



## Semiconvergence and Relaxation Parameters for Projected SIRT Algorithms

Elfving, Tommy; Hansen, Per Christian; Nikazad, Touraj

*Published in:*  
S I A M Journal on Scientific Computing

*Link to article, DOI:*  
[10.1137/110834640](https://doi.org/10.1137/110834640)

*Publication date:*  
2012

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Elfving, T., Hansen, P. C., & Nikazad, T. (2012). Semiconvergence and Relaxation Parameters for Projected SIRT Algorithms. S I A M Journal on Scientific Computing, 34(4), A2000-A2017. DOI: 10.1137/110834640

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## SEMICONVERGENCE AND RELAXATION PARAMETERS FOR PROJECTED SIRT ALGORITHMS\*

TOMMY ELFVING<sup>†</sup>, PER CHRISTIAN HANSEN<sup>‡</sup>, AND TOURAJ NIKAZAD<sup>§</sup>

**Abstract.** We give a detailed study of the semiconvergence behavior of projected nonstationary simultaneous iterative reconstruction technique (SIRT) algorithms, including the projected Landweber algorithm. We also consider the use of a relaxation parameter strategy, proposed recently for the standard algorithms, for controlling the semiconvergence of the projected algorithms. We demonstrate the semiconvergence and the performance of our strategies by examples taken from tomographic imaging.

**Key words.** projected Landweber iteration, simultaneous iterative reconstruction technique, relaxation parameters, semiconvergence, nonnegativity constraints, tomographic imaging

**AMS subject classifications.** 65F10, 65R32

**DOI.** 10.1137/110834640

**1. Introduction.** Computational tomography, as well as many other large-scale ill-posed imaging problems, lead to large linear systems of equations (often inconsistent) with noisy data, of the form

$$(1.1) \quad Ax \simeq b, \quad b = \bar{b} + \delta b, \quad A \in \mathbb{R}^{m \times n},$$

where  $\bar{b}$  denotes the exact data and  $\delta b$  is the perturbation consisting of additive noise. The matrix  $A$  comes from the discretization of an ill-posed linear problem such as the Radon transform. The numerical solution of these systems calls for the use of iterative regularization methods, and it is often necessary to incorporate additional constraints to the reconstruction. In this paper we assume that the solution must belong to a convex set, e.g., the positive orthant corresponding to nonnegativity constraints. Such constraints incorporate prior physical knowledge about the solution, and therefore they typically lead to smaller reconstruction errors (see Figure 2.1 for an example). Some applications of projected iterative methods in seismology, image restoration, nonnegative matrix factorization, matrix completion, and supervised learning can be found in [1], [2], [3], [7], [23], [28], [31], [32].

We focus on regularizing iterations with semiconvergence, where the iteration number plays the role of the regularizing parameter. In the early stages the iteration vector approaches a regularized solution, while continuing the iteration leads to iteration vectors deteriorated by noise; cf. [1], [19], [20], [22], [27]. The semiconvergence of projected iterative methods has been noted in several papers [2], [3], [7]. It is also discussed in [19] and analyzed in an infinite dimensional setting by Eicke [16].

---

\*Submitted to the journal's Methods and Algorithms for Scientific Computing section May 19, 2011; accepted for publication (in revised form) April 13, 2012; published electronically July 17, 2012.

<http://www.siam.org/journals/sisc/34-4/83464.html>

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

<sup>‡</sup>Department of Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark (pch@imm.dtu.dk). The author is supported by grant 274-07-0065 from the Danish Research Council for Technology and Production Sciences.

<sup>§</sup>Department of Mathematics, Iran University of Science and Technology, Narmak, Tehran, 16846-13114, Iran (tnikazad@yahoo.com).

The main goal of our paper is to analyze semiconvergence for a class of nonstationary iteration methods in  $\mathbb{R}^n$ , often referred to as simultaneous iterative reconstruction technique (SIRT) [17], [21], including Landweber and Cimmino iterations, with a focus on their constrained companion versions that include projection in each iteration.

In a previous paper [18] we studied semiconvergence properties of unconstrained SIRT methods using a constant relaxation parameter. We also proposed two new nonstationary relaxation strategies and provided an asymptotic convergence analysis. Here we extend the semiconvergence analysis in two ways: (i) we include the constrained versions of SIRT, and (ii) we allow a certain family of nonstationary relaxation parameters which includes the two strategies already proposed in [18]. We also show that the bound for the noise-error using one of the strategies is smaller than the bound for the other. This theoretical result is confirmed by our numerical results.

