SIAM J. MATRIX ANAL. APPL.
Vol. 24, No. 1, pp. 40–58

BLOCK-ITERATIVE ALGORITHMS WITH DIAGONALLY SCALED OBLIQUE PROJECTIONS FOR THE LINEAR FEASIBILITY PROBLEM*

YAIR CENSOR† AND TOMMY ELFVING‡

Abstract. We formulate a block-iterative algorithmic scheme for the solution of systems of linear inequalities and/or equations and analyze its convergence. This study provides as special cases proofs of convergence of (i) the recently proposed component averaging (CAV) method of Censor, Gordon, and Gordon [*Parallel Comput.*, 27 (2001), pp. 777–808], (ii) the recently proposed block-iterative CAV (BICAV) method of the same authors [*IEEE Trans. Medical Imaging*, 20 (2001), pp. 1050–1060], and (iii) the simultaneous algebraic reconstruction technique (SART) of Andersen and Kak [*Ultrasonic Imaging*, 6 (1984), pp. 81–94] and generalizes them to linear inequalities. The first two algorithms are projection algorithms which use certain generalized oblique projections and diagonal weighting matrices which reflect the sparsity of the underlying matrix of the linear system. The previously reported experimental acceleration of the initial behavior of CAV and BICAV is thus complemented here by a mathematical study of the convergence of the algorithms.

Key words. block-iterative algorithms, component averaging (CAV), block-iterative CAV, simultaneous algebraic reconstruction technique, oblique projections, linear feasibility problem

AMS subject classifications. 90C25, 90C30

PII. S089547980138705X

1. Introduction.

Recently Censor, Gordon, and Gordon proposed and studied new iterative schemes for linear equations: In [7] the CAV (component averaging) method was presented as a simultaneous projection algorithm and in [8] BICAV was proposed as a block-iterative companion to CAV. In these methods the sparsity of the matrix is explicitly used when constructing the iteration formula. Using this new scaling we observed considerable improvement compared to traditionally scaled iteration methods. In [7] a proof of convergence was given for unity relaxation only, whereas no proofs at all were given for the block-iterative case [8].

The purpose of this paper is to describe a generalization to linear inequalities (with linear equations as a special case) of the Censor, Gordon, and Gordon schemes and study its convergence. It is shown that for the *consistent* case the block-iterative scheme (of which the fully simultaneous method is a special case) converges. For the inconsistent case we consider only linear equations and show that the simultaneous scheme converges to a weighted least squares solution. The treatment of the consistent case is based on our paper [6], in which an accelerated version of the fully simultaneous method with orthogonal projections for linear inequalities was proposed and studied.

*Received by the editors April 4, 2001; accepted for publication (in revised form) by M. Hanke October 16, 2001; published electronically May 15, 2002. This research was supported by research grants 293/97 and 592/00 from the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

http://www.siam.org/journals/simax/24-1/38705.html

†Department of Mathematics, University of Haifa, Mt. Carmel, Haifa 31905, Israel (yair@math.haifa.ac.il). The research of this author was supported by NIH grant HL-28438 at the Medical Image Processing Group (MIPG), Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA.

‡Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden (toelf@mai.liu.se). The research of this author was supported by the Swedish Natural Science Research Council under project M650-19981853/2000.

Recent relevant work of Byrne [5] and Jiang and Wang [19] is referred to at the end of Examples 7.1 and 7.2, respectively.

2. The CAV algorithm: Motivation and review. To motivate this work, let us consider linear equations and denote the hyperplanes

$$H_i := \left\{x \in \mathbb{R}^n \mid \langle a^i, x\rangle = b_i\right\} \tag{2.1}$$

for $i = 1, 2, \ldots, m$, where $\langle\cdot,\cdot\rangle$ is the inner product and $a^i = (a^i_j)^n_{j=1} \in \mathbb{R}^n$, $a^i \neq 0$, and $b_i \in \mathbb{R}$ are given vectors and given real numbers, respectively. Then the *orthogonal (nearest Euclidean distance) projection* $P_i(z)$ of any $z \in \mathbb{R}^n$ onto H_i is

$$P_i(z) \;=\; z + \frac{b_i - \langle a^i, z\rangle}{\|a^i\|_2^2} a^i\,, \tag{2.2}$$

where $\|\cdot\|_2$ is the Euclidean norm.

In Cimmino's simultaneous projections method [11] (see also, e.g., Censor and Zenios [9, Algorithm 5.6.1] with relaxation parameters and with equal weights $w_i = 1/m$), the next iterate x^{k+1} is the average of the projections of x^k on the hyperplanes H_i, as follows.

Algorithm 2.1 (Cimmino).
Initialization: $x^0 \in \mathbb{R}^n$ *is arbitrary.*
Iterative Step: *Given* x^k*, compute*

$$x^{k+1} \;=\; x^k + \frac{\lambda_k}{m}\sum_{i=1}^{m}\left(P_i(x^k) - x^k\right), \tag{2.3}$$

where $\{\lambda_k\}_{k\geq 0}$ *are relaxation parameters.*

Expanding the iterative step (2.3) according to (2.2) produces, for every component $j = 1, 2, \ldots, n$,

$$x_j^{k+1} \;=\; x_j^k + \frac{\lambda_k}{m}\sum_{i=1}^{m}\frac{b_i - \langle a^i, x^k\rangle}{\|a^i\|_2^2} a^i_j\,. \tag{2.4}$$

When the $m \times n$ system matrix $A = (a^i_j)$ is sparse, only a relatively small number of the elements $\{a^1_j, a^2_j, \ldots, a^m_j\}$ in the jth column of A are nonzero, but in (2.4) the sum of their contributions is divided by the relatively large m. This observation led Censor, Gordon, and Gordon [7] to consider replacement of the factor $1/m$ in (2.4) by a factor that depends only on the *nonzero* elements in the set $\{a^1_j, a^2_j, \ldots, a^m_j\}$. For each $j = 1, 2, \ldots, n$, denote by s_j the number of nonzero elements of column j of the matrix A, and replace (2.4) by

$$x_j^{k+1} \;=\; x_j^k + \frac{\lambda_k}{s_j}\sum_{i=1}^{m}\frac{b_i - \langle a^i, x^k\rangle}{\|a^i\|_2^2} a^i_j\,. \tag{2.5}$$

Certainly, if A is sparse, then the s_j values will be much smaller than m. But this posed a theoretical difficulty. The iterative step (2.4) is a special case of

$$x^{k+1} \;=\; x^k + \lambda_k\sum_{i=1}^{m} w_i\frac{b_i - \langle a^i, x^k\rangle}{\|a^i\|_2^2} a^i\,, \tag{2.6}$$

where the fixed weights $\{w_i\}_{i=1}^m$ must be positive for all i and $\sum_{i=1}^m w_i = 1$. The attempt to use $1/s_j$ as weights in (2.5) does not fit into the scheme (2.6), unless one can prove convergence of the iterates of a fully simultaneous iterative scheme with component-dependent (i.e., j-dependent) weights of the form

$$x_j^{k+1} = x_j^k + \lambda_k \sum_{i=1}^m w_{ij} \frac{b_i - \langle a^i, x^k\rangle}{\|a^i\|_2^2} a_j^i \tag{2.7}$$

for all $j = 1, 2, \ldots, n$.

To derive a proof of convergence for (2.7), Censor, Gordon, and Gordon modified it further by replacing the orthogonal projections onto the hyperplanes H_i by certain *oblique projections* induced by appropriately defined weight matrices, as will be explained next. Consider a hyperplane $H := \{x \in \mathbb{R}^n \mid \langle a, x\rangle = b\}$, with $a = (a_j) \in \mathbb{R}^n$, $b \in \mathbb{R}$, and $a \neq 0$. Let G be an $n \times n$ symmetric positive definite matrix and let $\|x\|_G^2 := \langle x, Gx\rangle$ be the associated *ellipsoidal norm*; see, e.g., Bertsekas and Tsitsiklis [4, Proposition A.28]. Given a point $z \in \mathbb{R}^n$, the *oblique projection of z onto H with respect to G* is the unique point $P_H^G(z) \in H$ for which

$$P_H^G(z) = \arg\min \{\|x - z\|_G \mid x \in H\}. \tag{2.8}$$

Solving this minimization problem leads to

$$P_H^G(z) = z + \frac{b - \langle a, z\rangle}{\|a\|_{G^{-1}}^2} G^{-1} a, \tag{2.9}$$

where G^{-1} is the inverse of G. For $G = I$, the identity matrix, (2.9) yields the orthogonal projection of z onto H, as given by (2.2); see, e.g., Ben-Israel and Greville [3, section 2.6].

In order to consider oblique projections onto H with respect to a diagonal matrix $G = \mathrm{diag}(g_1, g_2, \ldots, g_n)$ for which some diagonal elements might be zero, the following definition is used.

