

Convergence properties of ART and SOR algorithms

L. Elsner¹, I. Koltracht², and P. Lancaster^{3,*}

¹Institute of Mathematics, University of Bielefeld, W-4800 Bielefeld,
Federal Republic of Germany

²Department of Mathematics, University of Connecticut, Storrs, CT 06268, USA

³Department of Mathematics and Statistics, University of Calgary, Calgary,
Alberta T2N 1N4, Canada

Received November 4, 1988

Summary. ART algorithms with relaxation parameters are studied for general (consistent or inconsistent) linear algebraic systems $Rx = f$, and a general convergence theorem is formulated. The advantage of severe underrelaxation is re-examined and clarified. The relationship to solutions obtained by applying SOR methods to the equation $RR^T y = f$ is investigated.

Subject classification. AMS(MOS) 65F10, 15A06, CR: G1.3

1 Introduction

In this paper we are concerned with iterative methods related to the ubiquitous equation $Rx = f$ where R is a given $m \times n$ real matrix (i.e. is in $\mathbb{R}^{m \times n}$) and f is a given real vector of length m (i.e. is in \mathbb{R}^m). More specifically, we are concerned with the cases when m and n are so large that storing the coefficients of R in a computer can be a problem, but the matrix has the property that the entries are easily generated as required. The equation $Rx = f$ may be consistent or inconsistent, and in the applications we have in mind is generally inconsistent.

Problems of this kind arise in a variety of applications that come under the broad heading of computerized tomography with incomplete data (see [DL], [He2], and [Na], for example), and have given rise to the development of a great variety of iterative algorithms based on updates of an approximate "solution" vector with a suitable linear combination of some of the rows (or just one row) of R . Prototypical among these algorithms is the Algebraic Reconstruction Technique (ART), which we briefly review together with some relaxation strategies.

Offprint requests to: P. Lancaster

* The work of this author was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada

Let $\mathbf{r}_1^T, \dots, \mathbf{r}_m^T$ (T denotes transposition) be the rows of R and $\mathbf{f}^T = [f_1, f_2, \dots, f_m]$. With a given starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$ we generate the sequence $\{\mathbf{x}^{(k)}\}_{k=0}^\infty$ by the relation

$$(1.1) \quad \mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \omega_k \|\mathbf{r}_k\|^{-2} (f_k - \mathbf{r}_k^T \mathbf{x}^{(k-1)}) \mathbf{r}_k,$$

$k = 1, 2, \dots$, where, when $k \equiv j \pmod{m}$ and $j = 1, 2, \dots, m$,

$$(1.2) \quad \mathbf{r}_k = \mathbf{r}_j, \quad f_k = f_j.$$

Note the scalar $f_k - \mathbf{r}_k^T \mathbf{x}^{(k-1)}$ appearing in (1.1) is just the residual error in the k th equation. The norm used in (1.1) (and elsewhere) is euclidean, and the real numbers $\{\omega_k\}_{k=0}^\infty$ are a sequence of relaxation parameters in the interval $(0, 2)$. With $\omega_k \equiv 1$ we obtain the classical ART algorithm (originating with Kaczmarz [K] in 1937). Algorithms of this kind can be obtained in several different ways, including minimization of residual vectors, filtering, gradient methods, or projection methods (see [He2] and [Tr1], for example). See also [EHL] for a treatment of these (and other) schemes.

There are also several variations that can be played on the theme of equations (1.1) and (1.2), including Richardson or SIRT methods ([Gi], [He2], [I]) and constrained systems ([He1] and [He2]). We confine attention to processes of type (1.1) and (1.2) and also related Successive Overrelaxation (SOR) methods for the system

$$(1.3) \quad RR^T \mathbf{y} = \mathbf{f}$$

(see [SB], and [Na], for example).

Excessive under-relaxation has been found to be beneficial in practice (see [HLL] and [He2]) and this has been explained to some extent by theory (see [CEG], and [Na]) when $\{\omega_k\}$ is a constant sequence. More general under-relaxation strategies have been investigated in [Tr1], [Tr2], and [B], in which ω_k depends only on the quotient obtained on division of k by m in (1.1). We are concerned with a cyclic choice of m parameters $\omega_1, \dots, \omega_m$. Thus, as in (1.2),

$$(1.4) \quad \omega_k = \omega_j \quad \text{when } k \equiv j \pmod{m},$$

and in the special case of a fixed relaxation parameter.

We first present a short and self-contained proof of the convergence of ART with our admissible relaxation strategy. The main new feature here is representation of the limit point(s) in terms of the residual vector $\mathbf{f} - R\mathbf{x}^1$ where $\mathbf{x}^1 = R^1 \mathbf{f}$, the "best-approximate" solution (in the l_2 -sense) and R^1 is the Moore-Penrose generalized inverse of R . This argument admits immediate generalization to a problem posed in Hilbert-space (see the Appendix). It demonstrates the robustness of these algorithms in the sense of independence from consistency of the original equation and the choice of initial vector $\mathbf{x}^{(0)}$. Our analysis also admits an improved explanation of the advantage to be gained by under-relaxation.

Our second major topic is the connection of ART algorithms with solutions of $R\mathbf{x} = \mathbf{f}$ obtained by applying SOR methods to the equation (1.3). In contrast with the algorithms of ART-type, it has been shown by O'Carroll [O'C] that the SOR algorithms actually diverge if the system $R\mathbf{x} = \mathbf{f}$ is inconsistent ($\mathbf{f} \notin \text{Im } R$). This suggests a serious disadvantage of the SOR strategy in the form of potential error accumulation, and is discussed in our formalism in Section 5.

Consider the $n \times n$ matrices Q_1, \dots, Q_m defined by

$$(1.5) \quad Q_j = I - \omega_j \|r_j\|^{-2} r_j r_j^T$$

and note that $\|Q_j\| = 1$ as long as $\omega_j \in (0, 2)$. It is easily seen that the “iteration matrix” for algorithms using equations (1.1), (1.2) and (1.4) is

$$(1.6) \quad A(\omega) = Q_m \dots Q_2 Q_1$$

where ω denotes the vector of relaxation parameters $\omega_1, \dots, \omega_m$. Thus $\|A(\omega)\| \leq 1$. Furthermore, if

$$\rho \in (\text{span}\{r_1, \dots, r_m\})^\perp = (\text{Im}(R^T))^\perp = \text{Ker } R,$$

then $A(\omega)\rho = \rho$, and we must generally admit that $1 \in \sigma(A(\omega))$, the spectrum of $A(\omega)$. It will be seen (and is well-known) that the speed of convergence depends on the quantity

$$(1.7) \quad \gamma(A) = \max\{|\lambda|; \lambda \in \{0\} \cup \sigma(A), \lambda \neq 1\}$$

(which is usually the magnitude of the sub-dominant eigenvalue of A). This quantity, $\gamma(A)$, will depend on ω as well as the ordering of the rows of R , and will also be investigated via the SOR connection.