We note that the role of the relaxation parameter is really twofold. The classical use of the step-length in optimization algorithms is to guarantee global (and fast) convergence. On the other hand, as demonstrated both theoretically and numerically in this paper, the step-length can also be used to suppress noise in connection with semiconvergence.

We start with a brief summary of the SIRT methods and their constrained versions followed by our analysis of semiconvergence. Next we discuss our parameter-choice strategies, and we finish with a few numerical examples from medical tomography. Throughout the paper  $\|\cdot\|$  denotes the vector and matrix 2-norm, and for  $z \in \mathbb{R}^m$  and a symmetric positive definite (SPD) matrix  $M \in \mathbb{R}^{m \times m}$  we define the weighted Euclidean norm  $\|z\|_M = \sqrt{z^T M z}$ . Moreover,  $M^{1/2}$  is the square root of  $M$ , and  $\rho(Q)$  is the spectral radius of the matrix  $Q$ .

**2. Projected SIRT methods.** The classical (unprojected) SIRT method and its projected version P-SIRT are summarized below.

*Algorithm SIRT.* Choose an arbitrary initial vector  $\hat{x}_0 \in \mathbb{R}^n$  and a sequence of positive relaxation parameters  $\{\lambda_k\}$ , and update the iteration vector  $\hat{x}_k$  by means of

$$(2.1) \quad \hat{x}_{k+1} = \hat{x}_k + \lambda_k A^T M (b - A \hat{x}_k), \quad k = 0, 1, 2, \dots$$

Several well-known methods can be written in the form (2.1) for appropriate choices of the matrix  $M$ . With  $M$  equal to the identity we get the classical Landweber method [24]. Cimmino's method [12] is obtained with  $M = M_C = \frac{1}{m} \text{diag}(1/\|a_i\|^2)$ , where  $a_i$  denotes the  $i$ th row of  $A$ . The component averaging (CAV) method [11] uses  $M = M_{CAV} = \text{diag}(1/\sum_{j=1}^n N_j a_{ij}^2)$ , where  $N_j$  is the number of nonzeros in the  $j$ th column of  $A$ . The original proposals of some unprojected SIRT methods use weights [17], but for simplicity we do not include weights here.

*Algorithm P-SIRT.* Let  $\mathcal{C}$  denote a closed convex set, and let  $P_{\mathcal{C}}$  be the metric projection onto the set  $\mathcal{C}$ . Choose an arbitrary initial vector  $x_0 \in \mathbb{R}^n$  and a sequence of positive relaxation parameters  $\{\lambda_k\}$ , and update the iteration vector  $x_k$  by means of

$$(2.2) \quad x_{k+1} = P_{\mathcal{C}}(x_k + \lambda_k A^T M (b - A x_k)), \quad k = 0, 1, 2, \dots$$

This is an instance of the gradient projection algorithm, for  $M = I$  often called the projected Landweber iteration, which aims at finding a solution of

$$(2.3) \quad \min_{x \in \mathcal{C}} \frac{1}{2} \|Ax - b\|_M^2.$$

The following convergence result is given in, e.g., [4] and [29].

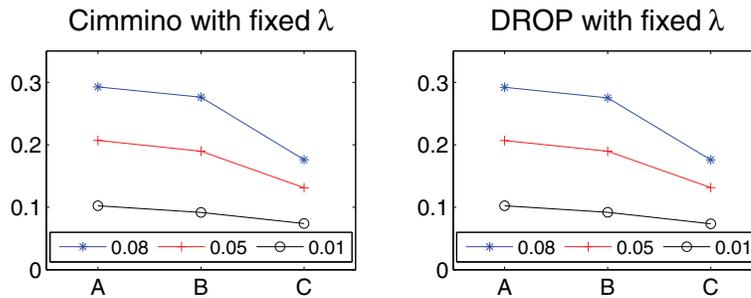


FIG. 2.1. Reconstruction errors for the Cimmino and DROP methods in their unprojected and projected versions, and with a fixed relaxation parameter. We used three different relative noise levels. A = unprojected algorithm with iterates  $\hat{x}_k$ ; B = projection applied to  $\hat{x}_k$  giving the vectors  $P_C(\hat{x}_k)$ ; C = projected algorithm with iterates  $x_k$ .