DEFINITION 2.1 (see [7]). *Let $G = \mathrm{diag}(g_1, g_2, \ldots, g_n)$ with $g_j \geq 0$ for all $j = 1, 2, \ldots, n$, let $H = \{x \in \mathbb{R}^n \mid \langle a, x\rangle = b\}$ be a hyperplane with $a = (a_j) \in \mathbb{R}^n$ and $b \in \mathbb{R}$, and assume that $g_j = 0$ if and only if $a_j = 0$. The* generalized oblique projection *of a point $z \in \mathbb{R}^n$ onto H with respect to G is defined, for all $j = 1, 2, \ldots, n$, by*

$$(P_H^G(z))_j := \begin{cases} z_j + \dfrac{b - \langle a, z\rangle}{\sum_{\substack{l=1\\ g_l \neq 0}}^n \frac{a_l^2}{g_l}} \cdot \dfrac{a_j}{g_j} & \text{if } g_j \neq 0, \\ z_j & \text{if } g_j = 0. \end{cases} \tag{2.10}$$

It is not difficult to verify that this $P_H^G(z)$ belongs to H, that it solves (2.8) if $\|x - z\|_G$ is replaced there by $\langle x - z, G(x - z)\rangle$, and that it is uniquely defined, although other solutions of (2.8) may exist due to the possibly zero-valued g_j's. This $P_H^G(z)$ reduces to (2.9) if $g_j \neq 0$ for all $j = 1, 2, \ldots, n$.

Consider next a set $\{G_i\}_{i=1}^m$ of real diagonal $n \times n$ matrices $G_i = \mathrm{diag}(g_{i1}, g_{i2}, \ldots, g_{in})$ with $g_{ij} \geq 0$ for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ and such that $\sum_{i=1}^m G_i = I$. Referring to the sparsity pattern of A, one needs the following definition.

DEFINITION 2.2 (see [7]). *A family $\{G_i\}_{i=1}^m$ of real diagonal $n \times n$ matrices with all diagonal elements $g_{ij} \geq 0$ and such that $\sum_{i=1}^m G_i = I$ is called* sparsity pattern oriented (SPO) *with respect to an* $m\times n$ *matrix* A *if, for every* $i=1,2,\ldots,m,\ g_{ij}=0$ *if and only if* $a^i_j=0$.

The CAV algorithm of [7] combined three features: (i) Each orthogonal projection onto H_i in (2.3) was replaced by a generalized oblique projection with respect to G_i. (ii) The scalar weights $\{w_i\}$ in (2.6) were replaced by the diagonal weighting matrices $\{G_i\}$. (iii) The actual weights were set inversely proportional to the number of nonzero elements in each column, as motivated by the discussion preceding (2.5). The iterative step resulting from the first two features has the form

$$x^{k+1} \;=\; x^k+\lambda_k\sum_{i=1}^{m}G_i\left(P^{G_i}_{H_i}(x^k)-x^k\right), \tag{2.11}$$

or, equivalently, substituting from (2.10) for each $P^{G_i}_{H_i}$, one gets the following.

Algorithm 2.2 (diagonal weighting (DWE); see [7]).
Initialization: $x^0\in\mathbb{R}^n$ *is arbitrary.*
Iterative Step: *Given* x^k, *compute* x^{k+1} *by using, for* $j=1,2,\ldots,n$, *the formula*

$$x^{k+1}_j \;=\; x^k_j+\lambda_k\sum_{\substack{i=1\\ g_{ij}\neq 0}}^{m}\frac{b_i-\langle a^i,x^k\rangle}{\sum_{\substack{l=1\\ g_{il}\neq 0}}^{n}\frac{(a^i_l)^2}{g_{il}}}\cdot a^i_j\,, \tag{2.12}$$

where $\{G_i\}_{i=1}^m$ *is a given family of diagonal SPO (with respect to* A*) weighting matrices as in Definition* 2.2, *and* $\{\lambda_k\}_{k\geq 0}$ *are relaxation parameters.*

Finally, the diagonal matrices $\{G_i\}_{i=1}^m$ are constructed in order to achieve the acceleration discussed above. Define

$$g_{ij}:=\begin{cases}\dfrac{1}{s_j} & \text{if } a^i_j\neq 0\,,\\[2ex] 0 & \text{if } a^i_j=0\,.\end{cases} \tag{2.13}$$

With this particular SPO family of G_i's one obtains the CAV algorithm.

Algorithm 2.3 (component averaging (CAV); see [7]).
Initialization: $x^0\in\mathbb{R}^n$ *is arbitrary.*
Iterative Step: *Given* x^k, *compute* x^{k+1} *by using, for* $j=1,2,\ldots,n$, *the formula*

$$x^{k+1}_j \;=\; x^k_j+\lambda_k\sum_{i=1}^{m}\frac{b_i-\langle a^i,x^k\rangle}{\sum_{l=1}^{n}s_l\left(a^i_l\right)^2}\cdot a^i_j\,, \tag{2.14}$$

where $\{\lambda_k\}_{k\geq 0}$ *are relaxation parameters and* $\{s_l\}_{l=1}^n$ *are as defined above.*

It was shown in [7] that Algorithm 2.2, with $\lambda_k=1$ for all $k\geq 0$, generates sequences $\{x^k\}_{k\geq 0}$ which always converge regardless of the initial point x^0 and independently from the consistency or inconsistency of the underlying system $Ax=b$. Moreover, it always converges to a minimizer of a certain proximity function.

3. The block-iterative component averaging algorithm (BICAV).

The basic idea of the block-iterative CAV (BICAV) algorithm is to break up the system $Ax=b$ into "blocks" of equations and treat each block according to the CAV methodology, passing cyclically over all the blocks. Throughout the following, T will be the number of blocks and, for $t=1,2,\ldots,T$, let the block of indices $B_t\subseteq\{1,2,\ldots,m\}$ be an ordered subset of the form $B_t=\{i^t_1,i^t_2,\ldots,i^t_{m(t)}\}$, where $m(t)$ is the number of elements in B_t. There is nothing preventing different blocks from containing common indices; we require, however, the following.

ASSUMPTION 3.1. *Every element of* $\{1, 2, \ldots, m\}$ *appears in at least one of the sets* B_t, $t = 1, 2, \ldots, T$.

For $t = 1, 2, \ldots, T$, let A_t denote the matrix formed by taking all the rows $\{a^i\}$ of A whose indices belong to the block of indices B_t, i.e.,

$$A_t := \begin{bmatrix} a^{i_1^t} \\ a^{i_2^t} \\ \vdots \\ a^{i_{m(t)}^t} \end{bmatrix}, \quad t = 1, 2, \ldots, T. \tag{3.1}$$

The iterative step of the BICAV algorithm, developed and experimentally tested by Censor, Gordon, and Gordon in [8], uses, for every block index $t = 1, 2, \ldots, T$, generalized oblique projections with respect to a family $\{G_i^t\}_{i \in B_t}$ of diagonal matrices which are SPO with respect to A_t. The same family is also used to perform the diagonal weighting. The resulting iterative step has the form

$$x^{k+1} \; = \; x^k + \lambda_k \sum_{i \in B_{t(k)}} G_i^{t(k)} \left(P_{H_i}^{G_i^{t(k)}}(x^k) - x^k \right), \tag{3.2}$$

where $\{t(k)\}_{k \geq 0}$ is a *control sequence* according to which the $t(k)$th block is chosen by the algorithm to be acted upon at the kth iteration, and thus, $1 \leq t(k) \leq T$ for all $k \geq 0$. The real numbers $\{\lambda_k\}_{k \geq 0}$ are user-chosen *relaxation parameters*. Substituting from (2.10) for each $P_{H_i}^{G_i^{t(k)}}$, one obtains the following.

ALGORITHM 3.1 (block-iterative diagonal weighting (BIDWE); see [8]).
Initialization: $x^0 \in \mathbb{R}^n$ *is arbitrary.*
Iterative Step: *Given* x^k, *compute* x^{k+1} *by using, for* $j = 1, 2, \ldots, n$, *the formula*

$$x_j^{k+1} \; = \; x_j^k + \lambda_k \sum_{\substack{i \in B_{t(k)} \\ g_{ij}^{t(k)} \neq 0}} \frac{b_i - \langle a^i, x^k \rangle}{\sum_{\substack{l=1 \\ g_{il}^{t(k)} \neq 0}}^{n} \frac{(a_l^i)^2}{g_{il}^{t(k)}}} \cdot a_j^i \,, \tag{3.3}$$

where, for each $t = 1, 2, \ldots, T$, $\{G_i^t\}_{i \in B_t}$ *is a given family of diagonal SPO (with respect to* A_t*) weighting matrices, as in Definition* 2.2, *the control sequence is cyclic, i.e.,* $t(k) = k \bmod T + 1$ *for all* $k \geq 0$, $\{\lambda_k\}_{k \geq 0}$ *are relaxation parameters, and* $G_i^t = \mathrm{diag}(g_{i1}^t, g_{i2}^t, \ldots, g_{in}^t)$.