In the paper by Smith, Solmon and Wagner [SSW], a bound is given on the norm of A restricted to the orthogonal complement of its spectral subspace corresponding to eigenvalue 1, say $c(A)$ (in the case $\omega_k \equiv 1$). Since $\gamma(A) \leq c(A) < 1$, the minimization of this bound can be used for finding certain suboptimal orderings of the rows of R (see [HS]). Here we wish to emphasize the importance of working with $\gamma(A)$, rather than its upper bound $c(A)$. A generalization of the bound $c(A)$ for products of paracontracting matrices (which include the matrices Q_j) is given by Nelson and Neumann [NN], and the quantity $c(A)$ is used by Natterer [Na] in a careful study of the effects of underrelaxation.

2 ART algorithms

We consider algorithms of the form (1.1), (1.2), (1.4) in which $x^{(0)}$ is chosen arbitrarily. Three lines of argument can be found in the literature. One seems to originate with Tanabe [Ta], and we give a short self-contained proof in that style (see also [CEG]). Our proof also admits generalization to a Hilbert-space setting, and we discuss this briefly in an Appendix to this paper. Another line of argument depends on a general theorem of Halperin [Ha] concerning the powers of a product of projection operators. This is used by Natterer [Na], for example, for the consistent problem with $x^{(0)} \in \text{Im } R^T$. The third approach uses the SOR connection and will be discussed below (see also [BE], [Na], for example).

For $s = 0, 1, 2, \dots$, write $\xi^{(s)} = x^{(sm)}$, the iterate obtained after s complete sweeps through the rows of R . It follows from (1.1) that

$$(2.1) \quad \xi^{(s+1)} = A(\omega)\xi^{(s)} + K(\omega)f$$

for $s = 0, 1, 2, \dots$, where $A(\omega)$ is defined by (1.5) and (1.6). Also

$$(2.2) \quad K(\omega) = \sum_{j=1}^m \omega_j \|r_j\|^{-2} Q_m Q_{m-1} \dots Q_{j+1} r_j e_j^T$$

and \mathbf{e}_j^\top denotes the j th unit coordinate vector in \mathbb{R}^m (and when $j = m$ we put $Q_m \cdots Q_{j+1} = \mathbf{I}$). It is easily verified that

$$(2.3) \quad K(\omega)R = \mathbf{I} - A(\omega).$$

Furthermore, if $\omega_1 = \omega_2 = \dots = \omega_m = \omega$ then as $\omega \rightarrow 0$

$$(2.4) \quad K(\omega) = \omega R^\top D + O(\omega^2)$$

where $D = \text{diag}[\|\mathbf{r}_1\|^{-2}, \dots, \|\mathbf{r}_m\|^{-2}]$.

Now let P be the orthogonal projector onto $\text{Im } R^\top$ along $\text{Ker } R$. Thus, $\mathbf{I} - P$ is onto $\text{Ker } R$ and along $\text{Im } R^\top$. The results of the first two lemmas are familiar (see [Ta], or [EHL]). The proof of the first is included for completeness and because it is short. The proof of the second is new and, being more technical, is relegated to the Appendix.

Lemma 1. *Im R^\top and Ker R are invariant under $A(\omega)$.*

Proof. Let P_j be the orthogonal projector onto \mathbf{r}_j then in (1.5) we have $Q_j = \mathbf{I} - \omega_j P_j, j = 1, 2, \dots, m$ and from (1.6) we deduce that if $\mathbf{x} \in \text{Im}(\mathbf{I} - A(\omega))$,

$$\mathbf{x} \in \sum_{j=1}^m \text{Im } P_j = \text{Im } R^\top.$$

Thus, $\text{Im}(\mathbf{I} - A) \subset \text{Im } R^\top$ and, since $\mathbf{I} - P$ annihilates $\text{Im } R^\top$, $(\mathbf{I} - P)(\mathbf{I} - A) = 0$, or

$$(\mathbf{I} - P)A(\omega) = \mathbf{I} - P.$$

But $R(\mathbf{I} - P) = 0$ and so equation (2.3) gives

$$(2.5) \quad A(\omega)(\mathbf{I} - P) = \mathbf{I} - P.$$

Hence $A(\omega)P = PA(\omega)$, as required. □

Lemma 2. *If $\omega_j \in (0, 2)$ for $j = 1, 2, \dots, m$ then $\|A(\omega)\| \leq 1$, and $\|A(\omega)P\| < 1$.*

This lemma requires a short technical proof that will be presented in the Appendix in a more general Hilbert space context. It is clear from the lemmas and equation (1.6) that the spectrum of the restriction of $A(\omega)$ to $\text{Im } P = \text{Im } R^\top$ is inside the open unit disc, while the restriction to $\text{Im}(\mathbf{I} - P) = \text{Ker } R$ has spectrum only at $\lambda = 1$. In fact, it is not difficult to show that $\text{Ker } R = \text{Ker}(\mathbf{I} - A)$ (see Corollary 4 of [Ta]).

We may now prove the convergence theorem. Recall that R^\dagger denotes the Moore–Penrose inverse of R , and we write $\mathbf{x}^\dagger = R^\dagger \mathbf{f}$. The corresponding residual vector is

$$(2.6) \quad \mathbf{g} \stackrel{\text{def}}{=} \mathbf{f} - R\mathbf{x}^\dagger = (\mathbf{I} - RR^\dagger)\mathbf{f},$$

and we note that, because RR^\dagger is the orthogonal projector onto $\text{Im } R$, $\mathbf{g} \in (\text{Im } R)^\perp = \text{Ker } R^\top$, whatever \mathbf{f} may be.

Theorem 1. *Let $\omega_j \in (0, 2)$ for $j = 1, 2, \dots, m$ and $\mathbf{x}^{(0)}$ be an arbitrary vector in \mathbb{R}^n . Then the sequence $\{\xi^{(s)}\}$ defined by (1.1) is convergent and*

$$(2.7) \quad \lim_{s \rightarrow \infty} \xi^{(s)} = \mathbf{x}^\dagger + (\mathbf{I} - P)\mathbf{x}^{(0)} + (\mathbf{I} - PA(\omega))^{-1}K(\omega)\mathbf{g}$$

where P is the orthogonal projector onto $\text{Im } R^T$ and \mathbf{g} is the residual vector, $\mathbf{g} = \mathbf{f} - R\mathbf{x}^1$.

Proof. It follows from (2.1) that

$$\xi^{(s+1)} = A^s(\omega)\mathbf{x}^{(0)} + (\mathbf{I} + A(\omega) + \dots + A^{s-1}(\omega))K(\omega)\mathbf{f}$$

and from (1.3) and (1.6)

$$\begin{aligned} \xi^{(s+1)} &= A^s(\omega)\mathbf{x}^{(0)} + (\mathbf{I} + A(\omega) + \dots + A^{s-1}(\omega))K(\omega)(R\mathbf{x}^1 + \mathbf{g}) \\ &= A^s(\omega)\mathbf{x}^{(0)} + (\mathbf{I} - A^s(\omega))\mathbf{x}^1 + (\mathbf{I} + A(\omega) + \dots + A^{s-1}(\omega))K(\omega)\mathbf{g}. \end{aligned}$$

In this equation put $A^s(\omega)\mathbf{x}^{(0)} = A^s(\omega)(P\mathbf{x}^{(0)} + (\mathbf{I} - P)\mathbf{x}^{(0)})$ and using (2.5) we obtain

$$A^s(\omega)\mathbf{x}^{(0)} = (A(\omega)P)^s\mathbf{x}^{(0)} + (\mathbf{I} - P)\mathbf{x}^{(0)}.$$