THEOREM 2.1. Let  $\rho = \rho(A^T M A)$ , and assume that  $0 < \epsilon \leq \lambda \leq (2 - \epsilon)/\rho$  and also that the solution set of (2.3) is nontrivial. Then the sequence generated by (2.2) with  $\lambda_k = \lambda$  converges to a solution of (2.3).

In [25] linear convergence of the projected Landweber method was shown, and in [23] an expression for the rate constant was derived.

The constrained problem (2.3) has a solution for any right-hand side  $b$  if and only if  $A(\mathcal{C})$ , the image set of  $\mathcal{C}$ , is closed; cf. [9, p. 444], [28, p. 442]. This is the case, e.g., if  $\mathcal{C}$  is bounded or when the objective function is coercive, i.e., the null space of  $A$  is nontrivial. Note that for this case the objective function is strictly convex and a unique solution exists. Although these results are interesting, it is important to remember that for ill-conditioned problems—even in the finite dimensional case considered here—the solution set is usually completely deprived of any physical meaning due to the inverted noise. However, numerical results [3], [28] show that the method exhibits semiconvergence so that during the first steps the iterates improve, while later due to noise propagation they start to deteriorate.

By introducing an SPD matrix  $S$  in the two SIRT algorithms (2.1) and (2.2) we obtain the iterations

$$(2.4) \quad \begin{aligned} \hat{x}_{k+1} &= \hat{x}_k + \lambda_k S A^T M (b - A \hat{x}_k), \\ x_{k+1} &= P_C(x_k + \lambda_k S A^T M (b - A x_k)), \end{aligned} \quad k = 0, 1, 2, \dots$$

The latter is known as the *scaled gradient projection method* [4, pp. 209–210]; if we choose  $M = M_C$  and  $S = \text{diag}(m/N_j)$  in the former, then we obtain the diagonally relaxed orthogonal projection (DROP) algorithm from [10]. When the matrix is dense,  $N_j = m$ , and Cimmino's method is obtained. When  $N_j < m$ , DROP takes larger stepsizes than Cimmino, which can much improve the initial rate of convergence [10], [11]. For recent developments and applications using the scaled gradient projection method see [3], [6], and [7].

Before continuing we briefly illustrate the importance of solving the constrained problem—instead of solving the unconstrained problem, possibly followed by a projection. For a certain test problem (see section 6) and with  $\mathcal{C}$  equal to the nonnegative orthant,  $M = M_C$  and  $S = \text{diag}(m/N_j)$ , we computed the reconstruction errors associated with the solutions  $\hat{x}_k$  from the unprojected algorithms (Cimmino and DROP), the projected vectors  $P_C(\hat{x}_k)$  (i.e., the projection onto  $\mathcal{C}$  of the unprojected iterates  $\hat{x}_k$ ), and the iterates  $x_k$  from the projected algorithms. Figure 2.1 shows the mini-

imum reconstruction errors for three different noise levels (the optimal iteration index  $k$  depends on the noise level). Clearly, the projected algorithms give the smallest reconstruction errors. Most often the set  $\mathcal{C}$  is chosen so as to satisfy a natural physical constraint from the application at hand; a typical example is nonnegativity arising, e.g., in tomography.

Below we study the semiconvergence behavior of Algorithm P-SIRT, as well as the scaled gradient projection method, using a decreasing sequence of relaxation parameters  $\lambda_k$ . 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 a matrix. In [18] we analyzed the behavior of these two functions and proposed new strategies to choose relaxation parameters in Algorithm SIRT in such a way that they control the propagated noise-error. We will show that this analysis, slightly modified, is valid also for the projected algorithms considered here.

**3. Analysis of semiconvergence for P-SIRT.** In this section we study the errors in the iterates  $x_k$  of Algorithm P-SIRT. Write the singular value decomposition (SVD) of  $M^{1/2}A$  as

$$(3.1) \quad M^{1/2}A = U \Sigma V^T,$$

where  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p, 0, \dots, 0) \in \mathbb{R}^{m \times n}$  with  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p > 0$ , and  $p$  is the rank of  $A$ . Then  $B \equiv A^T M A = (M^{1/2}A)^T (M^{1/2}A) = V \Sigma^T \Sigma V^T$  and  $\rho(B) = \sigma_1^2$ .