Finally, in order to achieve the acceleration, the diagonal matrices $\{G_i^t\}_{i \in B_t}$ are constructed as in the original CAV algorithm [7], but with respect to each A_t. Let s_j^t be the number of nonzero elements $a_j^i \neq 0$ in the jth column of A_t and define

$$g_{ij}^t := \begin{cases} \dfrac{1}{s_j^t} & \text{if } a_j^i \neq 0, \\[2ex] 0 & \text{if } a_j^i = 0. \end{cases} \tag{3.4}$$

It is easy to verify that, for each $t = 1, 2, \ldots, T$, $\sum_{i \in B_t} G_i^t = I$ holds for these matrices. With these particular SPO families of G_i^t's one obtains the block-iterative algorithm.

ALGORITHM 3.2 (block-iterative component averaging (BICAV); see [8]).
Initialization: $x^0 \in \mathbb{R}^n$ *is arbitrary.*
Iterative Step: *Given* x^k, *compute* x^{k+1} *by using, for* $j = 1, 2, \ldots, n$, *the formula*

$$x_j^{k+1} \;=\; x_j^k + \lambda_k \sum_{i \in B_{t(k)}} \frac{b_i - \langle a^i, x^k \rangle}{\sum_{l=1}^n s_l^{t(k)} (a_l^i)^2} \cdot a_j^i \,, \tag{3.5}$$

where $\{\lambda_k\}_{k \geq 0}$ *are relaxation parameters,* $\{s_l^t\}_{l=1}^n$ *are as defined above, and the control sequence is cyclic, i.e.,* $t(k) = k \bmod T + 1$ *for all* $k \geq 0$.

For the case $T = 1$ and $B_1 = \{1, 2, \ldots, m\}$, Algorithm 3.2 becomes fully simultaneous, i.e., it is the CAV algorithm of [7]. For $T = m$ and $B_t = \{t\},\ t = 1, 2, \ldots, m$, BICAV simply becomes the well-known ART (algebraic reconstruction technique) (see, e.g., Herman [17]), also known as Kaczmarz's algorithm [20] (see also, e.g., [9, Algorithm 5.4.3]).

4. The algorithmic schemes that cover the CAV and BICAV algorithms.

We consider the system of linear inequalities

$$Ax \leq b, \tag{4.1}$$

where A is a real $m \times n$ matrix. We partition A into row blocks, precisely as done at the beginning of section 3. The right-hand-side vector b is partitioned similarly with b^t denoting those elements of b whose indices belong to the block of indices B_t,

$$b^t := \begin{bmatrix} b_{i_1^t} \\ b_{i_2^t} \\ \vdots \\ b_{i_{m(t)}^t} \end{bmatrix}, \quad t = 1, 2, \ldots, T. \tag{4.2}$$

The classical partitioning with fixed nonoverlapping blocks of equal sizes results by taking $m(t) = l,\ t = 1, 2, \ldots, T$, with $l \times T = m$. For each $i = 1, 2, \ldots, m$, the closed half-space

$$L_i := \{\ x \in \mathbb{R}^n \mid \langle a^i, x \rangle \leq b_i\ \} \tag{4.3}$$

has (2.1) as its bounding hyperplane. Define $L := \cap_{i=1}^m L_i$ and note that L is a closed convex set in $\mathbb{R}^n$. The task of finding a member of L, i.e., a solution of (4.1), is called the *linear feasibility problem*, which is a special case of the *convex feasibility problem*; see, e.g., Bauschke and Borwein [2] or [9, Chapter 5].

It is well known and easy to verify that the orthogonal projection $P_{L_i}(z)$ of a point $z \in \mathbb{R}^n$ onto L_i is

$$P_{L_i}(z) = z + c_i(z) a^i, \text{ where } \ c_i(z) = \min \left\{ 0, \frac{b_i - \langle a^i, z \rangle}{||a^i||_2^2} \right\}. \tag{4.4}$$

Note that if $z \notin L_i$, then $c_i(z) < 0$; otherwise $c_i(z) = 0$. Further define

$$I_t(z) := \{\ i \mid i_1^t \leq i \leq i_{m(t)}^t \text{ and } c_i(z) < 0\ \} \tag{4.5}$$

as the set of indices of the half-spaces in the tth block which are violated by z. We also introduce diagonal matrices $\{D_t\}_{t=1}^T$, corresponding to the blocks $\{A_t\}_{i=1}^T$,

$$(D_t(z))_{jj} = \begin{cases} 1 & \text{if } j \in I_t(z), \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Let $\{M_t\}_{t=1}^T$ be some given positive definite and symmetric matrices with nonnegative elements. Define

$$M_t(z) = D_t(z) M_t D_t(z), \quad t = 1, 2, \ldots, T. \tag{4.7}$$

If $\{x^k\}_{k\geq 0}$ is a sequence of vectors, then we use the following abbreviations: $c_i(x^k) \equiv c_i^k$, $I_t(x^k) \equiv I_t^k$, $D_t(x^k) \equiv D_t^k$, and $M_t(x^k) \equiv M_t^k$. We propose now the block-iterative algorithmic scheme which will work as an algorithmic structure that covers the CAV and BICAV algorithms and extends them from methods for solving linear equations to methods for solving the linear feasibility problem (i.e., both linear equations and linear inequalities). We use T to denote matrix transposition, but no ambiguity with the index T can arise.

ALGORITHM 4.1 (block-iterations for linear inequalities).
Initialization: *$x^0 \in \mathbb{R}^n$ is arbitrary.*
Iterative Step: *Given x^k, compute*

$$x^{k+1} = x^k + \lambda_k A_{t(k)}^T M_{t(k)}^k (b^{t(k)} - A_{t(k)} x^k), \tag{4.8}$$

where $\{\lambda_k\}_{k\geq 0}$ are relaxation parameters*, and $\{t(k)\}_{k\geq 0}$ is the* control sequence *governing which block is taken up at the kth iteration.*

For the choice $T = 1$ there is only one block, and we get the fully simultaneous version of Algorithm 4.1. In fact this method is then identical to Algorithm 2 of Censor and Elfving [6]. In addition to the cyclic control sequence, defined and used in Algorithms 3.1 and 3.2 above, we consider here two additional control sequences. These additional controls are problem-dependent. Denote by $d(x, L_i)$ the Euclidean distance between a point $x \in \mathbb{R}^n$ and the set L_i and define

$$\Phi(x) := \{\sup d(x, L_i) \mid 1 \leq i \leq m\}. \tag{4.9}$$

DEFINITION 4.1. (i) *We say that a sequence $\{t(k)\}_{k\geq 0}$ such that $1 \leq t(k) \leq T$ for all $k \geq 0$ is an* approximately remotest block control *sequence (with respect to the sequence $\{x^k\}_{k\geq 0}$, the family of sets $\{L_i\}_{i=1}^m$, and the blocks $\{B_t\}_{t=1}^T$) if, for every $k \geq 0$, there exists an $i \in B_{t(k)}$ such that*

$$\lim_{k\to\infty} d(x^k, L_i) = 0 \ \textit{implies that} \ \lim_{k\to\infty} \Phi(x^k) = 0. \tag{4.10}$$

(ii) *We say that a sequence $\{t(k)\}_{k\geq 0}$ such that $1 \leq t(k) \leq T$ for all $k \geq 0$ is a* remotest block control *sequence (with respect to the sequence $\{x^k\}_{k\geq 0}$, the family of sets $\{L_i\}_{i=1}^m$, and the blocks $\{B_t\}_{t=1}^T$) if, for every $k \geq 0$, there exists an $i \in B_{t(k)}$ such that*

$$\lim_{k\to\infty} d(x^k, L_i) = \Phi(x^k). \tag{4.11}$$

Every remotest block control is an approximately remotest block control. If all blocks consist of a single index, then these two definitions coincide with the definitions of the *approximately remotest set control* and the *remotest set control*, respectively, of Gubin, Polyak, and Raik [16, section 1] (see also [9, section 5.1]). We will prove the next result in what follows.