Also, $P\mathbf{x}^1 = \mathbf{x}^1$ since $\mathbf{x}^1 \in \text{Im } R^T$. Thus, $A^s(\omega)\mathbf{x}^1 = (A(\omega)P)^s\mathbf{x}^1$ and

$$\begin{aligned} \xi^{(s+1)} &= \mathbf{x}^1 + (A(\omega)P)^s(\mathbf{x}^{(0)} - \mathbf{x}^1) + (\mathbf{I} - P)\mathbf{x}^{(0)} \\ &\quad + (\mathbf{I} + A(\omega) + \dots + A^{s-1}(\omega))K(\omega)\mathbf{g}. \end{aligned}$$

Then we see that, for $j = 1, 2, \dots, m$, $PP_j = P_j$, so that $PQ_j = Q_jP$ and hence $PK(\omega) = K(\omega)$. It follows that

$$\begin{aligned} \xi^{(s+1)} &= \mathbf{x}^1 + (A(\omega)P)^s(\mathbf{x}^{(0)} - \mathbf{x}^1) + (\mathbf{I} - P)\mathbf{x}^{(0)} \\ &\quad + (\mathbf{I} - PA(\omega))^{-1}(\mathbf{I} - (PA(\omega))^s K)(\omega)\mathbf{g}. \end{aligned}$$

Now use Lemma 2 and take the limit as $s \rightarrow \infty$ to obtain (2.7). □

3 Discussion of the theorem

(a) In general, the algorithm of Theorem 1 is applied in order to find, or estimate \mathbf{x}^1 . First observe that by choosing $\mathbf{x}^{(0)} \in \text{Im } R^T$ (a linear combination of $\mathbf{r}_1, \dots, \mathbf{r}_m$) we have $(\mathbf{I} - P)\mathbf{x}^{(0)} = 0$ in equation (1.7) and the limit is therefore independent of $\mathbf{x}^{(0)}$. Now it is easily seen (and already well-known) that each subsequence of $\{\mathbf{x}^{(j)}\}_{j=0}^\infty$ obtained by taking indices j that are congruent mod (m) will be convergent. Thus, the ART algorithm of equations (1.1), (1.2) and (1.4) converges cyclically from any initial $\mathbf{x}^{(0)} \in \text{Im } R^T$, and the limits will depend on the ordering of the rows of R . Let Π denote the set of all permutations of indices $\{1, 2, \dots, m\}$ and for any $\pi \in \Pi$ let

$$b_\pi = \|(\mathbf{I} - PA_\pi(\omega))^{-1}PK_\pi(\omega)\|,$$

and $\{\xi_\pi^{(s)}\}$ denote the sequence generated as in (2.1) after applying π to the rows of R . It follows from (2.7) that when $\mathbf{x}^{(0)} \in \text{Im } R^T$ all cyclic limits of the ART algorithm will lie in the sphere with the centre \mathbf{x}^1 and the radius $b\|\mathbf{g}\|$, where $b = \max_{\pi \in \Pi} b_\pi$:

$$\left\| \lim_{s \rightarrow \infty} \xi_\pi^{(s)} - \mathbf{x}^1 \right\| \leq b\|\mathbf{g}\|, \quad \pi \in \Pi.$$

Thus, if $\|\mathbf{g}\|$ is small (the measurement errors do not drive \mathbf{f} too far from $\text{Im } R$), then even though the iterations $\mathbf{x}^{(k)}$ will not converge in the usual sense, they will ultimately oscillate in a small sphere with centre \mathbf{x}^1 .

Recall that the vector \mathbf{g} is the residual vector for the system $R\mathbf{x} = \mathbf{y}$ evaluated at \mathbf{x}^1 . Since RR^1 is the orthogonal projection onto $\text{Im } R$, it follows that $\mathbf{g} = (\mathbf{I} - RR^1)\mathbf{f}$ is orthogonal to $\text{Im } R$ and so provides a good measure of the departure of \mathbf{f} from $\text{Im } R$. It can be argued that this clustering of the limit points is one of the main reasons for the success of ART algorithms. In computerized tomography, even though measurement errors make the system inconsistent, a recognizable image can be produced provided the diameter of the cluster is not too large.

(b) It follows from the above discussion that if $\mathbf{f} \in \text{Im } R$ (the system is consistent) then for any $\pi \in \Pi$

$$\lim_{s \rightarrow \infty} \xi_\pi^{(s)} = \mathbf{x}^1,$$

that is $\{\mathbf{x}^{(k)}\}_{k=1}^\infty$ converges to \mathbf{x}^1 in the usual sense and independently of the ordering of rows of R . However, the *speed* of convergence will generally depend on the ordering of the rows as well as ω .

(c) If $\text{Im } R^T = \mathbb{R}^n$, then for any $\pi \in \Pi$ and for any $\xi_\pi^{(0)} \in \mathbb{R}^n$,

$$\lim_{s \rightarrow \infty} \xi_\pi^{(s)} = \mathbf{x}^1 + (K(\omega)R)^{-1} K(\omega)\mathbf{g}.$$

Indeed, in this case $P = \mathbf{I}$ and the result follows immediately from (2.3) and (2.7).

(d) Let ω be a constant sequence (and so replace ω by ω) and let us investigate the efficacy of underrelaxation. We need a lemma:

Lemma 3. *The rational matrix function $(\mathbf{I} - A(\omega)P)^{-1}$ has a simple pole at $\omega = 0$.*

Proof. As $\omega \rightarrow 0$ it is easily seen that

$$A(\omega)P = P - \omega A_1 + O(\omega^2)$$

where $A_1 = R^TDR$ (with D as in equation (2.4)). Thus, $(\mathbf{I} - A(\omega)P)^{-1}$ certainly has a pole at $\omega = 0$. Let the order be $k \geq 1$. Then there are matrices B_j , $j = -k, -k + 1, \dots, 0, 1, \dots$ such that

$$(3.1) \quad (B_{-k}\omega^{-k} + \dots + B_{-1}\omega^{-1} + B_0 + B_1\omega + \dots)(\mathbf{I} - P + \omega A_1 + \dots) = \mathbf{I}$$

for sufficiently small $|\omega| \neq 0$. If $k > 1$ then $B_{-k}(\mathbf{I} - P) = 0$ and $B_{-k}A_1 + B_{-k+1}(\mathbf{I} - P) = 0$. As $A_1P = A_1$ this implies $B_{-k}A_1 = 0$. But $\text{Im } A_1 = \text{Im } P$ and so $\text{Im } P \subset \text{Ker } B_{-k}$, or $B_{-k}P = 0$. Then $B_{-k} = B_{-k}P = 0$. Thus $k = 1$ and, in a deleted neighbourhood of $\omega = 0$, we have an expansion

$$(3.2) \quad (\mathbf{I} - A(\omega)P)^{-1} = B_{-1}\omega^{-1} + B_0 + B_1\omega + \dots$$

with $B_{-1} \neq 0$. □

Proposition. *Let $\mathbf{x}^{(0)} \in \text{Im } R^T$. Then as $\omega \rightarrow 0$*

$$(3.3) \quad \lim_{s \rightarrow \infty} \xi^{(s)} = \mathbf{x}^1 + B_{-1}R^TD\mathbf{g} + O(\omega),$$

and $B_{-1}R^TD\mathbf{g} = 0$ if and only if $D\mathbf{g} \in \text{Ker } R^T$.