Let  $\bar{x}$  be the solution to (2.3) with the noise-free right-hand side,

$$\bar{x} = \underset{x \in \mathcal{C}}{\text{argmin}} \frac{1}{2} \|A x - \bar{b}\|_M^2.$$

Moreover, let  $x_k$  and  $\bar{x}_k$  denote the iterates of Algorithm P-SIRT using the noisy and the noise-free right-hand side, respectively. Then the error in the  $k$ th iterate clearly satisfies  $x_k - \bar{x} = x_k - \bar{x}_k + \bar{x}_k - \bar{x}$ , and therefore

$$(3.2) \quad \|x_k - \bar{x}\| \leq \|x_k - \bar{x}_k\| + \|\bar{x}_k - \bar{x}\|.$$

Hence the error decomposes into two components—the *noise error*  $x_k - \bar{x}_k$  and the *iteration error*  $\bar{x}_k - \bar{x}$  (in [19, p. 157], where unprojected Landweber iteration is treated, they are called “approximation error” and “data error”). It is the interplay between these two error-terms that explains the semiconvergence of the method. We split the analysis of these errors into two cases, depending on the rank of  $A$ , and we start with the case  $\text{rank}(A) = n$ .

**3.1. The full column-rank problem.** First we present an elementary result that is useful for our analysis.

LEMMA 3.1. *Let  $q = \|I - \lambda B\|$ . Assume that  $\text{rank}(A) = p = n$  and  $\sigma_1 > \sqrt{2}\sigma_n$ . Further assume that  $\lambda$  fulfills  $0 < \epsilon \leq \lambda \leq (2 - \epsilon)/\sigma_1^2$ . Then*

$$(3.3) \quad q = \begin{cases} 1 - \lambda \sigma_n^2, & 0 < \lambda \leq \frac{2}{\sigma_1^2 + \sigma_n^2}, \\ \lambda \sigma_1^2 - 1, & \frac{2}{\sigma_1^2 + \sigma_n^2} \leq \lambda < 2/\sigma_1^2. \end{cases}$$

*Proof.* Let  $a = 1 - \lambda \sigma_n^2$  and  $b = 1 - \lambda \sigma_1^2$ . Using the SVD, one finds  $q = \|I - \lambda B\| = \max(|a|, |b|)$ . We need to check four possible sign combinations of  $a$  and  $b$ .

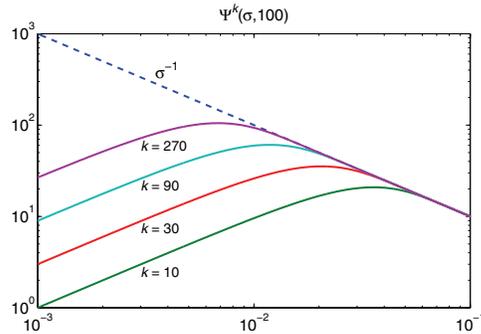


FIG. 3.1. The function  $\Psi^k(\sigma, \lambda)$  as a function of  $\sigma$  for  $\lambda = 100$  and  $k = 10, 30, 90,$  and  $270$ . The dashed line shows  $1/\sigma$ .

- The case  $(+, +)$ . Then  $0 < \lambda \leq 1/\sigma_1^2$  ( $< 1/\sigma_n^2$ ) and we find  $q = a$ .
- The case  $(-, -)$ . Then  $\lambda \geq 1/\sigma_n^2$  ( $> 1/\sigma_1^2$ ). Using Theorem 2.1, we require that  $1/\sigma_n^2 < 2/\sigma_1^2$ , which gives  $\sigma_1 < \sqrt{2}\sigma_n$ , contradicting the assumptions of this lemma.
- The case  $(-, +)$ . Then  $\lambda \geq 1/\sigma_n^2$  and  $\lambda \leq 1/\sigma_1^2$ , which is a contradiction.
- The case  $(+, -)$ . Then  $0 < \lambda \leq 1/\sigma_n^2$  and  $\lambda \geq 1/\sigma_1^2$ . Assume first that  $q = a$  so that  $a > -b$ , which implies  $1/\sigma_1^2 \leq \lambda \leq 2/(\sigma_1^2 + \sigma_n^2)$  ( $< 1/\sigma_n^2$ ). Next, if  $q = -b$  so that  $-b > a$ , one gets  $2/(\sigma_1^2 + \sigma_n^2) \leq \lambda \leq 1/\sigma_n^2$ .  $\square$

*Remark 3.2.* The above results may be obtained by a single graph of  $q$ ; see Figure 4.4 in [30].