THEOREM 4.1. *Assume that $L \neq \emptyset$ and that the relaxation parameters are restricted to*

$$0 < \epsilon \leq \lambda_k \leq (2-\epsilon)/\rho(A_{t(k)}^T M_{t(k)}^k A_{t(k)}) \text{ for all } k \geq 0, \tag{4.12}$$

where ϵ is an arbitrarily small but fixed constant and $\{M_t\}_{t=1}^T$ are given symmetric and positive definite matrices with nonnegative elements. If $\{t(k)\}_{k\geq 0}$ is a cyclic control or an approximately remotest block control, then any sequence $\{x^k\}_{k\geq 0}$, generated by Algorithm 4.1*, converges to a solution of the system* (4.1)*.*

We also formulate the corresponding block-iterative algorithmic scheme for linear equalities

$$Ax = b. \tag{4.13}$$

ALGORITHM 4.2 (block-iterations for linear equalities).
Initialization: *$x^0 \in \mathbb{R}^n$ is arbitrary.*
Iterative Step: *Given x^k, compute*

$$x^{k+1} \; = \; x^k + \lambda_k A_{t(k)}^T M_{t(k)} (b^{t(k)} - A_{t(k)} x^k), \tag{4.14}$$

where $\{\lambda_k\}_{k\geq 0}$ are relaxation parameters*, and $\{t(k)\}_{k\geq 0}$ is the* control sequence *governing which block is taken up at the kth iteration.*

For this algorithm the following theorem will be proven in the next section.

THEOREM 4.2. *Assume that $H := \cap_{i=1}^m H_i \neq \emptyset$ and that the relaxation parameters are restricted to*

$$0 < \epsilon \leq \lambda_k \leq (2-\epsilon)/\rho(A_{t(k)}^T M_{t(k)} A_{t(k)}) \text{ for all } k \geq 0, \tag{4.15}$$

where ϵ is an arbitrarily small but fixed constant and $\{M_t\}_{t=1}^T$ are given symmetric and positive definite matrices with nonnegative elements. If $\{t(k)\}_{k\geq 0}$ is a cyclic control or an approximately remotest block control, then any sequence $\{x^k\}_{k\geq 0}$, generated by Algorithm 4.2*, converges to a solution of the system* (4.13)*. If, in addition, $x^0 \in R(A^T)$ (the range of A^T), then $\{x^k\}_{k\geq 0}$ converges to the solution of* (4.13)*, which has minimal Euclidean norm.*

5. Proofs of the convergence theorems. In proving Theorem 4.1 we use a convergence theory developed by Gubin, Polyak, and Raik [16]; see Bauschke and Borwein [2, Theorem 2.16 and Remark 2.17], which also contains a review and generalizations.

DEFINITION 5.1. *A sequence $\{x^k\}_{k\geq 0}$ is called* Fejér-monotone with respect to the set L *if, for every $x \in L$,*

$$||x^{k+1} - x||_2 \leq ||x^k - x||_2 \text{ for all } k \geq 0. \tag{5.1}$$

It is easy to verify that every Fejér-monotone sequence is bounded. The convergence theory of Gubin, Polyak, and Raik applies to convex closed sets in general. For the sets L_i, defined here, their theorem is the following.

THEOREM 5.1. *Let $L = \cap_{i=1}^m L_i \neq \emptyset$. If, for a sequence $\{x^k\}_{k\geq 0}$, the following conditions hold, then $\lim_{k\to\infty} x^k = x^* \in L$:*

(i) *$\{x^k\}_{k\geq 0}$ is Fejér-monotone with respect to L, and*

(ii) *$\lim_{k\to\infty} \Phi(x^k) = 0$.*

Theorem 4.1 will be proved by establishing the conditions of Theorem 5.1. First we establish, in the next proposition, condition (i) of Theorem 5.1.

PROPOSITION 5.2. *Under the assumptions of Theorem 4.1, any sequence $\{x^k\}_{k\geq 0}$, generated by Algorithm 4.1, is Fejér-monotone with respect to L, provided that $x^k \notin L$ for all $k \geq 0$.*

Proof. We use the notation

$$r^{t(k),k} := b^{t(k)} - A_{t(k)}x^k \quad \text{and} \quad d^{t(k),k} = M_{t(k)}^k r^{t(k),k}. \tag{5.2}$$

Let $x \in L$ (i.e., $b - Ax \geq 0$), and define $e^k := x^k - x$. Then, by (4.8),

$$e^{k+1} = e^k + \lambda_k A_{t(k)}^T d^{t(k),k}. \tag{5.3}$$

It follows that

$$||e^{k+1}||_2^2 = ||e^k||_2^2 + \lambda_k^2 ||A_{t(k)}^T d^{t(k),k}||_2^2 + 2\lambda_k \langle A_{t(k)}^T d^{t(k),k}, e^k\rangle. \tag{5.4}$$

From $x \in L_{i_j^{t(k)}}$ we obtain (recall that $b_j^{t(k)}$ is the jth component of the block $b^{t(k)}$ of the vector b)

$$r_j^{t(k),k} = b_j^{t(k)} - \langle a^{i_j^{t(k)}}, x^k\rangle \geq -\langle a^{i_j^{t(k)}}, e^k\rangle, \quad j = 1, 2, \ldots, m(t(k)). \tag{5.5}$$

Hence we have for the last summand on the right-hand side of (5.4) that

$$\begin{aligned} \langle A_{t(k)}^T d^{t(k),k}, e^k\rangle &= -\sum_{j=1}^{m(t(k))} d_j^{t(k),k} \langle -a^{i_j^{t(k)}}, e^k\rangle \\ &\leq -\sum_{j=1}^{m(t(k))} d_j^{t(k),k} r_j^{t(k),k} = -\langle d^{t(k),k}, r^{t(k),k}\rangle, \end{aligned} \tag{5.6}$$

provided that

$$d_j^{t(k),k} \leq 0 \quad \text{for } j = 1, 2, \ldots, m(t(k)) \text{ and for all } k \geq 0. \tag{5.7}$$

To see that (5.7) holds, observe that

$$d_j^{t(k),k} = \left(M_{t(k)}^k r^{t(k),k}\right)_j = \left(D_{t(k)}^k M_{t(k)} D_{t(k)}^k r^{t(k),k}\right)_j = \left(D_{t(k)}^k\right)_{jj} \sum_{s\in I_{t(k)}^k} m_s^{i_j^{t(k)}} r_s^{t(k),k}, \tag{5.8}$$

where $\{m_s^{i_j^{t(k)}}\}$ are the entries of the $i_j^{t(k)}$th row of $M_{t(k)}$, which are nonnegative by assumption, and observe that $r_s^{t(k),k} < 0$ whenever $s \in I_{t(k)}^k$.

Turning now to the second summand in the right-hand side of (5.4), we decompose the semidefinite matrix $M_{t(k)}^k$ as $M_{t(k)}^k = W^T W$ and use the well-known inequality

$$\langle Qy, y\rangle \leq \rho(Q)\langle y, y\rangle, \tag{5.9}$$

which holds for any symmetric and positive semidefinite matrix Q (where $\rho(Q)$ denotes the spectral radius of the matrix Q; see, e.g., Demmel [12, equation (5.2)]), to obtain

$$\begin{aligned} ||A_{t(k)}^T d^{t(k),k}||_2^2 &= \langle A_{t(k)}^T M_{t(k)}^k r^{t(k),k}, A_{t(k)}^T M_{t(k)}^k r^{t(k),k}\rangle \\ &= \langle M_{t(k)}^k A_{t(k)} A_{t(k)}^T M_{t(k)}^k r^{t(k),k}, r^{t(k),k}\rangle \\ &= \langle (W A_{t(k)} A_{t(k)}^T W^T) W r^{t(k),k}, W r^{t(k),k}\rangle \\ &\le \rho(W A_{t(k)} A_{t(k)}^T W^T)\langle W r^{t(k),k}, W r^{t(k),k}\rangle \\ &= \rho(A_{t(k)}^T M_{t(k)}^k A_{t(k)})\langle d^{t(k),k}, r^{t(k),k}\rangle. \qquad (5.10) \end{aligned}$$

Substituting (5.10) and (5.6) into (5.4), we get

$$||e^{k+1}||_2^2 \le ||e^k||_2^2 + \lambda_k(\lambda_k \rho(A_{t(k)}^T M_{t(k)}^k A_{t(k)}) - 2)\langle d^{t(k),k}, r^{t(k),k}\rangle, \qquad (5.11)$$

where $\langle d^{t(k),k}, r^{t(k),k}\rangle = \langle W r^{t(k),k}, W r^{t(k),k}\rangle \ge 0$. Now using (4.12), the desired conclusion $||e^{k+1}|| \le ||e^k||$ follows. ∎

Note that if $I_{t(k)}^k = \emptyset$ (i.e., $A_{t(k)} x^k \le b^{t(k)}$), then $D_{t(k)}^k = 0$, and hence $d^{t(k),k} = 0$ so that the second summand in the right-hand side of (5.11) disappears. The next proposition establishes condition (ii) of Theorem 5.1.

