Electronic Transactions on Numerical Analysis.
Volume 37, pp. 321-336, 2010.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

SEMI-CONVERGENCE AND RELAXATION PARAMETERS FOR A CLASS OF SIRT ALGORITHMS*

TOMMY ELFVING†, TOURAJ NIKAZAD‡, AND PER CHRISTIAN HANSEN§

Abstract. This paper is concerned with the Simultaneous Iterative Reconstruction Technique (SIRT) class of iterative methods for solving inverse problems. Based on a careful analysis of the semi-convergence behavior of these methods, we propose two new techniques to specify the relaxation parameters adaptively during the iterations, so as to control the propagated noise component of the error. The advantage of using this strategy for the choice of relaxation parameters on noisy and ill-conditioned problems is demonstrated with an example from tomography (image reconstruction from projections).

Key words. SIRT methods, Cimmino and DROP iteration, semi-convergence, relaxation parameters, tomographic imaging

AMS subject classifications. 65F10, 65R32

1. Introduction. Large-scale discretizations of ill-posed problems (such as imaging problems in tomography) call for the use of iterative methods, because direct factorization methods are infeasible. In particular, there is an interest in regularizing iterations, where the iteration vector can be considered as a regularized solution with the iteration index playing the role of the regularizing parameter. Initially the iteration vector approaches a regularized solution, while continuing the iteration often leads to iteration vectors corrupted by noise. This behavior is called *semi-convergence* by Natterer [20, p. 89]; for analysis of the phenomenon, see, e.g., [2, 3, 13, 14, 15, 21, 22].

This work focuses on a class of non-stationary iteration methods, often referred to as Simultaneous Iterative Reconstruction Techniques (SIRT), including Landweber and Cimmino iteration. These methods incorporate a relaxation parameter, and the convergence rate of the initial iterations depends on the choice of this parameter. In principle, one can use a fixed parameter λ which is found by “training,” i.e., by adjusting it to yield near-optimal convergence for one or more test problems. However, this approach is time consuming and its success depends strongly on the resemblance of the test problems to the given problem.

An attractive alternative is to choose the relaxation parameter automatically in each iteration, in such a way that fast semi-convergence is obtained. In this paper we study the semi-convergence for the SIRT methods, and we use our insight to propose two new methods for adaptively choosing the relaxation parameter.

First, we introduce some notation. Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ be a given matrix and right-hand side, and consider the linear system of equations (which may be inconsistent)

$$Ax \simeq b, \qquad b = \bar{b} + \delta b. \tag{1.1}$$

We assume that the matrix A comes from discretization of some ill-posed linear problem, such as the Radon transform (used, e.g., when modeling reconstruction problems in medicine and biology). We also assume that the noise in the right-hand side is additive, i.e., b consists

*Received July 13, 2009. Accepted for publication July 6, 2010. Published online October 10, 2010. Recommended by F. Sgallari.

†Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden (toelf@mai.liu.se).

‡Department of Mathematics, Iran University of Science and Technology, Narmak, Tehran, 16846-13114, Iran (tonik@mai.liu.se).

§Department of Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark (pch@imm.dtu.dk). This author is supported by grant no. 274-07-0065 from the Danish Research Council for Technology and Production Sciences.

ETNA
Kent State University
http://etna.math.kent.edu

of a noise-free component $\bar{b}$ plus a noise component δb. We consider SIRT methods of the following form for solving (1.1).

ALGORITHM SIRT. Initialization: $x^0 \in \mathbb{R}^n$ is arbitrary. Iteration: update the iteration vector x^k by means of

$$x^{k+1} = x^k + \lambda_k A^T M(b - Ax^k), \qquad k = 0, 1, 2, \ldots. \tag{1.2}$$

Here $\{\lambda_k\}_{k\geq 0}$ are relaxation parameters and M is a given symmetric positive definite (SPD) matrix that depends on the particular method.

Let $\|x\| = \sqrt{x^T x}$ denote the 2-norm, and let $\|x\|_M = \sqrt{x^T M x}$ denote a weighted Euclidean norm (recall that M is assumed to be SPD). Also, let $M^{1/2}$ denote the square root of M, and let $\rho(Q)$ denote the spectral radius of a matrix Q. The following convergence results can be found, e.g., in [5] and [17, Theorem II.3].

THEOREM 1.1. *Let $\rho = \rho(A^T MA)$ and assume that $0 \leq \epsilon \leq \lambda_k \leq (2-\epsilon)/\rho$. If $\epsilon > 0$, or $\epsilon = 0$ and $\sum_{k=0}^{\infty} \min(\rho\lambda_k, 2 - \rho\lambda_k) = +\infty$, then the iterates of* ALGORITHM SIRT *converge to a solution x^* of $\min \|Ax - b\|_M$. If $x^0 \in R(A^T)$ then x^* is the unique solution of minimal Euclidean norm.*

Several well-known fully simultaneous methods can be written in the form of ALGORITHM SIRT for appropriate choices of the matrix M. With M equal to the identity we get the classical Landweber method [18]. Cimmino's method [8] is obtained with $M = \frac{1}{m}\mathrm{diag}(1/\|a_i\|^2)$ where a_i denotes the ith row of A. The CAV method of Censor, Gordon, and Gordon [7] uses $M = \mathrm{diag}(1/\sum_{j=1}^{n} N_j a_{ij}^2)$ where N_j is the number of nonzeroes in the jth column of A. Moreover, if we augment the iterative step (1.2) with a row scaling,

$$x^{k+1} = x^k + \lambda_k S A^T M(b - Ax^k), \qquad k = 0, 1, 2, \ldots,$$

with $S = \mathrm{diag}(m/N_j)$, then we obtain the simultaneous version of the DROP algorithm [6]. The original proposals of some SIRT methods use weights, but for simplicity we do not include weights here.

We now give a short summary of the contents of the paper. In Section 2 we study the semi-convergence behavior of ALGORITHM SIRT using a constant relaxation parameter and show that the total error can be decomposed into two parts, the iteration-error and the noise-error. These two errors can be represented by two functions both depending on the iteration index, the relaxation parameter, and the singular values of the matrix. We derive some results on the behavior of these two functions. Based on this analysis, in Section 3 we propose two new strategies to choose relaxation parameters in ALGORITHM SIRT. The parameters are computed so as to control the propagated noise-error. In the last section we compare our new strategies with two other strategies, using an example taken from image reconstruction from projections (tomography).

2. Analysis of semi-convergence. In order to analyze the mechanism of semi-convergence, in this section we take a closer look at the errors in the regularized solution using ALGORITHM SIRT with a *constant* relaxation parameter λ. Hence we study the iteration scheme

$$x^{k+1} = x^k + \lambda A^T M(b - Ax^k), \qquad k = 0, 1, 2, \ldots. \tag{2.1}$$

2.1. The error in the kth iterate. For convenience we define the quantities

$$B = A^T MA \qquad \text{and} \qquad c = A^T Mb,$$

ETNA
Kent State University
http://etna.math.kent.edu

and we write the singular value decomposition (SVD) of $M^{1/2}A$ as

$$M^{1/2}A = U\Sigma V^T,$$

where $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_p, 0, \dots, 0) \in \mathbb{R}^{m\times n}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p > 0$, and p is the rank of A. Therefore,

$$B = (M^{1/2}A)^T(M^{1/2}A) = V\Sigma^T\Sigma V^T$$

and $\rho(B) = \sigma_1^2$. From (2.1) we then obtain

$$x^k = (I - \lambda B)x^{k-1} + \lambda c = \lambda \sum_{j=0}^{k-1} (I-\lambda B)^j c + (I-\lambda B)^k x^0.$$