Proof. It is easily seen from (3.1) that $-B_{-1}A_1 = P = -A_1B_{-1}$. In particular,

$$\text{Ker } B_{-1} \subset \text{Ker } P = \text{Ker } R = (\text{Im } R^T)^\perp.$$

Thus, $B_{-1}R^T\mathbf{x} = \mathbf{0}$ if and if $\mathbf{x} \in \text{Ker } R^T$.

Using (3.2) and (2.4) in (2.7) we obtain (3.3). Furthermore, $B_{-1}R^TD\mathbf{g} = \mathbf{0}$ if and only if $D\mathbf{g} \in \text{Ker } R^T$. \square

Natterer considers this limiting case when $R\mathbf{x} = \mathbf{f}$ is consistent, i.e. $\mathbf{f} \in \text{Im } R^T$ (see [Na]). With this hypothesis it follows from (2.6) that $\mathbf{g} = \mathbf{0}$ and so $B_{-1}R^TD\mathbf{g} = \mathbf{0}$ in (3.3). Now, in any case, $\mathbf{g} \in \text{Ker } R^T$ so $B_{-1}R^TD\mathbf{g} = \mathbf{0}$ provided $D = I$. In other words, when the rows of R are normalized to have unit length. This is the case considered by Censor, Eggermont, and Gordon [CEG].

Now it is easily seen that, if the starting vector $\mathbf{x}^{(0)}$ is fixed, the sequence $\{\mathbf{x}^{(j)}\}_{j=0}^\infty$, and hence the subsequence $\{\mathbf{z}^{(s)}\}$, are invariant under row-scaling of R . Thus the *image* produced is independent of the scaling and (3.3) implies that

$$\mathbf{x}^1 + B_{-1}R^TD\mathbf{g} = \mathbf{x}_D^1,$$

the best approximate solution after row-normalization, which is approached in the limit, with or without normalization. Note that $\mathbf{x}_D^1 = (D^{1/2}R)^1D^{1/2}\mathbf{g}$ and, in general, $\mathbf{x}_D^1 \neq \mathbf{x}^1$ because $(D^{1/2}R)^1 \neq R^1D^{-1/2}$. As one might expect, the best approximate solution of $R\mathbf{x} = \mathbf{f}$ depends on the relative sizes of the residuals $\mathbf{r}_j^T\mathbf{x} - f_j$, $j = 1, 2, \dots, m$. (The effects of row normalization on SIRT algorithms have been discussed by van der Sluis and van der Vorst, [VV].)

These results show that, when $\|\mathbf{g}\|$ is small in an appropriate sense, the cluster of limit points described in item (a) above will have a diameter that decreases to zero as $\omega \rightarrow 0$. Indeed, all limit points converge to \mathbf{x}_D^1 as $\omega \rightarrow 0$.

(e) For any given ordering of the rows of R the speed of convergence of the ART algorithm will depend on the choice of the relaxation parameters $\omega_j \in (0, 2)$. The optimal ω minimizes the spectral radius of the restriction of $A(\omega)$ to $\text{Im } R^T$, $\gamma(A(\omega))$. As we have noted, it is not sufficient to minimize the norm of this restriction, $c(A(\omega))$, which provides only an upper bound for $\gamma(A(\omega))$. This will be demonstrated explicitly for the simplest nontrivial example in Section 6, i.e. when $m = 2$, a case that we introduce here. Using a single parameter, ω , we show that in this case $c(A(\omega)) \geq c(A(1))$ when $\omega \in (0, 2)$; a property that is not shared by the function $\gamma(A(\omega))$.

Example. Let $m = 2$ and

$$A(\omega) = (I - \omega\mathbf{r}_2\mathbf{r}_2^T)(I - \omega\mathbf{r}_1\mathbf{r}_1^T),$$

where $\|\mathbf{r}_1\| = \|\mathbf{r}_2\| = 1$. It is known ([GV], §12.4) that the singular values of $A(1)$ are equal to $1, \dots, 1, |\mathbf{r}_1^T\mathbf{r}_2|, 0$. Since the spectrum of $A(1)$ coincides with the spectra of

$$(I - \mathbf{r}_2\mathbf{r}_2^T)(I - \mathbf{r}_1\mathbf{r}_1^T)(I - \mathbf{r}_2\mathbf{r}_2^T) = (I - \mathbf{r}_2\mathbf{r}_2^T)(I - \mathbf{r}_1\mathbf{r}_1^T)(I - \mathbf{r}_1\mathbf{r}_1^T)(I - \mathbf{r}_2\mathbf{r}_2^T) = AA^T,$$

the eigenvalues of A are equal to $1, \dots, 1, (\mathbf{r}_1^T\mathbf{r}_2)^2, 0$. Thus $c(A(1)) = |\mathbf{r}_1^T\mathbf{r}_2|$, and $\gamma(A(1)) = (\mathbf{r}_1^T\mathbf{r}_2)^2$.

Using the minimax characterization of singular values ([GV], §8.3), it can be shown that $c(A(\omega)) \geq c(A(1))$ for any $\omega \in (0, 2)$. In contrast, it will be seen in Section 5 that, unless $\mathbf{r}_1 = \mathbf{r}_2$, $\gamma(A(\omega))$ is minimized when $\omega \neq 1$.

4 An equivalent characterization of $\gamma(A(\omega))$

In this section a second characterization for $\gamma(A(\omega))$ is derived, where $A(\omega)$ is defined by equations (1.5) and (1.6) with $\omega_1 = \dots = \omega_m = \omega$. Also, we assume that the rows of R are normalized, so that

$$A(\omega) = (\mathbf{I} - \omega \mathbf{r}_m \mathbf{r}_m^T) \dots (\mathbf{I} - \omega \mathbf{r}_1 \mathbf{r}_1^T).$$

Since $\gamma(A(\omega)) = \gamma(A^T(\omega))$ we consider for convenience $A^T(\omega)$. Clearly, there are (possibly non-unique) numbers β_{ij} , $1 \leq i, j \leq m$, such that

$$(4.1) \quad \mathbf{r}_i^T A^T(\omega) = \mathbf{r}_i^T + \beta_{i1} \mathbf{r}_1^T + \dots + \beta_{im} \mathbf{r}_m^T.$$

It is straightforward to check that β_{ij} defined by the following recursive relation:

$$\begin{aligned} \beta_{i1} &= -\omega c_{i1} \\ \beta_{i2} &= -\omega [c_{i2} + \beta_{i1} c_{12}] \\ &\vdots \\ \beta_{im} &= -\omega [c_{im} + \beta_{i1} c_{1m} + \dots + \beta_{i, m-1} c_{m-1, m}] \end{aligned}$$

where $c_{ij} = \mathbf{r}_i^T \mathbf{r}_j = c_{ji}$ satisfy (4.1). Denote $\mathcal{B}(\omega) = [\beta_{ij}]_{i,j=1}^m$ and write the Gram matrix $RR^T = \mathbf{I} - L - L^T$, where

$$L = - \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ c_{12} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ c_{1m} & c_{2m} & \dots & c_{m-1, m} & 0 \end{bmatrix}.$$

Then it follows from the recursive relations for β_{ij} that

$$(4.2) \quad \mathcal{B}(\omega) = -\omega(RR^T - \mathcal{B}(\omega)L^T).$$

Hence

$$(4.3) \quad \mathcal{B}(\omega) = -\omega RR^T (\mathbf{I} - \omega L^T)^{-1}.$$

Adding \mathbf{I} to both sides we obtain the triangular factorization

$$(4.4) \quad \mathbf{I} + \mathcal{B}(\omega) = \{(1 - \omega)\mathbf{I} + \omega L\} (\mathbf{I} - \omega L^T)^{-1}.$$

Given the importance of the function $\gamma(A(\omega))$, the significance of the matrix $\mathcal{B}(\omega)$ for our analysis is apparent from the next theorem.