For inverse problems  $\sigma_n \ll \sigma_1$  and hence  $2/(\sigma_1^2 + \sigma_n^2) \approx 2/\sigma_1^2$ . Therefore, we will consider only the case  $q = a = 1 - \lambda\sigma_n^2$  in what follows.

In our analysis we will assume that the sequence of relaxation parameters is nonnegative and nonascending:

$$(3.4) \quad 0 < \lambda_{i+1} \leq \lambda_i.$$

The following theorem considers the noise-error, while the iteration error is considered in Theorem 3.8; see also Figure 3.1.

**THEOREM 3.3.** *The noise-error in Algorithm P-SIRT is bounded above by*

$$(3.5) \quad \|x_k - \bar{x}_k\| \leq \frac{\sigma_1 \lambda_0}{\sigma_n \lambda_{k-1}} \Psi^k(\sigma_n, \lambda_{k-1}) \|M^{1/2} \delta b\|,$$

with

$$(3.6) \quad \Psi^k(\sigma, \lambda) \equiv \frac{1 - (1 - \lambda\sigma^2)^k}{\sigma}.$$

*Proof.* We have

$$x_k - \bar{x}_k = P_C(x_{k-1} + \lambda_{k-1} A^T M(b - A x_{k-1})) - P_C(\bar{x}_{k-1} + \lambda_{k-1} A^T M(\bar{b} - A \bar{x}_{k-1})).$$

Hence, using the nonexpansivity of the projection operator,

$$\|x_k - \bar{x}_k\| \leq \|(I - \lambda_{k-1} B)(x_{k-1} - \bar{x}_{k-1}) + \lambda_{k-1} A^T M \delta b\|.$$

Defining

$$(3.7) \quad e_k^N = \|x_k - \bar{x}_k\|, \quad \delta = \|A^T M \delta b\|, \quad q_k = \|I - \lambda_k B\|,$$

we get

$$(3.8) \quad e_k^N \leq q_{k-1} e_{k-1}^N + \delta \lambda_{k-1}.$$

It follows by induction that

$$e_k^N \leq \prod_{i=0}^{k-1} q_i e_0^N + \delta \sum_{s=1}^{k-1} \lambda_{s-1} \prod_{i=s}^{k-1} q_i + \delta \lambda_{k-1}.$$

Since  $x_0 = \bar{x}_0$ , and using (3.4), it follows that

$$e_k^N \leq \delta \lambda_0 \sum_{s=1}^{k-1} \prod_{i=s}^{k-1} q_i + \delta \lambda_{k-1}.$$

By Lemma 3.1 we have  $q_i = 1 - \lambda_i \sigma_n^2$  when  $0 < \lambda_i < 2/(\sigma_1^2 + \sigma_n^2)$ . Hence, using (3.4), we have  $q_i \leq 1 - \lambda_k \sigma_n^2 = q_k$  for  $k \geq i$ . We stress that this property is crucial throughout our proof. Thus

$$(3.9) \quad e_k^N \leq \delta \lambda_0 \sum_{s=0}^{k-1} q_{k-1}^s$$

$$(3.10) \quad = \delta \lambda_0 \frac{1 - q_{k-1}^k}{1 - q_{k-1}} = \delta \lambda_0 \frac{1 - (1 - \lambda_{k-1} \sigma_n^2)^k}{\lambda_{k-1} \sigma_n^2}.$$

Using the bound  $\delta \leq \sigma_1 \|M^{1/2} \delta b\|$  the upper bound in (3.5) emerges.  $\square$

*Remark 3.4.* A similar bound for the noise-error in the unprojected case, assuming a constant relaxation parameter, was derived in [18, (3.2)], without the factor  $\sigma_1/\sigma_n$ . The reason for the presence of this factor here is that we need to use norms and inequalities to utilize the nonexpansiveness of the projection operator. For the same reason, the bound (3.5) might be quite pessimistic since it now holds for any closed convex set  $\mathcal{C}$ . Another reason is that the estimates used to arrive at (3.9) are rather crude.