Using the SVD of B we can write

$$\sum_{j=0}^{k-1}(I-\lambda B)^j = V E_k V^T,$$

where we have introduced the diagonal matrix,

$$E_k = \mathrm{diag}\left(\frac{1-(1-\lambda\sigma_1^2)^k}{\lambda\sigma_1^2}, \dots, \frac{1-(1-\lambda\sigma_p^2)^k}{\lambda\sigma_p^2}, 0, \dots, 0\right). \tag{2.2}$$

Without loss of generality we can assume $x^0 = 0$ (for a motivation see [13, p. 155]), and then we obtain the following expression for the kth iterate,

$$\begin{aligned} x^k &= V(\lambda E_k)V^T c = V(\lambda E_k)\Sigma^T U^T M^{1/2} b \\ &= \sum_{i=1}^{p} \left(1-(1-\lambda\sigma_i^2)^k\right) \frac{u_i^T M^{1/2}(\bar{b} + \delta b)}{\sigma_i} v_i, \end{aligned} \tag{2.3}$$

where u_i and v_i are the columns of U and V. The quantities $\phi_i = 1 - (1-\lambda\sigma_i^2)^k$ are sometimes called the *filter factors*; see, e.g., [14, p. 138].

Let us now consider the minimum-norm solution $\bar{x}$ to the weighted least squares problem with the noise-free right-hand side,

$$\bar{x} = \mathrm{argmin}_x \|Ax - \bar{b}\|_M.$$

Using the SVD it is straightforward to show that

$$\bar{x} = V E\, \Sigma^T U^T M^{1/2}\bar{b}, \qquad E = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_p^2}, 0, \dots, 0\right). \tag{2.4}$$

Using (2.3) and (2.4) we now obtain an expression for the error in the kth iterate,

$$\begin{aligned} x^k - \bar{x} &= V(\lambda E_k)\Sigma^T U^T M^{1/2}(\bar{b} + \delta b) - V E \Sigma^T U^T M^{1/2}\bar{b} \\ &= V\left((\lambda E_k - E)\Sigma^T U^T M^{1/2}\bar{b} + \lambda E_k \Sigma^T U^T M^{1/2}\delta b\right), \end{aligned} \tag{2.5}$$

and from (2.2) we note that

$$(\lambda E_k - E)\Sigma^T = -\mathrm{diag}\left(\frac{(1-\lambda\sigma_1^2)^k}{\sigma_1}, \dots, \frac{(1-\lambda\sigma_p^2)^k}{\sigma_p}, 0, \dots, 0\right), \tag{2.6}$$

ETNA
Kent State University
http://etna.math.kent.edu

and

$$\lambda E_k \Sigma^T = \operatorname{diag}\left(\frac{1-(1-\lambda\sigma_1^2)^k}{\sigma_1}, \dots, \frac{1-(1-\lambda\sigma_p^2)^k}{\sigma_p}, 0, \dots, 0 \right). \tag{2.7}$$

If we define

$$\bar\beta = U^T M^{1/2} \bar b \qquad \text{and} \qquad \delta\beta = U^T M^{1/2} \delta b,$$

then we can write the error in the SVD basis as

$$V^T (x^k - \bar x) = (\lambda E_k - E)\Sigma^T \bar\beta + \lambda E_k \Sigma^T \delta\beta.$$

For our analysis below, let us introduce the functions

$$\Phi^k(\sigma,\lambda) = \frac{(1-\lambda\sigma^2)^k}{\sigma} \qquad \text{and} \qquad \Psi^k(\sigma,\lambda) = \frac{1-(1-\lambda\sigma^2)^k}{\sigma}. \tag{2.8}$$

Then the jth component of the error $V^T(x^k - \bar x)$ in the SVD basis is given by

$$v_j^T (x^k - \bar x) = -\Phi^k(\sigma_j,\lambda)\bar\beta_j + \Psi^k(\sigma_j,\lambda)\delta\beta_j, \qquad j = 1,\dots,p. \tag{2.9}$$

We see that this component has two contributions, the first term is the *iteration error* and the second term is the *noise error* (other names are used in the literature; for example, the terms "approximation error" and "data error" are used in [13, p. 157]). It is the interplay between these two terms that explains the semi-convergence of the method. Note that for $\lambda\sigma_j^2 \ll 1$ we have $\Psi^k(\sigma_j,\lambda) \approx k\lambda\sigma_j$ showing that k and λ play the same role for suppressing the noise; the same observation is made in [2, p. 145].

2.2. Analysis of the noise-error. We first establish some elementary properties of the functions $\Phi^k(\sigma,\lambda)$ and $\Psi^k(\sigma,\lambda)$.

PROPOSITION 2.1. *Assume that*

$$0 < \epsilon \le \lambda \le 2/\sigma_1^2 - \epsilon \qquad \text{and} \qquad 0 < \sigma < \frac{1}{\sqrt{\lambda}}. \tag{2.10}$$

(a) *Let λ and σ be fixed. As functions of k, $\Phi^k(\sigma,\lambda)$ is decreasing and convex and $\Psi^k(\sigma,\lambda)$ is increasing and concave.*
(b) *For all $k \ge 0$, $\Phi^k(\sigma,\lambda) > 0$, $\Psi^k(\sigma,\lambda) \ge 0$, $\Phi^k(\sigma,0) = 1/\sigma$, and $\Psi^k(\sigma,0) = 0$.*
(c) *Let λ be fixed. For all $k \ge 0$, as function of σ, $\Phi^k(\sigma,\lambda)$ is decreasing.*

Proof. Let $y = y(\sigma) = 1 - \lambda\sigma^2$. Then (2.10) implies that

$$0 < y \le 1 - \epsilon\sigma_p^2 < 1. \tag{2.11}$$

To prove (a) note that $\Phi^k(\sigma,\lambda) = y^k/\sigma$ and $\Psi^k(\sigma,\lambda) = (1-y^k)/\sigma$. Denote by Φ_k^k and Φ_{kk}^k the first and second derivative of $\Phi^k(\sigma,\lambda)$ with respect to k, respectively. Then $\Phi_k^k = (\ln y/\sigma)y^k$ and $\Phi_{kk}^k = ((\ln y)^2/\sigma)y^k$, and hence $\Phi_k^k < 0$ and $\Phi_{kk}^k > 0$ when $y \in (0,1)$. Since $\Psi^k(\sigma,\lambda) = 1/\sigma - \Phi^k(\sigma,\lambda)$ the result for $\Psi^k(\sigma,\lambda)$ follows directly. (b) follows directly from (2.8) and (2.11). To prove (c), let $1/\sqrt{\lambda} > \sigma'' > \sigma' \ge \sigma_p$. Then $y(\sigma''), y(\sigma') \in (0,1)$, and it follows that $\Phi^k(\sigma',\lambda) > \Phi^k(\sigma'',\lambda)$. ☐

REMARK 2.2. The upper bound for σ in (2.10) is $\hat\sigma = 1/\sqrt{\lambda}$. When $0 < \epsilon \le \lambda \le 1/\sigma_1^2$ then clearly $\hat\sigma \ge \sigma_1$. And when $1/\sigma_1^2 < \lambda < 2/\sigma_1^2$ then $\hat\sigma \ge 1/\sqrt{2/\sigma_1^2} = \sigma_1/\sqrt{2}$. Hence $\hat\sigma \ge \sigma_1/\sqrt{2}$ for all relaxation parameters λ satisfying (2.10).