Theorem 2. For any number $\lambda \neq 1$, $\lambda \in \sigma(A(\omega))$ if and only if $\lambda \in \sigma(\mathbf{I} + \mathcal{B}(\omega))$.

Proof. The relation (4.1) can be written in the form

$$(4.5) \quad RA^T(\omega) = (\mathbf{I} + \mathcal{B}(\omega))R.$$

Suppose $\lambda \neq 1$ and $\lambda \in \sigma(A(\omega))$. Then there is an $\mathbf{x} \neq \mathbf{0}$ such that $A^T(\omega)\mathbf{x} = \lambda\mathbf{x}$. It follows that $R\mathbf{x} \neq \mathbf{0}$, for otherwise $\mathbf{r}_i^T \mathbf{x} = 0$ for each i , and hence $A(\omega)\mathbf{x} = \mathbf{x}$,

contrary to hypothesis. Thus, it follows from (4.5) that

$$(\mathbf{I} + \mathcal{B}(\omega))(\mathbf{R}\mathbf{x}) = \lambda(\mathbf{R}\mathbf{x})$$

and $\lambda \in \sigma(\mathbf{I} + \mathcal{B}(\omega))$.

Conversely, let $\lambda \in \sigma(\mathbf{I} + \mathcal{B}(\omega))$, $\lambda \neq 1$. Then there is a $\mathbf{y} \neq \mathbf{0}$ such that $(\mathbf{I} + \mathcal{B}^T(\omega))\mathbf{y} = \lambda\mathbf{y}$. Clearly $\mathcal{B}^T(\omega)\mathbf{y} \neq \mathbf{0}$ or we contradict $\lambda \neq 1$. It follows from equation (4.3) that also $\mathbf{R}^T\mathbf{y} \neq \mathbf{0}$, and from (4.5) we obtain

$$A(\mathbf{R}^T\mathbf{y}) = \lambda(\mathbf{R}^T\mathbf{y}) . \quad \square$$

Note, in particular, that the theorem implies $\gamma(A(\omega)) = \gamma(\mathbf{I} + \mathcal{B}(\omega))$.

Remarks

1. $1 \notin \sigma(\mathbf{I} + \mathcal{B}(\omega))$ if and only if $\mathbf{r}_1, \dots, \mathbf{r}_m$ are linearly independent. Indeed, $(\mathbf{I} + \mathcal{B}^T(\omega))\mathbf{x} = \mathbf{x}$ implies that $\mathcal{B}^T(\omega)\mathbf{x} = \mathbf{0}$ and by (4.3) $\mathbf{R}\mathbf{R}^T\mathbf{x} = \mathbf{0}$. Thus $\mathbf{R}^T\mathbf{x} = \mathbf{0}$. On the other hand if $\mathbf{R}^T\mathbf{x} = \mathbf{0}$ then $\mathcal{B}^T(\omega)\mathbf{x} = \mathbf{0}$ and $(\mathbf{I} + \mathcal{B}^T(\omega))\mathbf{x} = \mathbf{x}$.
2. $1 \notin \sigma(A(\omega))$ if and only if $\text{Im } \mathbf{R}^T = \mathbb{R}^n$. This is obvious.
3. In case $\omega = 1$, $\mathbf{I} + \mathcal{B}^T(1) = (\mathbf{I} - L)^{-1}L^T$.

5 Relaxation with a sequence of parameters

Now let us return to the possibility of choosing m relaxation parameters in cyclic order. Thus, with $\omega_j \in (0, 2)$, $j = 1, 2, \dots, m$,

$$A^T(\omega) = (\mathbf{I} - \omega_1 \mathbf{r}_1 \mathbf{r}_1^T) \dots (\mathbf{I} - \omega_m \mathbf{r}_m \mathbf{r}_m^T) .$$

Repeating the argument of Theorem 2 it is easy to see that except (possibly) for the number 1, $A(\omega_1, \dots, \omega_m)$ and $\mathbf{I} + \mathcal{B}(\omega_1, \dots, \omega_m)$ have the same eigenvalues, and (cf. equation (4.4)),

$$\mathbf{I} + \mathcal{B}(\omega_1, \dots, \omega_m) = ((\mathbf{I} - W) + LW)(\mathbf{I} - L^T W)^{-1}$$

where $W = \text{diag}[\omega_1, \dots, \omega_m]$.

In case $\omega_1 = 1$ and $\omega_2, \dots, \omega_m$ are arbitrary numbers in the interval $(0, 2)$, the first row of the matrix $\mathbf{R}A^T$ is zero. Indeed,

$$\mathbf{r}_1^T(\mathbf{I} - \mathbf{r}_1 \mathbf{r}_1^T) \dots (\mathbf{I} - \omega_m \mathbf{r}_m \mathbf{r}_m^T) = \mathbf{0} .$$

Rewriting the equality (3.4) for submatrices with indices $i, j = 2, \dots, m$, $\mathcal{B}'(\omega_2, \dots, \omega_m) = \{\beta_{ij}\}_{i,j=2}^m$, $W' = \text{diag}[\omega_2, \dots, \omega_m]$ and

$$L' = \begin{bmatrix} 0 & 0 & \dots & 0 \\ c_{23} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ c_{2m} & c_{3m} & \dots & 0 \end{bmatrix}$$

we have

$$\mathbf{I} + \mathcal{B}'(\omega_2, \dots, \omega_m) = (\mathbf{I} - W' + L'W' + \mathbf{c}\mathbf{c}^T W')(\mathbf{I} - (L')^T W')^{-1}$$

where $\mathbf{c}^T = [c_{12}, \dots, c_{1m}]$. It follows now from (4.5), as in Theorem 2, that

$$\sigma(A(1, \omega_2, \dots, \omega_m)) \setminus \{1\} = \sigma(\mathbf{I}_{m-1, m-1} + \mathcal{B}'(\omega_2, \dots, \omega_m)) \setminus \{1\} .$$

This reduction by one dimension can be useful for analyzing cases with small m . For example for $m = 2$

$$A(1, \omega) = (\mathbf{I} - \omega \mathbf{r}_2 \mathbf{r}_2^T)(\mathbf{I} - \mathbf{r}_1 \mathbf{r}_1^T) .$$

Thus $\mathbf{I} + \mathcal{B}'(\omega) = \mathbf{I} - \omega + c_{12}^2 \omega$ and therefore $\gamma(A(1, \omega)) = 1 - \omega + c_{12}^2 \omega$. Clearly for $\omega = 1, \gamma(A) = c_{12}^2$. In case the angle between \mathbf{r}_1 and \mathbf{r}_2 is larger than $\pi/4$, one can take $\omega = 1/(1 - c_{12}^2) \in (0, 2)$ such that $\gamma(A(1, 1/(1 - c_{12}^2))) = 0$. If the angle is less than or equal to $\pi/4$, then $\gamma(A(1, \omega))$ we decrease as ω approaches 2.