PROPOSITION 5.3. *Under the assumptions of Theorem* 4.1*, any sequence* $\{x^k\}_{k\ge 0}$*, generated by Algorithm* 4.1*, has the property*

$$\lim_{k\to\infty} \Phi(x^k) = 0. \qquad (5.12)$$

Proof. Fejér-monotonicity, guaranteed by Proposition 5.2, implies that the sequence $\{||e^k||_2\}_{k\ge 0}$ is monotonically decreasing, and thus converging. It follows then from (5.11) that

$$\lim_{k\to\infty} \langle d^{t(k),k}, r^{t(k),k}\rangle = 0. \qquad (5.13)$$

But

$$\langle d^{t(k),k}, r^{t(k),k}\rangle = \langle M_{t(k)}^k r^{t(k),k}, r^{t(k),k}\rangle = \langle M_{t(k)} D_{t(k)}^k r^{t(k),k}, D_{t(k)}^k r^{t(k),k}\rangle, \qquad (5.14)$$

and thus

$$\lim_{k\to\infty} D_{t(k)}^k r^{t(k),k} = 0. \qquad (5.15)$$

Using (4.4),

$$\left(D_{t(k)}^k r^{t(k),k}\right)_j = c_{i_j^{t(k)}}^k ||a^{i_j^{t(k)}}||_2^2, \quad j = 1, 2, \ldots, m(t(k)), \qquad (5.16)$$

leads to

$$\begin{aligned} d(x^k, L_{i_j^{t(k)}}) &= ||P_{L_{i_j^{t(k)}}}(x^k) - x^k||_2 \\ &= ||c_{i_j^{t(k)}}^k a^{i_j^{t(k)}}||_2 = \left|\left(D_{t(k)}^k r^{t(k),k}\right)_j\right| / ||a^{i_j^{t(k)}}||_2 \qquad (5.17) \end{aligned}$$

for all $j = 1, 2, \ldots, m(t(k))$. This shows, by (5.15), that

$$\lim_{k\to\infty} d(x^k, L_i) = 0 \text{ for all } i \in B_{t(k)}. \tag{5.18}$$

If $\{t(k)\}_{k\geq 0}$ is an approximately remotest block control, then the required result follows directly from (5.18) and Definition 4.1(i) and Assumption 3.1. For a cyclic control we argue as follows. From (4.8) and (4.7) we get

$$\begin{aligned} ||x^{k+1} - x^k||_2 &= \lambda_k ||A^T_{t(k)} D^k_{t(k)} M_{t(k)} D^k_{t(k)} r^{t(k),k}||_2 \\ &\leq \lambda_k ||A^T_{t(k)} D^k_{t(k)} M^{1/2}_{t(k)}||_2 \cdot ||M^{1/2}_{t(k)}||_2 ||D^k_{t(k)} r^{t(k),k}||_2. \end{aligned} \tag{5.19}$$

Therefore, using (4.12) and the fact that, for any matrix Q, it is true that $\rho(Q^TQ) = ||Q^T||_2^2$ (see, e.g., Demmel [12, Fact 9, p. 23]), we obtain

$$||x^{k+1} - x^k||_2 \leq \theta_1 \theta_2^{-1} ||D^k_{t(k)} r^{t(k),k}||_2, \tag{5.20}$$

where

(5.21)
$$\theta_1 := 2\max\{||M_i^{1/2}||_2 \mid 1 \leq i \leq T\} \text{ and } \theta_2 := \max\{||A_i^T D_i^k M_i^{1/2}||_2 \mid 1 \leq i \leq T\}.$$

The max in the expression of θ_2 exists and is independent of k because of the way these matrices were defined. If $\theta_2 = 0$, then, by (4.8), $x^{k+1} = x^k$. If, on the other hand, $\theta_2 \neq 0$, then θ_2 is bounded away from zero and, thus, (5.15) and (5.20) yield

$$\lim_{k\to\infty} ||x^{k+1} - x^k||_2 = 0. \tag{5.22}$$

Let $\epsilon > 0$ be such that for all $k \geq K$, we have $||x^{k+1} - x^k||_2 \leq \epsilon/T$. To reach the required conclusion (5.12) we look at $d(x^k, L_i) = ||P_{L_i}(x^k) - x^k||_2$ and observe that if $i \in B_{t(k)}$, then (5.18) shows that $||P_{L_i}(x^k) - x^k||_2 \leq \epsilon$ for all $k \geq K$. Otherwise, if $i \notin B_{t(k)}$, the cyclicality of $\{t(k)\}_{k\geq 0}$ guarantees that there exists a τ such that $1 \leq \tau < T$ and $i \in B_{t(k+\tau)}$. Then,

$$\begin{aligned} d(x^k, L_i) &= ||x^k - P_{L_i}(x^k)||_2 \leq ||x^k - P_{L_i}(x^{k+\tau})||_2 \\ &\leq ||x^k - x^{k+\tau}||_2 + ||x^{k+\tau} - P_{L_i}(x^{k+\tau})||_2 \\ &\leq ||x^k - x^{k+1}||_2 + \cdots + ||x^{k+\tau-1} - x^{k+\tau}||_2 + ||x^{k+\tau} - P_{L_i}(x^{k+\tau})||_2 \\ &\leq (T-1)(\epsilon/T) + \epsilon = \epsilon \end{aligned} \tag{5.23}$$

for all $k \geq K$. Therefore, $\Phi(x^k) \leq \epsilon$ for all $k \geq K$, and, using Assumption 3.1, the result follows. $\square$

So, we see that the last two propositions, combined with Theorem 4.1, imply the truth of Theorem 4.1.

Proof of Theorem 4.2. Theorem 4.2 follows from Theorem 4.1. To simplify the discussion we deal only with the case that the weight matrices $\{M_t\}$ are positive diagonal matrices. This assumption actually holds in all three examples given in section 7. The general case can be proved along lines similar to the following argument. Any equation $\langle a^i, x\rangle = b_i$ can be written as a pair of inequalities $\langle a^i, x\rangle \leq b_i$ and $\langle -a^i, x\rangle \leq -b_i$. Now for a given linear system $Ax = b$, where $A \in \mathbf{R}^{m\times n}$, and given diagonal weight matrices $\{M_t\}$ we construct the inequalities $\tilde{A}x \leq \tilde{b}$ as follows:

$$\tilde{a}^{2i-1} = a^i, \quad \tilde{a}^{2i} = -a^i, \quad \tilde{b}_{2i-1} = b_i, \quad \tilde{b}_{2i} = -b_i, \qquad i = 1, 2, \ldots, m. \tag{5.24}$$

Denoting the (i,j)th element of a matrix A by $(A)_{i,j}$, we also set

$$\left(\tilde{M}_t\right)_{2i-1,2i-1} = \left(\tilde{M}_t\right)_{2i,2i} = (M_t)_{i,i} \quad \text{for all } i = 1,2,\ldots,m. \tag{5.25}$$

Recall that $\tilde{M}_t^k = D_t(x^k)\tilde{M}_t D_t(x^k)$, where the matrix $D_t(z)$ is defined in (4.6). Then, for any x^k, one can verify that

$$\tilde{A}_{t(k)}^T \tilde{M}_{t(k)}^k (\tilde{b}^{t(k)} - \tilde{A}_{t(k)} x^k) = A_{t(k)}^T M_{t(k)} (b^{t(k)} - A_{t(k)} x^k) \tag{5.26}$$

so that the two iteration formulas (4.8) and (4.14) generate the same sequence of iterates, provided they are initialized with the same vector. It is also true, for any x^k, that

$$\rho(\tilde{A}_{t(k)}^T \tilde{M}_{t(k)}^k \tilde{A}_{t(k)}) = \rho(A_{t(k)}^T M_{t(k)} A_{t(k)}); \tag{5.27}$$

hence Theorem 4.2 follows. □

6. The inconsistent case. When there is just one block, i.e., $t = T = 1$, the resulting methods are fully simultaneous. We consider here the inconsistent case behavior only for linear equations. Let $M_1 = M$, $c = A^T M b$, and $\Gamma = A^T M A$. Then the iteration (4.14) can be written as

$$x^{k+1} = x^k + \lambda_k (c - \Gamma x^k). \tag{6.1}$$

This is the nonstationary *Richardson iteration method*; cf. Young [24, p. 361]. We observe that $c \in R(\Gamma)$ (the range of Γ) and, if we assume that $\hat{x}$ satisfies $c = \Gamma \hat{x}$, then $\hat{x} = \arg\min ||Ax - b||_M$ (with $||x||_M^2 = \langle x, Mx \rangle$). Let $u^k = \hat{x} - x^k$ and note that, with $v^k = c - \Gamma x^k$, it is true that $v^k = \Gamma u^k$. It follows that

$$u^k = \prod_{j=0}^{k-1} (I - \lambda_j \Gamma) u^0. \tag{6.2}$$

Assume first that Γ is a positive definite matrix. Then any sequence $\{x^k\}_{k \geq 0}$ generated by Algorithm 4.2, as given by (6.1), is convergent for any x^0 if and only if

$$\lim_{k \to \infty} \prod_{j=0}^{k-1} (I - \lambda_j \Gamma) = 0. \tag{6.3}$$

Since $|| \prod_{j=0}^{k-1} (I - \lambda_j \Gamma) ||_2 \leq \prod_{j=0}^{k-1} \rho(I - \lambda_j \Gamma)$, it follows that any sequence $\{x^k\}_{k \geq 0}$, generated by Algorithm 4.2, as given by (6.1), converges to a weighted least squares solution if $0 < \epsilon \leq \lambda_k \leq (2 - \epsilon)/\rho(\Gamma)$. In case Γ is only positive semidefinite we have a similar result. All of these observations lead to the following theorem.