For small values of k the noise-error, expressed via $\Psi^k(\sigma,\lambda)$, is negligible and the iteration approaches the exact solution. When the noise-error reaches the same order of magnitude as the approximation error, the propagated noise-error is no longer hidden in the iteration vector, and the total error starts to increase. The typical overall error behavior is illustrated in Figure 4.1 in Section 4, which shows convergence histories for the Cimmino and DROP methods with a fixed λ. We next investigate the noise-error further.

PROPOSITION 2.3. *Assume that* (2.10) *of Proposition 2.1 holds, and let* λ *be fixed. For all* $k \geq 2$ *there exists a point* $\sigma_k^* \in (0, 1/\sqrt{\lambda})$ *such that*

$$\sigma_k^* = \arg \max_{0<\sigma<1/\sqrt{\lambda}} \Psi^k(\sigma,\lambda).$$

Moreover, σ_k^* *is unique and given by*

$$\sigma_k^* = \sqrt{\frac{1-\zeta_k}{\lambda}}, \tag{2.12}$$

where ζ_k *is the unique root in* $(0,1)$ *of*

$$g_{k-1}(y) = (2k-1)y^{k-1} - (y^{k-2} + \cdots + y + 1) = 0. \tag{2.13}$$

Proof. Denote by Ψ' the derivative of Ψ^k with respect to σ. Then

$$\begin{aligned}
\frac{1}{\lambda}\Psi'(\sigma,\lambda) &= 2k(1-\lambda\sigma^2)^{k-1} - \frac{1-(1-\lambda\sigma^2)^k}{\lambda\sigma^2} \\
&= 2k(1-\lambda\sigma^2)^{k-1} - \frac{1-(1-\lambda\sigma^2)^k}{1-(1-\lambda\sigma^2)} \\
&= (2k-1)y^{k-1} - (y^{k-2} + \cdots + y + 1) = g_{k-1}(y),
\end{aligned}$$

with $y = 1-\lambda\sigma^2$. The function g_{k-1} is continuous with $g_{k-1}(0) = -1$ and $g_{k-1}(1) = k$. Hence there exists at least one point $\zeta_k \in (0,1)$ such that $g_{k-1}(\zeta_k) = 0$. For ease of notation, in the rest of this proof we put $g = g_{k-1}$, $z = \zeta_k$, and $\sigma^* = \sigma_k^*$.

Now $z = 1-\lambda\sigma^2$ so that the point $\sigma^* = \sqrt{\frac{1-z}{\lambda}}$ is a critical point of Ψ^k obviously lying in the open interval $(0, 1/\sqrt{\lambda})$. We now demonstrate the uniqueness of z. It is easy to see that the following equality holds,

$$\begin{aligned}
\frac{g(y)}{y-z} = (2k-1)y^{k-2} &+ ((2k-1)z-1)y^{k-3} \\
&+((2k-1)z^2 - z - 1)y^{k-4} + \cdots \\
&+((2k-1)z^{k-2} - z^{k-3} - \cdots - z - 1) \equiv Q(y).
\end{aligned}$$

Now $Q(0) = g(0)/(-z) = 1/z > 0$. To complete the proof, we show that $Q(y)$ increases for $y > 0$. Let $0 < t < 1$ and $\alpha > 0$ such that $t + \alpha < 1$. Then

$$\begin{aligned}
Q(t+\alpha) - Q(t) = (2k-1)&\left((t+\alpha)^{k-2} - t^{k-2}\right) \\
&+ ((2k-1)z - 1)\left((t+\alpha)^{k-3} - t^{k-3}\right) \\
&+ \left((2k-1)z^2 - z - 1\right)\left((t+\alpha)^{k-4} - t^{k-4}\right) + \cdots \\
&+ \left((2k-1)z^{k-3} - z^{k-4} - \cdots - z - 1\right)\left((t+\alpha) - t\right).
\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

Since $g(z) = (2k-1)z^{k-1} - (z^{k-2} + z^{k-3} + \cdots + z + 1) = 0$, we have

$$\begin{aligned} z^{k-2}\left((2k-1)z - 1\right) &= (z^{k-3} + \cdots + z + 1), \\ z^{k-3}\left((2k-1)z^2 - z - 1\right) &= (z^{k-4} + \cdots + z + 1), \\ &\vdots \\ z^2\left((2k-1)z^{k-3} - z^{k-4} - \cdots - z - 1\right) &= z + 1. \end{aligned}$$

It follows that $Q(t+\alpha) - Q(t) > 0$. Hence z and σ^* are unique. Since $g(y) = (y-z)Q(y)$ and $Q(y) > 0$, we can conclude that

$$g(y) > 0, \text{ when } y > z, \text{ and } g(y) < 0 \text{ when } y < z. \tag{2.14}$$

Now $y < z$ implies that $1 - \lambda\sigma^2 < z$ or $\sigma > \sqrt{(1-z)/\lambda}$, i.e., $\Psi' = \lambda g(y) < 0$ when $\sigma > \sigma^*$ and vice versa. This shows that σ^* is indeed a maximum point of $\Psi^k(\sigma, \lambda)$. ∎

PROPOSITION 2.4. *The sequence* $\{\zeta_k\}_{k\geq 2}$ *defined in Proposition 2.3 satisfies* $0 < \zeta_k < \zeta_{k+1} < 1$, *and* $\lim_{k\to\infty} \zeta_k = \zeta = 1$.

Proof. By Proposition 2.3, $0 < \zeta_k < 1$. Using (2.13), we obtain

$$g_k(y) = (2k+1)y^k - 2ky^{k-1} + g_{k-1}(y). \tag{2.15}$$

We next show $g_k(\zeta_k) < 0$ which by (2.14) (with $g = g_k$ and $z = \zeta_{k+1}$ so that $g(z) = 0$) implies $\zeta_k < \zeta_{k+1}$. Using (2.13) and the geometric series formula, it follows that

$$g_{k-1}(y) = (2k-1)y^{k-1} - \frac{1 - y^{k-1}}{1 - y}. \tag{2.16}$$

With $y = \frac{2k}{2k+1}$, we then get

$$g_{k-1}\left(\frac{2k}{2k+1}\right) = (2k-1)\left(\frac{2k}{2k+1}\right)^{k-1} - \frac{1 - \left(\frac{2k}{2k+1}\right)^{k-1}}{1 - \frac{2k}{2k+1}} = \frac{2(2k)^k - (2k+1)^k}{(2k+1)^{k-1}},$$

which is positive if $2(2k)^k - (2k+1)^k > 0$ or, equivalently, when $2^{1/k} > 1 + 1/(2k)$. One can easily show that $2^x - \frac{x}{2} - 1 > 0$ for $x > 0$. So for $x = 1/k$ follows $g_{k-1}\left(\frac{2k}{2k+1}\right) > 0$. Thus by (2.14)

$$\zeta_k < \frac{2k}{2k+1}. \tag{2.17}$$

It follows, also using (2.15), that

$$\begin{aligned} g_k(\zeta_k) &= (2k+1)(\zeta_k)^k - 2k(\zeta_k)^{k-1} + g_{k-1}(\zeta_k) \\ &= (\zeta_k)^{k-1}\left((2k+1)\zeta_k - 2k\right) < 0. \end{aligned}$$