6 Relation to the SOR algorithm

It follows from equation (4.4) that, in the case of a single relaxation parameter and with row normalization, the matrix $\mathbf{I} + \mathcal{B}^T(\omega)$ is the iteration of the SOR algorithm for the equation $R R^T \mathbf{y} = \mathbf{f}$ (see [SB], p. 546, for example). Therefore (see also equation (2.4) of Nicolaides, [Ni])

$$(6.1) \quad \mathbf{I} + \mathcal{B}^T(\omega) = (\mathbf{I} - \omega \mathbf{e}_m \mathbf{r}_m^T R^T) \dots (\mathbf{I} - \omega \mathbf{e}_1 \mathbf{r}_1^T R^T).$$

We observe that $A(\omega)$ is expressed in terms of the orthogonal projectors $\mathbf{r}_k \mathbf{r}_k^T$ and $\mathbf{I} + \mathcal{B}^T(\omega)$ is expressed in a similar way in terms of the (generally) non-orthogonal projectors $\mathbf{e}_k \mathbf{r}_k^T R^T$. For brevity, let us denote $\mathbf{I} + \mathcal{B}^T(\omega)$ by $T(\omega)$. We conclude from Theorem 2 that

$$(6.2) \quad \sigma(A(\omega)) \setminus \{1\} = \sigma(T(\omega)) \setminus \{1\}.$$

Let us show that the matrix $T(\omega)$ is in fact an iteration matrix for a certain iterative process of ART type in \mathbb{R}^m . (This is also pointed out by Björck and Elving [BE]). If $\mathbf{x}^{(0)} \in \text{Im } R^T$ it follows from (1.1) that $\mathbf{x}^{(k)} \in \text{Im } R^T$ for all k . Hence there exist $\mathbf{y}^{(k)} \in \mathbb{R}^m$ (in general not unique), such that

$$\mathbf{x}^{(k)} = R^T \mathbf{y}^k, \quad k = 0, 1, 2, \dots$$

Suppose we start with some $\mathbf{y}^{(0)} \in \mathbb{R}^m$ and define iterations

$$(6.3) \quad \mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} + \alpha_k \mathbf{e}_{k+1}, \quad k = 0, 1, 2, \dots$$

where \mathbf{e}_k denotes the k -th coordinate vector in \mathbb{R}^m and

$$(6.4) \quad \alpha_k = \omega (f_{k+1} - \mathbf{r}_{k+1}^T R^T \mathbf{y}^{(k)}) .$$

Then clearly, for $k = 0, 1, 2, \dots$,

$$R^T \mathbf{y}^{(k)} = R^T \mathbf{y}^{(k+1)} + \omega (f_{k+1} - \mathbf{r}_{k+1}^T R^T \mathbf{y}^{(k)}) \mathbf{r}_{k+1} .$$

Denoting $\mathbf{x}^{(k)} = R^T \mathbf{y}^{(k)}$ we get

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \omega (f_{k+1} - \mathbf{r}_{k+1}^T \mathbf{x}^{(k)}) \mathbf{r}_{k+1}$$

or

$$\mathbf{x}^{(k+1)} = (\mathbf{I} - \omega \mathbf{r}_{k+1} \mathbf{r}_{k+1}^T) \mathbf{x}^{(k)} + \omega f_{k+1} \mathbf{r}_{k+1}$$

for $k = 0, 1, 2, \dots$, which is essentially the same as equation (1.1) with ω fixed.

Thus, the process (6.3)–(6.4) can be considered as an alternative method of solution of $R \mathbf{x} = \mathbf{f}$. The extra work needed to compute $R^T \mathbf{y}^{(k)}$ can be justified when

R^T is sparse, because the updates in (6.3) are in one position only. (Note that the vector $x^{(k)} = R^T y^{(k)}$ is computed in the process of finding α_k .)

Denoting $\eta^{(s)} = y^{(sn)}$, $s = 0, 1, 2, \dots$ we get

$$(6.5) \quad \eta^{(s+1)} = T(\omega)\eta^{(s)} + \tilde{K}(\omega)\mathbf{f}$$

where $\tilde{K}(\omega)$ is an $m \times m$ matrix dependent on ω . It is known from the SOR theory (p. 546 of [SB]) that (recalling $\|\mathbf{r}_i\| = 1$),

$$(6.6) \quad \tilde{K}(\omega) = \omega(\mathbf{I} - \omega L)^{-1}.$$

In contrast to the Kaczmarz algorithm of Theorem 1, the next result shows that the SOR procedure converges if and only if the original system $R\mathbf{x} = \mathbf{f}$ (or the system $RR^T\mathbf{y} = \mathbf{f}$) is consistent. The conclusion originates with O'Carroll [O'C], but it is re-formulated here to express the limiting vector in more explicit form consistent with the presentation of Theorem 1.

Theorem 3. *Let $\omega \in (0, 2)$ and $\eta^{(0)}$ be any vector in \mathbb{R}^m . Then the iteration (6.5) converges if and only if $\mathbf{f} \in \text{Im } R$, and in this case*

$$(6.7) \quad \lim_{s \rightarrow \infty} \eta^{(s)} = (RR^T)^{\dagger}\mathbf{f} + Q(\omega)\eta^{(0)}.$$

Remarks

(a) In case $\mathbf{f} \notin \text{Im}(RR^T)$ the iterations $\eta^{(s)}$ will diverge. However their projections onto $\text{Im}(\mathbf{I} - Q)$ will converge. Indeed, writing $\mathbf{f} = (RR^T)\mathbf{y}^{\dagger} + \mathbf{h}$, it is found that

$$\lim_{s \rightarrow \infty} (\mathbf{I} - Q)\eta^{(s)} = \mathbf{y}^{\dagger} + (\mathbf{I} - (\mathbf{I} - Q)T(\omega))^{-1}(\mathbf{I} - Q)\tilde{K}(\omega)\mathbf{h}.$$

This implies cyclic convergence for the sequence $\{\mathbf{y}^{(s)}\}_{s=1}^{\infty}$ which depends on the ordering of the rows of R . Clearly

$$R^T \lim_{s \rightarrow \infty} \eta^{(s)} = \mathbf{x}^{\dagger} + (\mathbf{I} - PA(\omega)^{-1}PK(\omega))\mathbf{g}$$

where P and \mathbf{g} are defined in Theorem 1. It is possible to show that if S is any other matrix satisfying the relations $S(\mathbf{I} - Q) = 0$ and $R^T S = R^T$, then $S\eta^{(s)}$ will converge and

$$\lim_{s \rightarrow \infty} S\eta^{(s)} = S\mathbf{y}^{\dagger} + S(\mathbf{I} - (\mathbf{I} - Q)T(\omega))^{-1}(\mathbf{I} - Q)\tilde{K}(\omega)\tilde{\mathbf{h}}.$$

