product space, *Mathematical Programming* **28** (1984) 96–115.
32. S. Reich, A limit theorem for projections, *Linear and Multilinear Algebra* **13** (1983) 281–290.
33. R.T. Rockafellar, *Convex Analysis,* (Princeton University Press, Princeton, New Jersey, 1970).