Hence $\lim_{k\to\infty} \zeta_k = \zeta \leq 1$. We next show that $\zeta = 1$. Using (2.16) with $y = \zeta_k$, and putting $\zeta_k = 1 - z_k$ (so that $0 < z_k < 1$), we get

$$g_{k-1}(1 - z_k) = 0 = (2k-1)(1-z_k)^{k-1} - \frac{1 - (1-z_k)^{k-1}}{z_k}.$$

ETNA
Kent State University
http://etna.math.kent.edu

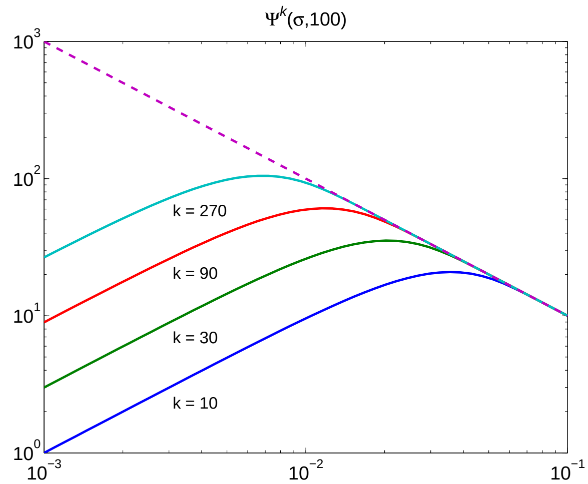

FIG. 2.1. *The function* $\Psi^k(\sigma, \lambda)$ *as a function of* σ*, for* $\lambda = 100$ *and* $k = 10, 30, 90,$ *and* 270*. The dashed line shows* $1/\sigma$*.*

It follows that $(1-z_k)^{k-1}((2k-1)z_k+1) = 1$ and so $(1-z_k)^{-(k-1)} = (2k-1)z_k+1 < 2k$. Therefore $0 < -\ln(1-z_k) = -\ln\zeta_k \leq \ln(2k)/(k-1) \to 0$ as $k \to \infty$. It follows that $\zeta = \lim \zeta_k = 1$. ☐

Figure 2.1 illustrates the behavior of $\Psi^k(\sigma, \lambda)$ as a function of σ, for a fixed λ. We see that σ_k^* (the argument to the maximum) decreases as k increases (which follows from (2.12) and Proposition 2.4). This property implies that the amount of noise in x^k (coming from δb) increases with the number of iterations k, because the contribution from δb becomes less damped. Furthermore, it is seen that the maximal value $\Psi^k(\sigma_k^*, \lambda)$ increases with k, which further increases the amount of noise in x^k. We now prove this last property.

PROPOSITION 2.5. *The value* $\Psi^k(\sigma_k^*, \lambda)$ *is an increasing function of* k.

Proof. By Proposition 2.1 (a), we have

$$\Psi^{k+1}(\sigma, \lambda) \geq \Psi^k(\sigma, \lambda), \; 0 < \sigma \leq \hat{\sigma}.$$

(This result assumes $\sigma \geq \sigma_p$, but it also holds for $\sigma > 0$, as is easily seen.) Hence,

$$\Psi^{k+1}(\sigma_{k+1}^*, \lambda) = \max_{0<\sigma\leq\hat{\sigma}} \Psi^{k+1}(\sigma, \lambda) \geq \max_{0<\sigma\leq\hat{\sigma}} \Psi^k(\sigma, \lambda) = \Psi^k(\sigma_k^*, \lambda). \qquad \square$$

To summarize: using the SVD of $M^{1/2}A$ we have, for constant λ, derived the expression (2.9) for the error in the SVD basis, in the kth iteration. The error depends on two functions Φ and Ψ, where Φ controls the iteration error and Ψ controls the noise error. Both functions depend on k, σ, and λ. In Proposition 2.3 we analyzed the behavior of Ψ as a function of σ. Based on this semi-convergence analysis we propose two new relaxation parameter strategies in the next section.

3. Choice of relaxation parameters. We first review two relaxation parameter strategies proposed in the literature. The first is the optimal choice strategy: this means finding that constant value of λ which gives rise to the fastest convergence to the smallest relative error in the solution. The value of λ is found by searching over the interval $(0, 2/\sigma_1^2)$. This strategy requires knowledge of the exact solution, so for real data one would first need to train the algorithm using simulated data; see [19].

TABLE 3.1
The unique root $\zeta_k \in (0,1)$ of the equation $g_{k-1}(y) = 0$, Eq. (2.13), as function of the iteration index k.

k	ζ_k	k	ζ_k	k	ζ_k	k	ζ_k
2	0.3333	9	0.8574	16	0.9205	23	0.9449
3	0.5583	10	0.8719	17	0.9252	24	0.9472
4	0.6719	11	0.8837	18	0.9294	25	0.9493
5	0.7394	12	0.8936	19	0.9332	26	0.9513
6	0.7840	13	0.9019	20	0.9366	27	0.9531
7	0.8156	14	0.9090	21	0.9396	28	0.9548
8	0.8392	15	0.9151	22	0.9424	29	0.9564

Another strategy is based on picking λ_k such that the error $\|x^* - x^k\|$ is minimized in each iteration, where x^* is a solution to $Ax = b$ (which we now assume to be consistent):

$$\lambda_k = \frac{(r^k)^T M\, r^k}{\|A^T M\, r^k\|_2^2}, \qquad r^k = b - A\, x^k.$$

This strategy is due to Dos Santos [12] (based on work by De Pierro [11]), where the convergence analysis is done for Cimmino's method. Similar strategies have also been proposed by Appleby and Smolarski [1] and Dax [10].

3.1. First strategy. We first propose the following rule for picking relaxation parameters in ALGORITHM SIRT:

$$\lambda_k = \begin{cases} \frac{\sqrt{2}}{\sigma_1^2} & \text{for } k = 0, 1 \\ \frac{2}{\sigma_1^2}(1 - \zeta_k) & \text{for } k \geq 2. \end{cases} \tag{3.1}$$

We will refer to this choice as Ψ_1-based relaxation. We note that the roots ζ_k of $g_{k-1}(y) = 0$, for $k \geq 2$, can easily be precalculated; see Table 3.1. The following theorem ensures that the iterates computed by this strategy converge to the weighted least squares solution when the iterations are carried out beyond the semi-convergence phase.

PROPOSITION 3.1. *The iterates produced using the Ψ_1-based relaxation strategy (3.1) converge toward a solution of $\min_x \|Ax - b\|_M$.*

Proof. Proposition 2.4 gives, for $k \geq 2$,

$$0 = \frac{2}{\sigma_1^2}(1 - \zeta) \leq \lambda_k = \frac{2}{\sigma_1^2}(1 - \zeta_k) \leq \frac{2}{\sigma_1^2}(1 - \zeta_2) < \frac{2}{\sigma_1^2}.$$

It follows that $\min(\sigma_1^2 \lambda_k, 2 - \sigma_1^2 \lambda_k) = \sigma_1^2 \lambda_k$ for k sufficiently large. Next we observe, using (2.17), that $\sum_{k \geq 2} \lambda_k = \frac{2}{\sigma_1^2} \sum_{k \geq 2} (1 - \zeta_k) > \frac{2}{\sigma_1^2} \sum_{k \geq 2} \frac{1}{2k+1} = \infty$. Convergence then follows using Theorem 1.1. $\square$