(b) Since $R^T(\mathbf{I} - \omega\mathbf{e}_k\mathbf{r}_k^T R^T) = (\mathbf{I} - \omega r_k r_k^T)R_k^T$ it follows that

$$R^T T(\omega) = A(\omega)R^T.$$

Hence $R^T \eta^{(s+1)} = R^T T(\omega)\eta^{(s)} + R^T \tilde{K}(\omega)\mathbf{f}$, or

$$\xi^{(s+1)} = A(\omega)\xi^{(s)} + R^T \tilde{K}(\omega)\mathbf{f}.$$

Comparing with (2.1) for any $\mathbf{f} \in \mathbb{R}^m$ we get (with normalized rows),

$$K(\omega) = R^T \tilde{K}(\omega) = \omega R^T(\mathbf{I} - \omega L)^{-1}.$$

(c) To illustrate the use of the connection between ART and SOR algorithms we shall find the ω which minimizes $\gamma(A(\omega))$ when $m = 2$. Thus

$$A(\omega) = (\mathbf{I} - \omega\mathbf{r}_2\mathbf{r}_2^T)(\mathbf{I} - \omega\mathbf{r}_1\mathbf{r}_1^T).$$

Let $P_k = R_k R_k^T$ denote the orthogonal projections onto $\text{Im } R_k$ for $k = 1, \dots, M$. Then the ART iteration matrix $A(\omega)$ for R can be written as follows:

$$A(\omega) = (I - \omega P_M) \dots (I - \omega P_1)$$

where P_k have the property that

$$(6.9) \quad \begin{aligned} P_1 \perp P_3, \dots, P_M, \\ P_2 \perp P_4, \dots, P_M, \\ \vdots \\ P_{M-2} \perp P_M. \end{aligned}$$

Thus if R has the block structure (6.8) such that (6.9) holds, we get

$$(6.10) \quad \omega_{\text{opt}} = 2 / (1 - \sqrt{1 - \gamma^2(A(1))})$$

This observation suggests a possible strategy of ordering and partially orthogonalizing rows of R in such a way that the optimal ω could be determined via (6.10). Starting with some orthogonal rows R_1 find all rows which are not orthogonal to R_1 and orthogonalize them. Then find all rows which are not orthogonal to all previously chosen rows and orthogonalize them. Continue until all rows of R are used.

7 Appendix

Convergence of the ART-algorithm in Hilbert space

Let $\mathcal{H}, \mathcal{H}_1, \dots, \mathcal{H}_m$ be Hilbert spaces and R_j be a bounded linear operator from \mathcal{H} to \mathcal{H}_j for $j = 1, 2, \dots, m$. Assume that each R_j is a transformation with closed range. Define

$$R = \begin{bmatrix} R_1 \\ \vdots \\ R_m \end{bmatrix} : \mathcal{H} \rightarrow \mathcal{H}_1 \times \mathcal{H}_2 \times \dots \times \mathcal{H}_m,$$

and R is seen to have closed range as well. Under these hypotheses, R_1, \dots, R_m , and R each have a Moore-Penrose inverse (see [Gr]), denoted by $R_1^{\dagger}, R^{\dagger}$, etc.. Furthermore, if we define $P_j = R_j^{\dagger} R_j$, P_j is the orthogonal projector onto $\text{Im } R_j^{\dagger} = \text{Im } R_j^{\dagger} = (\text{Ker } R_j)^{\perp}$. Given an

$$f = (f_1, f_2, \dots, f_m) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \dots \times \mathcal{H}_m,$$

the equation $Rx = f$ obviously generalizes the problem of the main text (where R_j corresponds to r_j^T , etc.). Note that the problem considered here, with m fixed transformations R_1, \dots, R_m is different from that considered in the Hilbert-space analysis of McCormick [M].

We remark that this problem is not posed just for the sake of generalization. The full Hilbert-space machinery is required for analysis of the tomography problem with m complete projections (see section V.4.3 of [Na]).

In this context the ART algorithm (1.1), (1.2), (1.4) takes the form: Given any $x^{(0)} \in \mathcal{H}$, generate $\{x^{(k)}\}$ by

$$\begin{aligned} x^{(k)} &= x^{(k-1)} + \omega_k R_k^1 (f_k - R_k x^{(k-1)}), \\ &= (I - \omega_k P_k) x^{(k-1)} + \omega_k R_k^1 f_k \end{aligned}$$

where, for $k \equiv j \pmod{m}$, $j = 0, 1, \dots, m - 1$,

$$R_k = R_{j+1}, \quad f_k = f_{j+1}, \quad \omega_k = \omega_{j+1}.$$

Writing $\xi^{(s)} = x^{(sm)}$, $s = 0, 1, 2, \dots$ we obtain

$$(A.1) \quad \xi^{(s+1)} = A(\omega) \xi^{(s)} + K(\omega) f$$

(as in equation (2.1)) where now

$$A(\omega) = (I - \omega_m P_m) \dots (I + \omega_2 P_2) (I - \omega_1 P_1)$$

and (as in (2.3)),

$$K(\omega)R = I - A(\omega).$$

Now the analysis proceeds in much the same way as the finite-dimensional case. We present the proof of one lemma that takes some care and then conclude the Appendix with a statement of the generalized form of Theorem 1. Note first that we now define P to be the orthogonal projector onto $\text{Im } R^*$, and that the conclusion of Lemma 1 of the main text carries over immediately to this setting.

Lemma 2A. *If $\omega_j \in (0, 2)$ for $j = 1, 2, \dots, m$ then $\|A(\omega)\| \leq 1$ and $\|A(\omega)P\| < 1$.*

Proof. For any projector P (that is not 0 or I), the spectrum is given by $\sigma(P) = \{0, 1\}$. It follows that $\sigma(I - \omega_j P_j) = \{1 - \omega_j, 1\}$ and, as $I - \omega_j P_j$ is normal, $\|I - \omega_j P_j\| = 1$ if $\omega_j \in (0, 2)$. It follows that $\|A(\omega)\| \leq 1$. Consequently, (see Lemma 3.1 of [Na]) we have the orthogonal decomposition $\mathcal{H} = \text{Ker}(I - A) \oplus \text{Im}(I - A)$, $I - A$ is one-to-one on its image, and $\text{Ker}(I - A) = \text{Ker}(I - A^*)$. Since $\text{Ker}(I - A) = \text{Ker } R$ it follows that $\text{Im}(I - A) = \text{Im } R^*$.