*Remark 3.5.* When  $\lambda_k = \lambda$  for all  $k \geq 0$ , the bound (3.5) becomes

$$(3.11) \quad \|x_k - \bar{x}_k\| \leq \frac{\sigma_1}{\sigma_n} \Psi^k(\sigma_n, \lambda) \|M^{1/2} \delta b\|,$$

and for  $\lambda \sigma_n^2 \ll 1$  we have

$$(3.12) \quad \|x_k - \bar{x}_k\| \approx \lambda k \sigma_1 \|M^{1/2} \delta b\|,$$

showing that  $k$  and  $\lambda$  play the same role in suppressing the noise. The same observation is made in [1, p. 145] for Algorithm SIRT, and the observation also applies to (3.13) in the previous remark.

*Remark 3.6.* If  $A$  is a bounded linear operator mapping from an infinite dimensional Hilbert space to another infinite dimensional Hilbert space, so that the range of  $A$  is not closed (e.g.,  $A$  is a compact operator), then  $\sigma_n \rightarrow 0$  for  $n \rightarrow \infty$ . Eicke [16, Thm. 3.4], using a constant relaxation parameter  $\lambda$ , derives the following bound:

$$(3.13) \quad \|x_k - \bar{x}_k\| \leq \lambda \delta k, \quad \delta = \|A^T M \delta b\|.$$

TABLE 3.1

The unique root  $\zeta_k \in (0, 1)$  of  $g_{k-1}(y) = 0$  (cf. (3.14)) as a function of the iteration index  $k$ .

$k$	$\zeta_k$								
2	0.3333	7	0.8156	12	0.8936	17	0.9252	22	0.9424
3	0.5583	8	0.8392	13	0.9019	18	0.9294	23	0.9449
4	0.6719	9	0.8574	14	0.9090	19	0.9332	24	0.9472
5	0.7394	10	0.8719	15	0.9151	20	0.9366	25	0.9493
6	0.7840	11	0.8837	16	0.9205	21	0.9396	26	0.9513
								31	0.9592

We note that when  $\lambda_k = \lambda$ , (3.9) in the above proof becomes

$$e_k^N \leq \delta \lambda \sum_{s=0}^{k-1} (1 - \lambda \sigma_n^2)^s,$$

and hence  $\lim \|x_k - \bar{x}_k\| = \delta \lambda k$  as  $\sigma_n \rightarrow 0$  similar to (3.13). Our bounds (3.5) and (3.11) can be seen as extensions of this result for  $\sigma_n > 0$  and, in the case of (3.5), for a nonstationary algorithm.

Next we give an alternative upper bound for the noise error, which we will use in section 5.3 to give bounds for two specific relaxation strategies that satisfy (3.4). Following [18], we consider the equation

$$(3.14) \quad g_{k-1}(y) = (2k-1)y^{k-1} - (y^{k-2} + \dots + y + 1) = 0,$$

which has a unique real root  $\zeta_k \in (0, 1)$ . The roots satisfy  $0 < \zeta_k < \zeta_{k+1} < 1$  and  $\lim_{k \rightarrow \infty} \zeta_k = 1$  (see [18, Propositions 2.3, 2.4]), and they can easily be precalculated; see Table 3.1.

**THEOREM 3.7.** *Assume that  $\sigma_1 \leq 1/\sqrt{\lambda_{k-1}}$ ; then*

$$(3.15) \quad \|x_k - \bar{x}_k\| \leq \frac{\sigma_1 \lambda_0}{\sigma_n \sqrt{\lambda_{k-1}}} \frac{1 - \zeta_k^k}{\sqrt{1 - \zeta_k}} \|M^{1/2} \delta b\|,$$

where  $\zeta_k$  is the unique root in  $(0, 1)$  of (3.14).