We will now briefly motivate the choice (3.1). Assume first that the relaxation parameter is kept fixed during the first k iterations,

$$\lambda_j = \lambda, \qquad j = 0, 1, 2, \ldots, k-1,$$

so that the results of Section 2 apply to the first k steps. Also, let x^k and $\bar{x}^k$ denote the iterates of ALGORITHM SIRT using noisy and noise-free data, respectively. Then the error in the kth iterate clearly satisfies

$$\|x^k - \bar{x}\| \leq \|\bar{x}^k - \bar{x}\| + \|x^k - \bar{x}^k\|.$$

ETNA
Kent State University
http://etna.math.kent.edu

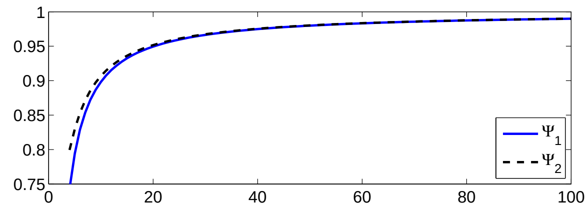

FIG. 3.1. *The ratio λ_{j+1}/λ_j as a function of j for the two parameter-choice strategies.*

Hence the error decomposes into two components, the iteration-error part $\|\bar{x}^k - \bar{x}\|$ and the noise-error part $\|x^k - \bar{x}^k\|$. Using Eqs. (2.3), (2.4), (2.6), and (2.7) we obtain

$$\bar{x}^k - \bar{x} = V(\lambda E_k - E)\Sigma^T U^T M^{1/2}\bar{b}, \qquad x^k - \bar{x}^k = V\lambda E_k \Sigma^T U^T M^{1/2}\delta b.$$

Hence the norm of the noise-error is bounded by

$$\|x^k - \bar{x}^k\| \leq \max_{1\leq i\leq p} \Psi^k(\sigma_i, \lambda)\, \|M^{1/2}\delta b\|. \tag{3.2}$$

To analyze (3.2) further, assume first that $\lambda \in (0, 1/\sigma_1^2]$. Then by Remark 2.2, we have $\hat{\sigma} \geq \sigma_1$, and it follows (with $k \geq 2$) that

$$\max_{1\leq i\leq p} \Psi^k(\sigma_i, \lambda) \leq \max_{0\leq\sigma\leq\sigma_1} \Psi^k(\sigma, \lambda) \leq \max_{0\leq\sigma\leq\hat{\sigma}} \Psi^k(\sigma, \lambda) = \Psi^k(\sigma_k^*, \lambda). \tag{3.3}$$

It follows using (2.8) and (2.12) that

$$\|x^k - \bar{x}^k\| \leq \Psi^k(\sigma_k^*, \lambda)\|M^{1/2}\delta b\| = \sqrt{\lambda}\, \frac{1-\zeta_k^k}{\sqrt{1-\zeta_k}}\|M^{1/2}\delta b\|. \tag{3.4}$$

Now consider the kth iteration step, and pick λ_k by the rule (3.1). Assuming that the relaxation parameters are such that $\lambda_{j+1}/\lambda_j \approx 1$, we can expect that Eq. (3.4) holds approximatively – the plot of the ratio λ_{j+1}/λ_j in Figure 3.1 indicates the validity of this assumption. Therefore by substituting (3.1) into (3.4) we get

$$\|x^k - \bar{x}^k\| \lesssim \frac{\sqrt{2}}{\sigma_1}(1-\zeta_k^k)\, \|M^{1/2}\delta b\|, \qquad k \geq 2. \tag{3.5}$$

This implies that the rule (3.1) provides an upper bound for the noise-part of the error. This is our heuristic motivation for using the relaxation rule: to monitor and control the noise part of the error.

We also need to consider the case $\lambda \in (1/\sigma_1^2, 2/\sigma_1^2)$, and it follows by Remark 2.2 that $\hat{\sigma} = 1/\sqrt{\lambda} \in (\sigma_1/\sqrt{2}, \sigma_1)$. This means that (3.3) only holds approximatively. However, using our proposed relaxation rule, we have that $\lambda_k \leq 1/\sigma_1^2$ for $k > 2$; see Table 3.1.

3.2. Second strategy. We next consider an alternative choice of relaxation parameters,

$$\lambda_k = \begin{cases} \frac{\sqrt{2}}{\sigma_1^2} & \text{for } k = 0, 1 \\ \frac{2}{\sigma_1^2}\frac{1-\zeta_k}{(1-\zeta_k^k)^2} & \text{for } k \geq 2. \end{cases} \tag{3.6}$$

The reason we also introduce (3.6) is that in our numerical tests we found that it usually gives faster convergence than (3.1). We will refer to this choice as Ψ_2-based relaxation. Reasoning

ETNA
Kent State University
http://etna.math.kent.edu

as above (substituting (3.6) into (3.4)) we get the following bound for the noise-error using Ψ_2-based relaxation,

$$\|x^k - \bar{x}^k\| \le \frac{\sqrt{2}}{\sigma_1}\|M^{1/2}\delta b\|, \qquad k \ge 2. \tag{3.7}$$

Figure 3.1 also shows the ratio λ_{j+1}/λ_j for this strategy.

We stress again that the two bounds (3.5) and (3.7) for the noise error are derived under the assumption of a fixed noise level. Further we assumed that λ_j is fixed for $j = 0, 1, \ldots, k-1$ and that λ_k is given by (3.1) or (3.6). We leave it as an open problem to derive rigorous upper bounds for the noise error without these assumptions. In this context we also mention the bound $\|x^k - \bar{x}^k\| \le \sqrt{k}\,\|M^{1/2}\delta b\|$ from [13, Lemma 6.2] when using a constant relaxation parameter in the algorithm.

Again we need to show convergence to the weighted least squares solution when iterating beyond the semi-convergence phase. To do this we shall use the following result.

LEMMA 3.2.

$$2^t > \frac{2-t-t^2}{2-2t-t^2}, \qquad 0 < t \le 1/4.$$

Proof. Let

$$h(t) = 2^t(2-2t-t^2) - (2-t-t^2).$$

We need to show that $h(t) > 0$ for $0 < t \le \frac{1}{4}$. The third derivative of h satisfies

$$h'''(t) = -2^t \log(2)\, p(t) < 0 \quad \text{for} \quad 0 < t \le \frac{1}{4},$$

where $p(t) = 6 - 2\log^2(2) + \log(64) + t\left(2\log^2(2) + \log(64)\right) + t^2\log^2(2)$ is a positive polynomial for $0 < t \le 1/4$. It follows that $h''(t)$ decreases as a function of t. Using $h''(0) = 2\log(2)(-2 + \log(2)) < 0$, we obtain

$$h''(t) < h''(0) < 0,$$

which means the function $h(t)$ is concave for $0 < t \le 1/4$. Therefore, $h(t) > 0$ for $0 < t \le 1/4$ since $h(0) = 0$ and $h(1/4) = (23\sqrt[4]{2} - 27)/16 > 0$. □

PROPOSITION 3.3. *The iterates produced using Ψ_2-based relaxation (3.6) converge toward a solution of* $\min_x \|Ax - b\|_M$.