Now assume $\|A(\omega)P\| = 1$ and look for a contradiction. Let

$$k^2 = 2 \max_{1 \leq j \leq m} (2\omega_j - \omega_j^2)^{-1}.$$

Give $\varepsilon > 0$ there is an $a_\varepsilon \in \mathcal{H}$ with $\|a_\varepsilon\| = 1$ and

$$\|APa_\varepsilon\| = \|PAa_\varepsilon\| > 1 - \varepsilon^2.$$

Thus for any j , $1 \leq j \leq m$,

$$(1 - \varepsilon^2) < \|PAa_\varepsilon\| \leq \|(I - \omega_j P_j)a_\varepsilon\| \leq 1.$$

But $\|(I - \omega_j P_j)a_\varepsilon\|^2 = \|a_\varepsilon\|^2 - (2\omega_j - \omega_j^2) \|P_j a_\varepsilon\|^2$, so

$$(1 - \varepsilon^2)^2 < 1 - (2\omega_j - \omega_j^2) \|P_j a_\varepsilon\|^2 \leq 1$$

and

$$\varepsilon^2 > \varepsilon^2 - \frac{1}{2}\varepsilon^4 > \frac{1}{2}(2\omega_j - \omega_j^2) \|P_j a_\varepsilon\|^2 \geq 0.$$

Thus,

$$\|P_j a_\varepsilon\| < \left(\frac{2}{2\omega_j - \omega_j^2} \right)^{1/2} \varepsilon < k\varepsilon$$

and there is a sequence $\{a_n\}$ such that $\|a_n\| = 1$, $\|PAa_n\| \rightarrow 1$ and $P_j a_n \rightarrow 0$ as $n \rightarrow \infty$. Since

$$PAa_n = P(I - \omega_m P_m) \dots (I - \omega_2 P_2)(I - \omega_1 P_1)a_n$$

it follows that $PAa_n = Pa_n + \sum_{j=1}^m K_j P_j a_n$ for certain bounded linear operators K_1, \dots, K_m , and hence

$$\lim_{n \rightarrow \infty} (I - A)Pa_n = 0.$$

Since $\{Pa_n\}$ is in $\text{Im } R^* = \text{Im}(I - A)$, and $I - A$ is one-to-one on its image, it follows that $Pa_n \rightarrow 0$ with $\|a_n\| = 1$ and $\|APA_n\| \rightarrow 1$ as $n \rightarrow \infty$. This contradicts the continuity of A . \square

Theorem 1A. *Let $\omega_j \in (0, 2)$ for $j = 1, 2, \dots, m$ and $x^{(0)} \in \mathcal{H}$. Then the sequence $\{\xi^{(s)}\}$ ($\xi^{(0)} = x^{(0)}$) defined by (A.1) is convergent and*

$$\lim_{s \rightarrow \infty} \xi^{(s)} = R^1 f + (I - P)x^{(0)} + (I - PA(\omega))^{-1} K(\omega)g$$

where P is the orthogonal projector onto $\text{Im } R^*$ and g is the residual vector: $g = f - R(R^1 f)$.

Acknowledgement. The authors thank an anonymous referee for helpful comments and criticism.

References

- [B] Baumeister, J.: On the adaptive solution of inconsistent systems of linear equations. Proc. of the 3rd seminar on "Model Optimization Exploration Geophysics", Berlin 1985
- [BE] Björck, A., Elving, T.: Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations. BIT 19, 145–163 (1979)
- [CEG] Censor, Y., Eggermont, P.P.B., Gordon, D.: Strong underrelaxation in Kaczmarz's method for inconsistent systems. Numer. Math. 41, 83–92 (1983)
- [DL] Dines, K.A., Little, R.J.: Computerized geophysical tomography. Proc. I.E.E.E. 67, 1065–1073 (1979)
- [EHL] Eggermont, P.P.B., Herman, G.T., Lent, A.: Iterative algorithms for large partitioned linear systems, with applications to image reconstruction. Linear Alg. Appl. 40, 57–67 (1981)
- [Gi] Gilbert, P.: Iterative methods for the three-dimensional reconstruction of an object from projections. J. Theor. Bio. 36, 105–117 (1972)
- [GV] Golub, G., Van Loan, C.F.: Matrix computations. Johns Hopkins Univ. Press. 1983
- [Gr] Groetsch, C.W.: Generalized Inverses of Linear Operator. New York: Marcel Dekker 1977
- [Ha] Halperin, I.: The product of projection operators. Acta Sci. Math. (Szeged) 23, 96–99 (1962)
- [HS] Hamaker, C., Solmon, D.C.: The angles between the null spaces of X-rays. J. Math. Anal. 62, 1–23 (1978)
- [He1] Herman, G.T.: A relaxation method for reconstructing objects from noisy X-rays. Math. Programming 8, 1–19 (1975)
- [He2] Herman, G.T.: Image reconstruction from projections. New York: Academic Press 1980
- [HLL] Herman, G.T., Lent, A., Lutz, P.H.: Relaxation methods for image reconstruction. Comm. of the A.C.M. 21, 152–158 (1978)
- [I] Ivansson, S.: Seismic borehole tomography—theory and computational methods. Proc. I.E.E.E. 74, 328–338 (1986)

- [K] Kaczmarz, S.: Angenaherte Aufloesung von Systemen linear Gleichungen. Bull. Int. Acad. Pol. Sci. Lett. A. 355–357 (1937)
- [M] McCormick, S.F.: The method of Kaczmarz and row orthogonalization for solving linear equations and least squares problems in Hilbert space. Indiana Univ. Math. J. **26**, 1137–1150 (1977)
- [NaS] Nashed, M.Z.: Continuous and semicontinuous analogues of iterative methods of Cimmino and Kaczmarz with applications to the inverse Radon transform. In: Herman, G.T., Natterer, F. (eds) Mathematical aspects of computerized tomography. Berlin Heidelberg New York: Springer 1981
- [Na] Natterer, F.: The mathematics of computerized tomography. New York: Wiley 1986
- [NN] Nelson, S., Neumann, M.: Generalization of the projection method with applications to SOR method for hermitian semidefinite linear systems. Numer. Math. **51**, 123–141 (1987)
- [Ni] Nicolaides, R.A.: On a geometrical aspect of SOR and the theory of consistent ordering for positive definite matrices. Numer. Math. **23**, 99–104 (1974)
- [O’C] O’Carroll, M.J.: Inconsistencies and SOR convergence for the discrete Neumann problem. J. Inst. Math. Applic. **11**, 343–350 (1973)
- [SSW] Smith, K.T., Solmon, D.C., Wagner, S.L.: Practical and mathematical aspects of the problem of reconstructing a function from radiographs. Bull. Amer. Math. Soc. **83**, 1227–1270 (1977)
- [SB] Stoer, J., Bulirsch, R.: Introduction to numerical analysis. New York Heidelberg Berlin: Springer 1976
- [Ta] Tanabe, K.: Projection method for solving a singular system of linear equations and its applications. Numer. Math. **17**, 203–214 (1971)
- [Tr1] Trummer, M.R.: Reconstruction pictures from projections: On the convergence of the ART algorithm with relaxation. Computing **26**, 189–195 (1981)
- [Tr2] Trummer, M.R.: A note on the ART of relaxation. Computing **33**, 349–352 (1984)
- [VV] van der Sluis, A., van der Vorst, H.A.: Numerical solution of large, sparse linear algebraic systems arising from tomographic problems. In: Nolet, G. (ed.) Seismic tomography, pp. 49–83. Dordrecht: Reidel 1987
- [Y] Yosida, K.: Functional analysis. New York Heidelberg Berlin: Springer 1968