THEOREM 6.1. *Assume that M is a positive definite matrix. If $0 < \epsilon \leq \lambda_k \leq (2-\epsilon)/\rho(A^T M A)$ for all $k \geq 0$, where ϵ is an arbitrarily small but fixed constant, then any sequence $\{x^k\}_{k \geq 0}$, generated by Algorithm* 4.2, *as given by* (6.1), *converges to a weighted least squares solution $\hat{x} = \arg\min ||Ax - b||_M$. If, in addition, $x^0 \in R(A^T)$, then $\{x^k\}_{k \geq 0}$ converges to the unique solution of minimal Euclidean norm among all weighted least squares solutions.*

The proof of Theorem 6.1 can essentially be found in, e.g., Eggermont, Herman, and Lent [13, p. 44]; see also Elfving [14, p. 4].

We do not give a proof of convergence for the case of linear inequalities. We note, however, that a variant of Algorithm 4.1 for $T = 1$ (Cimmino's method; see Example 7.3 below) was shown to converge locally for the inconsistent case by Iusem and De Pierro [18] and to converge globally in that case by Combettes [10].

7. Applications. In this section we will consider only diagonal matrices $M_t = \operatorname{diag}\{\mu_j^t \mid j = 1, 2, \ldots, m(t)\}$ with positive diagonal elements. For such diagonal matrices let

$$W_t := A_t^T M_t A_t \text{ for all } t = 1, 2, \ldots, T \quad \text{and} \quad W_{t(k)}^k := A_{t(k)}^T M_{t(k)}^k A_{t(k)} \text{ for all } k \geq 0, \tag{7.1}$$

and note the expansions

$$W_{t(k)}^k = \sum_{j \in I_{t(k)}^k} \mu_j^{t(k)} a^{i_j^{t(k)}} \left(a^{i_j^{t(k)}}\right)^T, \quad W_{t(k)} = \sum_{j=1}^{m(t(k))} \mu_j^{t(k)} a^{i_j^{t(k)}} \left(a^{i_j^{t(k)}}\right)^T. \tag{7.2}$$

Hence the iterative step of Algorithm 4.1 takes the form

$$x^{k+1} = x^k + \lambda_k \sum_{j \in I_{t(k)}^k} \mu_j^{t(k)} \left(b_j^{t(k)} - \langle a^{i_j^{t(k)}}, x^k \rangle\right) a^{i_j^{t(k)}}, \tag{7.3}$$

and the iterative step of Algorithm 4.2 becomes

$$x^{k+1} = x^k + \lambda_k \sum_{j=1}^{m(t(k))} \mu_j^{t(k)} \left(b_j^{t(k)} - \langle a^{i_j^{t(k)}}, x^k \rangle\right) a^{i_j^{t(k)}}. \tag{7.4}$$

Also note that, by (7.2), for all $k \geq 0$,

$$\rho\big(W_{t(k)}^k\big) \leq \rho\big(W_{t(k)}\big). \tag{7.5}$$

In the following examples we show that several algorithms, including the BICAV and simultaneous algebraic reconstruction technique (SART) algorithms, are in fact special cases of the algorithmic schemes studied in the previous sections.

Example 7.1. The BICAV (Algorithm 3.2) and CAV (Algorithm 2.3) are both algorithms for equalities and of the form (7.4) with

$$\mu_j^{t(k)} = \frac{1}{\|a^{i_j^{t(k)}}\|_{S_{t(k)}}^2} = \frac{1}{\sum_{\nu=1}^n s_\nu^{t(k)} \left(a_\nu^{i_j^{t(k)}}\right)^2}, \quad j = 1, 2, \ldots, m(t(k)). \tag{7.6}$$

Here $\{t(k)\}_{k \geq 0}$ is the control sequence, $s_\nu^{t(k)}$ is the number of nonzero elements in the νth column of the block $A_{t(k)}$, and $S_{t(k)} := \operatorname{diag}\{s_\nu^{t(k)} \mid \nu = 1, 2, \ldots, n\}$. We first study the upper bound on the relaxation parameters for CAV, i.e., allowing one block only so that $t = T = 1$ and $m(1) = m$; cf. (3.1). The following result (Lemma 7.1) is due to Dr. Arnold Lent [22] (see the acknowledgments at the end of this paper).

LEMMA 7.1. *Let $t = T = 1$ and $m(1) = m$, let $M := \operatorname{diag}\{\mu_j \mid j = 1, 2, \ldots, m\}$ with $\mu_j = \mu_j^1$ obtained from* (7.6) *for $t = t(k) = 1$, and let $A_1 = A$, $s_\nu^1 = s_\nu$, $S_1 = S$, and $W := A^T M A$. Then $\rho(W) \leq 1$.*

Proof. Let a_j^i be the element in the ith row and jth column of A and write, by (7.6),

$$(\mu_i)^{-1} = \sum_{j=1}^n s_j \left(a_j^i\right)^2, \quad i = 1, 2, \ldots, m. \tag{7.7}$$

Let (λ, v) be an eigenpair (i.e., eigenvalue and eigenvector) of W so that $A^T MAv = \lambda v$ or $AA^T MAv = \lambda M^{-1} MAv$, or, with $w := MAv$, $AA^T w = \lambda M^{-1} w$. Hence $||A^T w||_2^2 = \lambda w^T M^{-1} w$ or, in component form, switching the order of summations and using (7.7),

(7.8)
$$||A^T w||_2^2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_j^i w_i \right)^2 = \lambda \sum_{i=1}^{m} w_i^2 \left(\sum_{j=1}^{n} s_j \left(a_j^i\right)^2 \right) = \lambda \sum_{j=1}^{n} s_j \left(\sum_{i=1}^{m} w_i^2 \left(a_j^i\right)^2 \right).$$

From Cauchy's inequality we have

$$\left(\sum_{i=1}^{m} a_j^i w_i \right)^2 \leq s_j \sum_{i=1}^{m} w_i^2 \left(a_j^i\right)^2, \tag{7.9}$$

and by summing both sides of (7.9) over j and comparing with (7.8), one finds that $\lambda \leq 1$. ∎

Remark 7.1. The critical estimate is (7.9). Let a, w, and e be three vectors of equal length. Denote by $z = a * w$ componentwise multiplication, i.e., $z_j = a_j w_j$ for all j. Further, let $e_j = 0$ if $z_j = 0$, and let $e_j = 1$ otherwise. Then

$$\langle a, w \rangle^2 = \langle e, z \rangle^2 \leq ||e||_2^2 \cdot ||z||_2^2 \leq s||z||_2^2, \tag{7.10}$$

where s is the number of nonzero elements in the vector a.

By applying Lemma 7.1 to each block A_t, $t = 1, 2, \ldots, T$, we obtain the following.

COROLLARY 7.1. *Let* $M_{t(k)} = \mathrm{diag}\{\mu_j^{t(k)} \mid j = 1, 2, \ldots, m(t(k))\}$, $k \geq 0$, *with* $\mu_j^{t(k)}$ *obtained from* (7.6), *and let* $W_{t(k)} = A_{t(k)}^T M_{t(k)} A_{t(k)}$. *Then* $\rho(W_{t(k)}) \leq 1$ *for all* $k \geq 0$.

The next theorems establish the convergence of the BICAV algorithm in the consistent case for linear equations and linear inequalities, respectively, with relaxation parameters within the interval $[\epsilon, 2 - \epsilon]$.