Proof. For $z \in (0,1)$, let

$$\gamma_k(z) = \frac{1-z}{(1-z^k)^2}.$$

Then

$$\gamma_{k+1}(z) = \frac{1-z}{(1-z^{k+1})^2} < \frac{1-z}{(1-z^k)^2} = \gamma_k(z). \tag{3.8}$$

We will next show that

$$\gamma_k(\zeta_{k+1}) < \gamma_k(\zeta_k). \tag{3.9}$$

ETNA
Kent State University
http://etna.math.kent.edu

Using $g_{k-1}(\zeta_k) = 0$ and $1-\zeta_k^k = (1-\zeta_k)(1+\zeta_k+\ldots+\zeta_k^{k-1})$, it can easily be shown that $\gamma_k(\zeta_k) = (2k(1-\zeta_k^k)\zeta_k^{k-1})^{-1}$. Then

$$\frac{d\gamma_k}{d\zeta_k} = \frac{(2k-1)\zeta_k^k + (1-k)}{2k(1-\zeta_k^k)^2\zeta_k^k} < 0,$$

provided that

$$\zeta_k < \left(\frac{k-1}{2k-1}\right)^{1/k} \equiv z_k. \tag{3.10}$$

The inequality (3.10) holds for $k = 2, 3$ since $\zeta_2 = 1/3$ and $\zeta_3 = \frac{1+\sqrt{21}}{10}$. For $k \geq 4$ we will show that

$$g_k(z_k) > 0. \tag{3.11}$$

This implies (3.10) since, by (2.14) and Proposition 2.4,

$$z_k > \zeta_{k+1} > \zeta_k.$$

Now (3.11) is equivalent to

$$(2k-1)(2k^2-2k-1)^k > (k-1)(2k^2-k-1)^k.$$

Since $-(2k^2-2k-1)^k > -(2k^2-k-1)^k$, it suffices to show that

$$2^{1/k} > \frac{2k^2-k-1}{2k^2-2k-1}.$$

This inequality follows from Lemma 3.2 by putting $t = 1/k$. Hence (3.9) holds. Using (3.8) and (3.9),

$$0 < \gamma_{k+1}(\zeta_{k+1}) < \gamma_k(\zeta_{k+1}) < \gamma_k(\zeta_k) < \gamma_2(\zeta_2) < 1.$$

It follows that

$$\lim_{k\to\infty} \gamma_k(\zeta_k) = \gamma \geq 0.$$

Therefore,

$$0 \leq \frac{2}{\sigma_1^2}\gamma \leq \lambda_k = \frac{2}{\sigma_1^2}\gamma_k(\zeta_k) \leq \frac{2}{\sigma_1^2}\gamma_2(\zeta_2) < \frac{2}{\sigma_1^2}.$$

If $\gamma > 0$ convergence follows by invoking the first condition in Theorem 1.1. Next assume $\gamma = 0$, and note that $1/(1-\zeta_k^k)^2 > 1$, since $\zeta_k < 1$. It also follows that $\min(\sigma_1^2\lambda_k, 2-\sigma_1^2\lambda_k) = \sigma_1^2\lambda_k$ for k large enough. Therefore, using (2.17), we have $\sum_{k\geq 2}\lambda_k > 2/\sigma_1^2 \sum_{k\geq 2} 1/(2k+1) = \infty$. Hence, convergence follows by the second condition in Theorem 1.1. (The same arguments could also have been used in the proof of Proposition 3.1, thus avoiding the need to show that $\zeta = 1$). $\square$

3.3. Modified strategies. Finally we will explore the possibility to accelerate the two new strategies, via the choice $\bar{\lambda}_k = \tau_k \lambda_k$ for $k \geq 2$, where τ_k a parameter to be chosen. Consider first the Ψ_1 strategy, for which we have

$$\bar{\lambda}_k = \tau_k \frac{2}{\sigma_1^2}(1-\zeta_k), \qquad k \geq 2. \tag{3.12}$$

We will assume that τ_k satisfies

$$0 < \epsilon_1 \leq \tau_k < (1-\zeta_k)^{-1}, \qquad k \geq 2. \tag{3.13}$$

It then follows that

$$0 \leq \epsilon_1 \frac{2}{\sigma_1^2}(1-\zeta_k) \leq \bar{\lambda}_k < \frac{2}{\sigma_1^2}.$$

Hence using that $\sum_{k\geq 2}(1-\zeta_k) = \infty$ (see proof of Proposition 3.1) we may invoke Theorem 1.1 to conclude convergence of the relaxation strategy (3.12) with (3.13). Note that with a constant value $\tau_k = \tau$ we must pick $\tau < (1-\zeta_2)^{-1} \approx 1.5$. If we allow τ_k to depend on k then Table 3.1 immediately leads to the following upper bounds: $\tau_2 < 1.5$, $\tau_3 < 2.264$, $\tau_4 < 3.048$, and $\tau_5 < 3.84$.

For the Ψ_2 strategy we take

$$\bar{\lambda}_k = \tau_k \frac{2}{\sigma_1^2} \frac{1-\zeta_k}{(1-\zeta_k^k)^2}, \qquad k \geq 2,$$

and with

$$0 < \epsilon_2 \leq \tau_k < (1-\zeta_k^k)^2/(1-\zeta_k), \qquad k \geq 2,$$

we maintain convergence (following the same reasoning as above). Using Table 3.1 we obtain the following upper bounds: $\tau_2 < 1.185$, $\tau_3 < 1.545$, $\tau_4 < 1.932$ and $\tau_5 < 2.33$. If we want to use a constant value we must take $\tau_k = \tau < 1.185$. Our numerical results indicate that picking $\tau > 1$ accelerates the convergence.

4. Computational results. We report some numerical tests with an example taken from the field of tomographic image reconstruction from projections (see [16] for a recent and illustrative treatment), using the SNARK93 software package [4] and the standard head phantom from Herman [15]. The phantom is discretized into 63×63 pixels, and we use 16 projections with 99 rays per projection. The resulting projection matrix A has, therefore, the dimensions 1376×3969 (so that the system of equations is highly under-determined). In addition to A, the software produces an exact right-hand side $\bar{b}$ and an exact solution $\bar{x}$. By using SNARK93's right-hand side $\bar{b}$, which is not generated as the product $A\bar{x}$, we avoid committing an inverse crime where the exact same model is used in the forward and reconstruction models.

We add independent Gaussian noise of mean 0 and various standard deviations to generate three different noise levels $\eta = \|\delta b\|/\|\bar{b}\| = 0.01$, 0.05, and 0.08. Our figures show the relative errors in the reconstructions, defined as

$$\text{relative error} = \|\bar{x} - x^k\|/\|\bar{x}\|,$$

as functions of the iteration index k.

Figure 4.1 shows relative error histories for the fixed-λ Cimmino and DROP methods, taking three different choices of λ and two different noise levels η for each method. Here

ETNA
Kent State University
http://etna.math.kent.edu

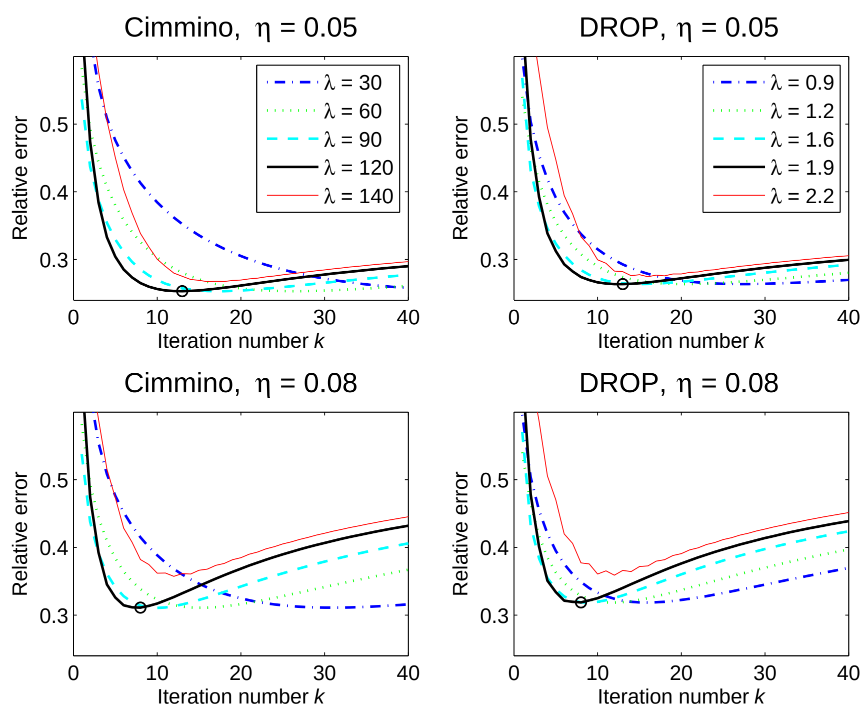

FIG. 4.1. *Fixed-λ Cimmino and DROP iterations, using five different choices of λ and two different noise levels η for each method. The choices $\lambda = 120$ and $\lambda = 1.9$ are optimal for Cimmino and DROP, respectively, and the circle shows the minimum.*

the upper bound $2/\sigma_1^2$ for the relaxation parameter equals 150.6 for Cimmino and 2.38 for DROP. The choices $\lambda = 120$ (for Cimmino) and $\lambda = 1.9$ (for DROP) are optimal, in the sense that they give the fastest convergence to the minimal error.