*Proof.* Using [18, Proposition 2.3], we obtain the following bound for the function  $\Psi^k(\sigma, \lambda)$  appearing in (3.5):

$$(3.16) \quad \begin{aligned} \max_{1 \leq i \leq n} \Psi^k(\sigma_i, \lambda_{k-1}) &\leq \max_{0 < \sigma \leq \sigma_1} \Psi^k(\sigma, \lambda_{k-1}) \\ &\leq \max_{0 < \sigma \leq 1/\sqrt{\lambda_{k-1}}} \Psi^k(\sigma, \lambda_{k-1}) \leq \sqrt{\lambda_{k-1}} \frac{1 - \zeta_k^k}{\sqrt{1 - \zeta_k}}. \end{aligned}$$

The assumption in the theorem implies

$$(3.17) \quad \sigma_1 \leq 1/\sqrt{\lambda_{k-1}} \Leftrightarrow \lambda_{k-1} \leq 1/\sigma_1^2.$$

Then by (3.5) and (3.16), and assuming (3.17), we obtain the bound in (3.15).  $\square$

We note that the case  $\lambda_{k-1} \in (1/\sigma_1^2, 2/\sigma_1^2)$  is discussed in [18, Remark 2.2]. Next we give a bound for the iteration error, again assuming that (3.4) holds.

**THEOREM 3.8.** *The iteration error in Algorithm P-SIRT is bounded above by*

$$(3.18) \quad \|\bar{x}_k - \bar{x}\| \leq \sigma_n \Phi^k(\sigma_n, \lambda_{k-1}) \|x_0 - \bar{x}\|,$$

where  $x_0$  is the initial vector and

$$(3.19) \quad \Phi^k(\sigma, \lambda) \equiv \frac{(1 - \lambda \sigma^2)^k}{\sigma}.$$

*Proof.* We use the following characterization of the constrained solution (see, e.g., [5, Proposition 3.3b]):

$$(3.20) \quad \bar{x} = P_{\mathcal{C}}(\bar{x} + \lambda A^T M(\bar{b} - A\bar{x})), \quad \lambda > 0.$$

Then (again using the nonexpansivity of the projection operator)

$$\|\bar{x}_k - \bar{x}\| \leq \|\bar{x}_{k-1} - \bar{x} - \lambda_{k-1} B(\bar{x}_{k-1} - \bar{x}) + (\lambda_{k-1} - \lambda) A^T M(\bar{b} - A\bar{x})\|.$$

With

$$e_k^{\text{it}} = \|\bar{x}_k - \bar{x}\|, \quad \mu_k = |\lambda_k - \lambda|, \quad c = \|A^T M(\bar{b} - A\bar{x})\|,$$

this can be written

$$(3.21) \quad e_k^{\text{it}} \leq q_{k-1} e_{k-1}^{\text{it}} + \mu_{k-1} c,$$

where  $q_k$  is defined in (3.7). Since (3.20) holds for any  $\lambda > 0$  (so that  $\bar{x}$  is independent of  $\lambda$ ), we may use a different value of  $\lambda$  in  $\mu_k$  for any index  $k \geq 0$ , so that  $\mu_k = 0$ ,  $k \geq 0$ . It follows from (3.21) that

$$e_k^{\text{it}} \leq q_{k-1}^k e_0^{\text{it}} = (1 - \lambda_{k-1} \sigma_n^2)^k \|\bar{x}_0 - \bar{x}\|,$$

and since  $\bar{x}_0 = x_0$ , this completes the proof.  $\square$

Figure 3.2 shows the actual noise-error and iteration error for a certain test problem (see section 6) and three noise levels for Algorithm P-SIRT with  $M = M_{\mathcal{C}}$  (Cimmino) with the optimal fixed  $\lambda$  (as defined in section 5.1). The dashed lines are proportional to the upper bounds in (3.11) and (3.18) except that we have *omitted* the factors  $\sigma_1/\sigma_n$  and  $\|x_0 - \bar{x}\|$ , which make the bounds unrealistically pessimistic. The important observation is that, after some initial iterations, our bounds track the behavior of the two types of errors, illustrating that our analysis (except for unrealistic scaling factors) indeed describes the actual behavior of the errors.

Note that the iteration error, in the top part of the figure, increases after about 30 iterations; this is caused by the system being inconsistent. For comparison purposes, the bottom part of the figure shows the behavior when we solve a consistent system with the same matrix and a modified right-hand side.