THEOREM 7.1 (BICAV for linear equalities). *Let* $0 < \epsilon \leq \lambda_k \leq 2 - \epsilon$ *for all* $k \geq 0$, *where* ϵ *is an arbitrarily small but fixed constant. If the system* (4.13) *is consistent, then any sequence* $\{x^k\}_{k \geq 0}$, *generated by Algorithm* 3.2 *(BICAV), converges to a solution of the system* (4.13). *If, in addition,* $x^0 \in R(A^T)$, *then* $\{x^k\}_{k \geq 0}$ *converges to the solution of* (4.13), *which has minimal Euclidean norm.*

Proof. The proof follows from Theorem 4.2 and Corollary 7.1. ∎

THEOREM 7.2 (BICAV for linear inequalities). *Let* $0 < \epsilon \leq \lambda_k \leq 2 - \epsilon$ *for all* $k \geq 0$, *where* ϵ *is an arbitrarily small but fixed constant. If the system* (4.1) *is consistent, then any sequence* $\{x^k\}_{k \geq 0}$, *generated by Algorithm* 4.1, *with* $M_{t(k)} = \mathrm{diag}\{\mu_j^{t(k)} \mid j = 1, 2, \ldots, m(t(k))\}$ *and* $\{\mu_j^{t(k)}\}$ *given by* (7.6), *converges to a solution of the system* (4.1).

Proof. The proof follows from Theorem 4.1, Corollary 7.1, and (7.5). ∎

The next theorem shows that any sequence $\{x^k\}_{k \geq 0}$, generated by the fully simultaneous Algorithm 2.3 (CAV), converges to a weighted least squares solution of the system of equations $Ax = b$, regardless of its consistency, for relaxation parameters in the interval $[\epsilon, 2 - \epsilon]$. Only the case of unity relaxation, i.e., $\lambda_k = 1$ for all $k \geq 0$, was proven by Censor, Gordon, and Gordon in [7], where CAV was first proposed and experimented with.

THEOREM 7.3 (CAV for linear equalities). *If $0 < \epsilon \leq \lambda_k \leq 2 - \epsilon$ for all $k \geq 0$, where ϵ is an arbitrarily small but fixed constant, then any sequence $\{x^k\}_{k\geq 0}$, generated by Algorithm 2.3 (CAV for linear equations), converges to a weighted least squares solution with weight matrix $M_1 = M_{CAV} = \mathrm{diag}\{1/||a^i||_S^2 \mid i = 1, 2, \ldots, m\}$ and with $S = \mathrm{diag}\{s_j \mid j = 1, 2, \ldots, n\}$, where s_j is the number of nonzero elements in the jth column of A. If, in addition, $x^0 \in R(A^T)$, then $\{x^k\}_{k\geq 0}$ converges to the unique solution of minimal Euclidean norm among all weighted least squares solutions.*

Proof. The proof follows from Theorem 6.1 and Lemma 7.1. □

Note that Theorems 7.1 and 7.2 assumed cyclic control of the blocks, as formulated in Algorithm 3.2; however, due to the analysis presented here, we may also allow approximately remotest block control of the blocks (by Theorems 4.2 and 6.1). Recently, and independently of our work, Byrne [5] derived convergence results analogous to Theorems 7.1 and 7.3, but only for the cyclic control and without explicit consideration of weighting. He also used Lent's result as expressed above in Lemma 7.1.

Example 7.2. The *simultaneous algebraic reconstruction technique (SART)* was proposed by Andersen and Kak [1] for solving the large and very sparse systems of linear equations arising from a fully discretized model of transmission computerized tomography problems; see also Kak and Slaney [21, section 7.4]. We show that a simplified version of SART falls within the convergence analysis presented here. First recall that the 1-*norm* of a vector $x \in \mathbb{R}^n$ is $||x||_1 = \sum_{j=1}^n |x_j|$ and that the induced matrix norm of an $m \times n$ matrix A is $||A||_1 = \max\{\sum_{i=1}^m |a_{ij}| \mid j = 1, 2, \ldots, n\}$. Let $a_c^{l,t}$ be the lth column of A_t. Then the iterative step of the original SART algorithm for linear equalities [1, equation (32)] (see also [23, equation (4)]) is

$$x_l^{k+1} = x_l^k + \frac{\lambda_k}{||a_c^{l,t(k)}||_1} \sum_{j=1}^{m(t(k))} \frac{b_j^{t(k)} - \langle a^{i_j^{t(k)}}, x^k \rangle}{||a^{i_j^{t(k)}}||_1} a_l^{i_j^{t(k)}}, \quad l = 1, 2, \ldots, n. \tag{7.11}$$

Note that in (7.11) it is tacitly assumed that all blocks A_t have nonzero columns. The formula (7.11) is slightly more general than the original algorithm in [1] since it allows (i) a relaxation parameter λ_k, (ii) a more flexible row-partitioning (originally the matrix was partitioned into nonoverlapping row blocks, where each block corresponds to all equations in one tomographic scan direction), (iii) arbitrary sign of the matrix elements (originally only nonnegative elements were considered), and (iv) apart from the cyclic control of blocks also the remotest block control.

We first note that (7.11) can be written in matrix-vector form, using our previous notation, as

$$x^{k+1} = x^k + \lambda_k D_{t(k)} A_{t(k)}^T M_{t(k)} (b^{t(k)} - A_{t(k)} x^k), \tag{7.12}$$

where

$$D_{t(k)} = \mathrm{diag}\{1/||a_c^{l,t(k)}||_1 \mid l = 1, 2, \ldots, n\} \tag{7.13}$$

and

$$M_{t(k)} = \mathrm{diag}\{1/||a^{i_j^t}||_1 \mid j = 1, 2, \ldots, m(t(k))\}. \tag{7.14}$$

We will not, however, analyze this iteration here. Instead, we consider a simplified version which fits into the class of methods (4.8) and (4.14), respectively. Let a_c^l be the lth column of A and put $D = \mathrm{diag}\{1/||a_c^l||_1 \mid l = 1, 2, \ldots, n\}$.

Replacing $D_{t(k)}$ by D in (7.12) we get

$$x^{k+1} \; = \; x^k + \lambda_k D A^T_{t(k)} M_{t(k)} (b^{t(k)} - A_{t(k)} x^k). \tag{7.15}$$

We will call the method which uses the iterative step (7.15) *block simplified SART (BSSART)*. The method is a scaled version of (7.4). To see this we put

$$y^k = D^{-1/2} x^k \quad \text{and} \quad \bar{A}_{t(k)} = A_{t(k)} D^{1/2}, \tag{7.16}$$

which converts (7.15) into

$$y^{k+1} \; = \; y^k + \lambda_k \bar{A}^T_{t(k)} M_{t(k)} (b^{t(k)} - \bar{A}_{t(k)} y^k), \tag{7.17}$$

which is of the form (7.4) (or equivalently (4.14)). Next observe that with $W_{t(k)} = D^{1/2} A^T_{t(k)} M_{t(k)} A_{t(k)} D^{1/2}$, we have

$$\rho(W_{t(k)}) = \rho(A^T_{t(k)} M_{t(k)} A_{t(k)} D) \leq ||A^T_{t(k)} M_{t(k)}||_1 \cdot ||A_{t(k)} D||_1 = 1. \tag{7.18}$$

It follows from Theorem 4.2 that y^k converges to some y^*. Since, by (7.16), every row of A is postmultiplied by $D^{1/2}$, we also conclude that $AD^{1/2} y^* = b$. Then, using (7.16),

$$\lim_{k \to \infty} x^k = D^{1/2} y^* = x^*. \tag{7.19}$$

Hence $Ax^* = b$.

Now consider BSSART adapted to inequalities, i.e., the iterative step

$$x^{k+1} \; = \; x^k + \lambda_k D A^T_{t(k)} M^k_{t(k)} (b^{t(k)} - A_{t(k)} x^k). \tag{7.20}$$

It is clear, using (7.5) and Theorem 4.2, that the above analysis also holds for the iteration (7.20). Hence the following companion results to Theorems 7.1 and 7.2 hold.

THEOREM 7.4 (BSSART for linear equalities). *Let* $0 < \epsilon \leq \lambda_k \leq 2 - \epsilon$ *for all* $k \geq 0$, *where* ϵ *is an arbitrarily small but fixed constant. If the system* (4.13) *is consistent, then any sequence* $\{x^k\}_{k \geq 0}$, *generated by the iterative step* (7.15) *(BSSART), converges to a solution of the system* (4.13).

THEOREM 7.5 (BSSART for linear inequalities). *Let* $0 < \epsilon \leq \lambda_k \leq 2 - \epsilon$ *for all* $k \geq 0$, *where* ϵ *is an arbitrarily small but fixed constant. If the system* (4.1) *is consistent, then any sequence* $\{x^k\}_{k \geq 0}$, *generated by the iterative step* (7.20) *(BSSART for inequalities), converges to a solution of the system* (4.1).