We clearly see the semi-convergence of the iterations studied in Section 2 and that, consequently, fewer iterations are needed to reach the minimum error when the noise level increases. We make two important observations: the number of iterations to reach the minimum error depends strongly on the choice of λ, and the minimum itself is practically independent of λ (except for the largest λ). This illustrates the necessity of using either a good fixed relaxation parameter, which requires an extensive study of model problems to find a close-to-optimal value, or a parameter-choice method that chooses λ_k such that fast semi-convergence is achieved automatically.

Figure 4.2 shows the relative error histories for the Ψ_1, Ψ_2, and optimal strategies using the Cimmino and DROP methods. For Cimmino's method we also include the strategy proposed by Dos Santos mentioned in the previous section. We observe a noise-damping effect using the Ψ_1 and Ψ_2 strategies. The zigzagging behavior of the Dos Santos strategy was also noted by Combette [9, pp. 479 and 504]; the reason seems to be that the strategy assumes consistent data.

We see that for low-noise data the Ψ_1 and Ψ_2 strategies are less efficient than the Dos Santos and optimal strategies. However, for larger noise levels (where the Dos Santos strategy leads to irregular convergence) our new methods produce an initial convergence rate, during the semi-convergent phase, which is close to that of the optimal strategy. Note that the relative error stays almost constant after the minimum has been obtained, showing that the Ψ_1 and

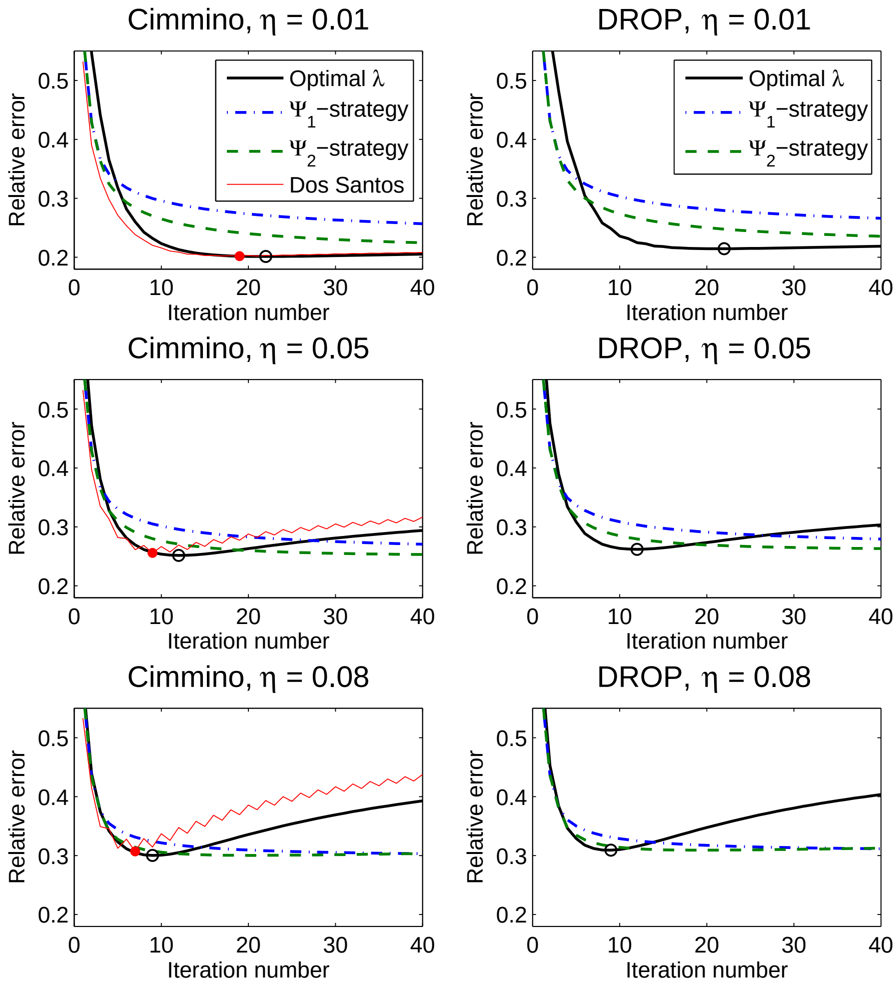

FIG. 4.2. *Relative error histories for different relaxation strategies in the Cimmino and DROP methods, for three different noise levels η. The circle and the dot show the minimum for the optimal strategy and the Dos Santos strategy, respectively.*

Ψ_2 strategies are indeed able to dampen the influence of the noise-error, as desired. The very flat minimum reduces the sensitivity of the solution to the particular stopping criterion used.

Figure 4.3 shows the relative error histories using Cimmino's method with the modified Ψ_1 and Ψ_2 strategies from Section 3.3, for four different choices of a constant τ. (Note that $\tau = 1$ correspond to the original Ψ_1 and Ψ_2 strategies.) As mentioned in Section 3.3, in the modified Ψ_2 strategy the theoretical upper limit for τ is 1.18, but Figure 4.3 shows it pays to allow a somewhat larger value and we found $\tau_2 = 1.5$ is a reasonable choice for the modified Ψ_2 strategy. (Note that $\tau = 2$ leads to instability during the first iterations.) Similarly, for the Ψ_1-strategy (where the upper bound for τ is 1.5), we found that $\tau = 2$ is a reasonable choice. Of course, one could here also consider other options, such as allowing τ to depend on the iteration index, or introducing a constant factor τ after more than two iterations. But since we are looking for a simple and self-contained method we have not considered these choices.

Finally in Figure 4.4, we compare the two best modified strategies together with the

ETNA
Kent State University
http://etna.math.kent.edu

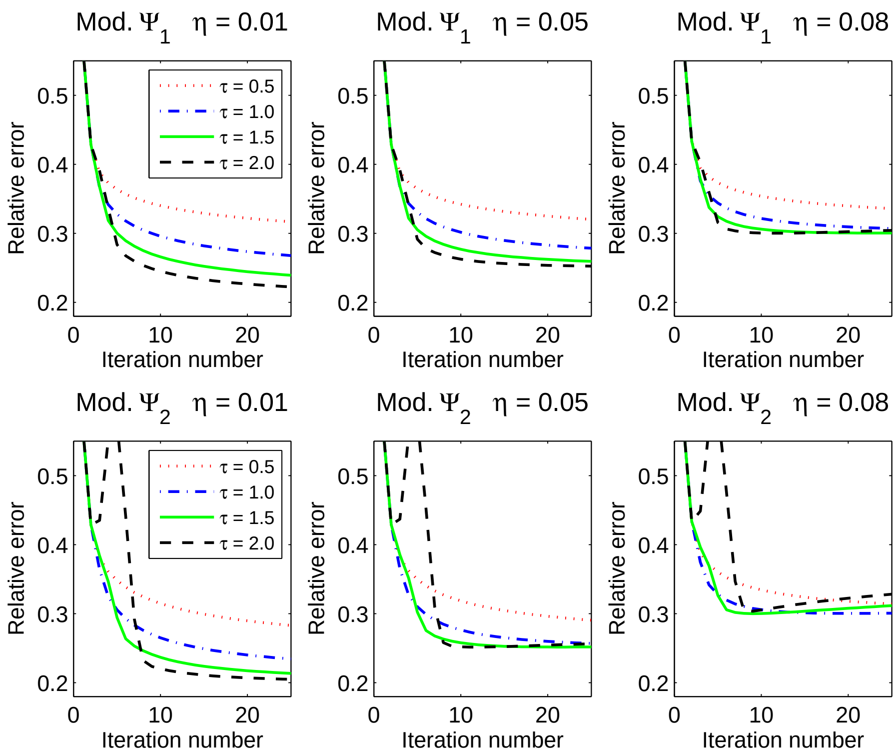

FIG. 4.3. *Relative error histories using the modified* Ψ_1 *and* Ψ_2 *relaxation strategies in Cimmino's method, for three different noise levels* η.

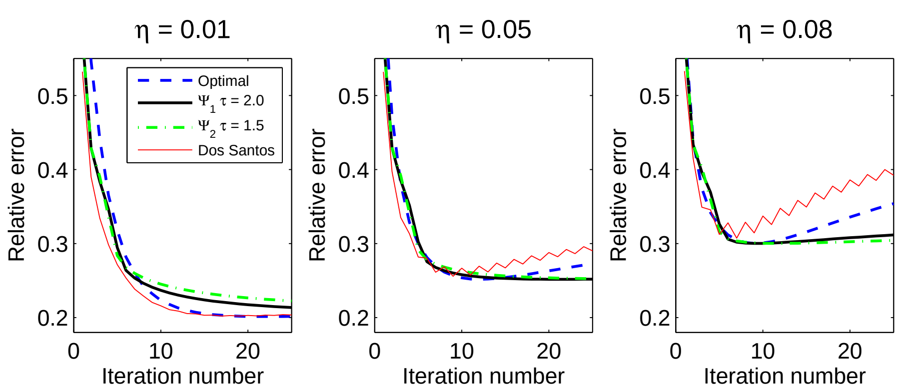

FIG. 4.4. *Comparison of the error histories for the modified* Ψ_1 *and* Ψ_2 *relaxation strategies in Cimmino's method with those of the optimal strategy and the Dos Santos strategy, for three different noise levels* η.

optimal strategy and the Dos Santos strategy. We see that for the larger noise levels, the modified Ψ_1 and Ψ_2 strategies give an initial convergence which is almost identical to that of the other two methods, but with much better damping of the noise propagation: once we reach the minimum in the error histories, then the error only increases slowly. Also, we avoid the erratic behavior of the Dos Santos strategy.

ETNA
Kent State University
http://etna.math.kent.edu

5. Conclusion. Using theoretical results for the semi-convergence of the SIRT algorithm with a fixed relaxation parameter, we derive two new strategies for choosing the parameter adaptively in each step in order to control the propagated noise component of the error. We prove that with these strategies, the SIRT algorithm will still converge in the noise-free case. Our numerical experiments show that if the noise is not too small, then the initial convergence of the SIRT algorithm with our strategies is competitive with the Dos Santos strategy (which leads to erratic convergence) as well as the optimal-choice strategy (which depends on a careful "training" of the parameter). The experiments also show that our strategies carry over to the DROP algorithm which is a weighted SIRT method.

Acknowledgment. We wish to thank Dr. Arne Enqvist for help in proving Lemma 3.2. We also wish to thank three anonymous referees for constructive criticism and helpful suggestions which improved our paper.

REFERENCES

[1] G. APPLEBY AND D. C. SMOLARSKI, *A linear acceleration row action method for projecting onto subspaces*, Electron. Trans. Numer. Anal., 20 (2005), pp. 253–275. http://etna.math.kent.edu/vol.20.2005/pp253-275.dir.
[2] M. BERTERO AND P. BOCCACCI, *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, 1998.
[3] P. BRIANZI, F. DI BENEDETTO, AND C. ESTATICO, *Improvement of space-invariant image deblurring by preconditioned Landweber iterations*, SIAM J. Sci. Comp., 30 (2008), pp. 1430–1458.
[4] J. A. BROWNE, G. T. HERMAN, AND D. ODHNER, *SNARK93: A Programming System for Image Reconstruction from Projections*, The Medical Imaging Processing Group (MIPG), Department of Radiology, University of Pennsylvania, Technical Report MIPG198, 1993.
[5] Y. CENSOR AND T. ELFVING, *Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 40–58.
[6] Y. CENSOR, T. ELFVING, G. T. HERMAN, AND T. NIKAZAD, *On diagonally-relaxed orthogonal projection methods*, SIAM J. Sci. Comput., 30 (2008), pp. 473–504.
[7] Y. CENSOR, D. GORDON, AND R. GORDON, *Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems*, Parallel Computing, 27 (2001), pp. 777–808.
[8] G. CIMMINO, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*, La Ricerca Scientifica, XVI, Series II, Anno IX, 1 (1938), pp. 326–333.
[9] P. L. COMBETTES, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, IEEE Trans. Image Process., 6 (1997), pp. 493–506.
[10] A. DAX, *Line search acceleration of iterative methods*, Linear Algebra Appl., 130 (1990), pp. 43–63.
[11] A. R. DE PIERRO, *Methodos de projeção para a resolução de sistemas gerais de equações algébricas lineares*, Thesis (tese de Doutoramento), Instituto de Matemática da UFRJ, Cidade Universitária, Rio de Janeiro, 1981.
[12] L. T. DOS SANTOS, *A parallel subgradient projections method for the convex feasibility problem*, J. Comput. Appl. Math., 18 (1987), pp. 307–320.
[13] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, The Netherlands, 1996.
[14] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[15] G. T. HERMAN, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Second ed., Springer, New York, 2009.
[16] G. T. HERMAN AND R. DAVIDI, *Image reconstruction from a small number of projections*, Inverse Problems, 24 (2008), 045011.
[17] M. JIANG AND G. WANG, *Convergence studies on iterative algorithms for image reconstruction*, IEEE Transactions on Medical Imaging, 22 (2003), pp. 569–579.
[18] L. LANDWEBER, *An iterative formula for Fredholm integral equations of the first kind*, Amer. J. Math., 73 (1951), pp. 615–624.
[19] R. MARABINI, C. O. S. SORZANO, S. MATEJ, J. J. FERNÁNDEZ, J. M. CARAZO, AND G. T. HERMAN, *3-D Reconstruction of 2-D crystals in real space*, IEEE Transactions on Image Processing, 13 (2004), pp. 549–561.
[20] F. NATTERER, *The Mathematics of Computerized Tomography*, John Wiley, New York, 1986.
[21] M. PIANA AND M. BERTERO, *Projected Landweber method and preconditioning*, Inverse Problems, 13 (1997), pp. 441–464.
[22] A. VAN DER SLUIS AND H. A. VAN DER VORST, *SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems*, Linear Algebra Appl., 130 (1990), pp. 257–303.