When $T = 1$, SART (7.12) and BSSART (7.15) coincide and can be written

$$x^{k+1} \; = \; x^k + \lambda_k D A^T M (b - Ax^k), \tag{7.21}$$

with $M = \text{diag}\{1/||a^j||_1 \mid j = 1, 2, \ldots, m\}$. Using the corresponding transformations as in (7.16),

$$y^k = D^{-1/2} x^k \quad \text{and} \quad \bar{A} = AD^{1/2}, \tag{7.22}$$

we find that

$$y^{k+1} \; = \; y^k + \lambda_k \bar{A}^T M (b - \bar{A} y^k). \tag{7.23}$$

It follows from Theorem 6.1, and by using (as above) the fact that $\rho(A^T MAD) \leq 1$, that

$$\lim_{k\to\infty} y^k = y^* \quad \text{such that } ||\bar{A}y^* - b||_M \quad \text{is minimal.} \tag{7.24}$$

But $\lim_{k\to\infty} x^k = D^{1/2}y^* = x^*$ so that x^* minimizes $||Ax - b||_M$. Also, by using

$$||y^*||_2 = ||D^{-1/2}D^{1/2}y^*||_2 = ||x^*||_{D^{-1}}, \tag{7.25}$$

it follows that x^* has minimal D^{-1}-norm. Hence the following result holds.

THEOREM 7.6. *If $0 < \epsilon \leq \lambda_k \leq 2 - \epsilon$ for all $k \geq 0$, where ϵ is an arbitrarily small but fixed constant, then any sequence $\{x^k\}_{k\geq 0}$, generated by Algorithm* (7.21)*, converges to a weighted least squares solution with weight matrix $M = \mathrm{diag}\{1/||a^i||_1 \mid i = 1, 2, \ldots, m\}$. If, in addition, $x^0 \in R(DA^T)$, then the limit point has minimal D^{-1}-norm.*

No proof of convergence was given in [1] or has, to the best of our knowledge, been published elsewhere since then. Recently, however, and independently of our work, Jiang and Wang [19] have also derived, under the additional assumption that the elements of the matrix A are nonnegative, Theorem 7.6.

Example 7.3. Block-Cimmino methods for linear equations and linear inequalities can also be viewed as special cases of Algorithms 4.1 and 4.2. To see this we define

$$\mu_j^{t(k)} = \frac{\theta_{i_j^{t(k)}}}{||a^{i_j^{t(k)}}||_2^2}, \quad j = 1, 2, \ldots, m(t(k)), \tag{7.26}$$

where $\theta_{i_j^{t(k)}} > 0$ and $\sum_{j=1}^{m(t(k))} \theta_{i_j^{t(k)}} = 1$. It follows, using (7.2), that $\rho(W_{t(k)}) = ||W_{t(k)}||_2 \leq \sum_{j=1}^{m(t(k))} \theta_{i_j^{t(k)}} = 1$ and that

$$\rho(W_{t(k)}^k) = ||W_{t(k)}^k||_2 \leq \sum_{j\in I_{t(k)}^k} \theta_{i_j^{t(k)}} \leq 1. \tag{7.27}$$

Therefore, also in this example, we may conclude convergence just as in Theorems 7.1, 7.2, and 7.3 with $M_1 = M_{CIM} = \mathrm{diag}\{\theta_i/||a^i||_2^2 \mid i = 1, 2, \ldots, m\}$ in Theorem 7.3. The geometric interpretation of this scaling is as follows. By (2.2),

$$P_{H_i}(x) - x = (b_i - \langle a^i, x\rangle)\frac{a^i}{||a^i||_2^2}, \tag{7.28}$$

so that

$$\sum_{i=1}^m \theta_i ||P_{H_i}(x) - x||_2^2 = \sum_{i=1}^m \frac{\theta_i(b_i - \langle a^i, x\rangle)^2}{||a^i||_2^2} = ||b - Ax||_{M_{CIM}}^2. \tag{7.29}$$

Cimmino's original algorithm for linear equations [11] is purely simultaneous ($T = 1$), i.e., of the form (2.4). An interesting detail is that $\lambda_k = 2$ is used by Cimmino, and for this a special convergence analysis is furnished. We also remark that for *inequalities* the requirement on the relaxation parameters can be relaxed, using (7.27), to

$$0 < \epsilon \leq \lambda_k \leq \frac{2 - \epsilon}{\sum_{j\in I_{t(k)}^k} \theta_{i_j^{t(k)}}}. \tag{7.30}$$

In fact, the choice $\lambda_k = 2/\sum_{j\in I^k_{t(k)}} \theta_{i_j^{t(k)}}$ is also allowed but requires a special analysis, which appears, for the fully simultaneous case $T = 1$, assuming consistency, in Censor and Elfving [6]. See also Bauschke and Borwein [2, Remark 6.48] for a correction. A similar analysis can be done also for the block-iterative case. Iusem and De Pierro [18] have shown that this method (with $T = 1$) also converges (locally) for the inconsistent case and generalized it to closed convex sets in $\mathbb{R}^n$. A generalization to global convergence in infinite dimensional Hilbert spaces was done by Combettes [10].

We finally mention that if all block sizes are equal to 1 ($m(t) = 1$) and linear equations are considered, then we get the algebraic reconstruction technique (ART) of Gordon, Bender, and Herman [15], also known as Kaczmarz's method. For more on the history of this method and many of its variants, see, for example, Herman [17] and Censor and Zenios [9].

Acknowledgments. We thank Charles Byrne, Dan Gordon, Rachel Gordon, Arnold Lent, Robert Lewitt, and Samuel Matej for their enlightening comments and discussions on this work. Special thanks are due to Charles Byrne for sharing with us his paper [5] on this topic and to Ming Jiang and Ge Wang for making available to us their paper [19]. Lemma 7.1 is due to Dr. Arnold Lent from AT&T Labs, who made it available to us during a blackboard discussion with the first author and Professor Charles Byrne from the Department of Mathematical Sciences at the University of Massachusetts at Lowell. Part of this work was done during visits of Y. Censor to the Department of Mathematics of the University of Linköping, Linköping, Sweden. The support and hospitality of Professor Åke Björck, head of the Numerical Analysis Group there at the time, are gratefully acknowledged.

REFERENCES

[1] A.H. Andersen and A.C. Kak, *Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm*, Ultrasonic Imaging, 6 (1984), pp. 81–94.

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

[3] A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*, John Wiley & Sons, New York, 1974.

[4] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[5] C. Byrne, *Notes on Block-Iterative or Ordered Subset Methods for Image Reconstruction*, Technical Report, 2000.

[6] Y. Censor and T. Elfving, *New methods for linear inequalities*, Linear Algebra Appl., 42 (1982), pp. 199–211.

[7] Y. Censor, D. Gordon, and R. Gordon, *Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems*, Parallel Comput., 27 (2001), pp. 777–808.

[8] Y. Censor, D. Gordon, and R. Gordon, *BICAV: An inherently parallel algorithm for sparse systems with pixel-dependent weighting*, IEEE Trans. Medical Imaging, 20 (2001), pp. 1050–1060.

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

[10] P.L. Combettes, *Inconsistent signal feasibility problems: Least-squares solutions in a product space*, IEEE Trans. Signal Process., SP-42 (1994), pp. 2955–2966.

[11] G. Cimmino, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*, Ric. Sci. Progr. Tecn. Econom. Naz., 1 (1938), pp. 326–333.

[12] J.W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.

[13] P.P.B. Eggermont, G.T. Herman, and A. Lent, *Iterative algorithms for large partitioned linear systems, with applications to image reconstruction*, Linear Algebra Appl., 40 (1981), pp. 37–67.

[14] T. Elfving, *Block-iterative methods for consistent and inconsistent linear equations*, Numer. Math., 35 (1980), pp. 1–12.

[15] R. Gordon, R. Bender, and G.T. Herman, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography*, J. Theoret. Biology, 29 (1970), pp. 471–481.

[16] L.G. Gubin, B.T. Polyak, and E.V. Raik, *The method of projections for finding the common point of convex sets*, Comput. Math. Math. Phys., 7 (1967), pp. 1–24.

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

[18] A.N. Iusem and A.R. De Pierro, *Convergence results for an accelerated nonlinear Cimmino algorithm*, Numer. Math., 49 (1986), pp. 367–378.

[19] M. Jiang and G. Wang, *Convergence of the simultaneous algebraic reconstruction technique (SART)* (invited talk), in Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2001.

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

[21] A.C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Classics Appl. Math. 33, SIAM, Philadelphia, PA, 2001; also available online from http://www.slaney.org/pct/pct-toc.html.

[22] A. Lent, *Private communication*.

[23] K. Mueller, R. Yagel, and J.J. Wheller, *Anti-aliased three-dimensional cone-beam reconstruction of low-contrast objects with algebraic methods*, IEEE Trans. Medical Imaging, 18 (1999), pp. 519–537.

[24] D. Young, *Iterative Solution of Linear Systems*, Academic Press, New York, 1971.