**3.2. The rank-deficient problem.** In tomographic reconstruction applications the matrix often fails to fulfill the assumption  $\text{rank}(A) = n$  needed in Lemma 3.1, and an important case is the underdetermined problem with  $m < n$ . In order to analyze Algorithm P-SIRT for such problems, we consider the following slightly modified problem, in the form of a regularized problem with regularization parameter  $\alpha$  and “balancing parameter”  $\mu = \|M\|$ :

$$(3.22) \quad \min_{x \in \mathcal{C}} \frac{1}{2} \left( \|Ax - b\|_M^2 + \alpha^2 \mu \|x\|^2 \right) = \min_{x \in \mathcal{C}} \frac{1}{2} \left\| \begin{pmatrix} A \\ \alpha I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_{\widehat{M}}^2,$$

where  $\widehat{M} = \text{diag}(M, \mu I)$ . The factor  $\mu$  is introduced to balance the terms  $\|Ax - b\|_M^2$  and  $\mu \|x\|^2$ ; for example,  $\mu = 1$  for Landweber and  $\mu = m^{-1} / \min_i \|a_i\|^2$  for Cimmino. Note that the negative gradient of the objective function in (3.22) is  $A^T M(b - Ax) - \alpha^2 \mu x$ . Hence one step of Algorithm P-SIRT takes the form

$$(3.23) \quad x_{k+1} = P_{\mathcal{C}} \left( x_k + \lambda_k (A^T M(b - Ax_k) - \alpha^2 \mu x_k) \right),$$

which is a standard projected SIRT step with a small correction.

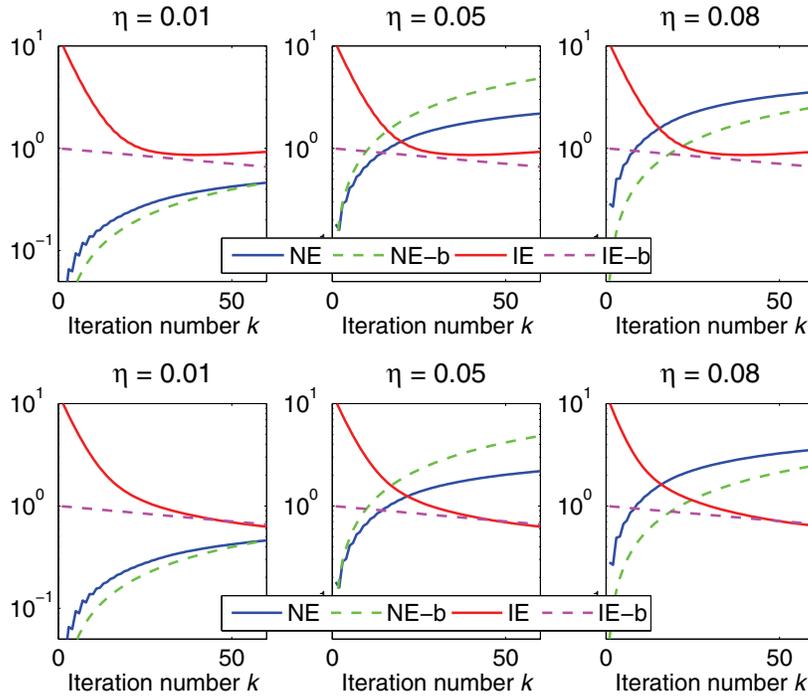


FIG. 3.2. Comparison of actual errors with our bounds for three different noise levels  $\eta = \|\delta b\|/\|\bar{b}\|$ . NE = noise-error; NE-b = our bound (3.11) with the factor  $\sigma_1/\sigma_n$  omitted; IE = iterations error; IE-b = our bound (3.18) with the factor  $\|x_0 - \bar{x}\|$  omitted. Top: Results for an inconsistent system of equations. Bottom: Similar results for a consistent system with the same coefficient matrix.

To analyze the noise- and iteration errors for the modified iteration in (3.22), we first consider the following error formula for the noise-error, which is derived similarly to (3.8):

$$\|x_k - \bar{x}_k\| \leq \hat{q}_{k-1} \|x_{k-1} - \bar{x}_{k-1}\| + \delta \lambda_{k-1},$$

in which  $\hat{q}_k = \|(1 - \lambda_k \alpha^2 \mu)I - \lambda_k B\|$ . We next relate  $\hat{q}_k$  to the SVD (3.1) by considering an extension of Lemma 3.1.

LEMMA 3.9. Let  $\hat{q} = \|(1 - \lambda \alpha^2 \mu)I - \lambda B\|$  and  $\alpha^2 \mu = \sigma_p^2$ , and assume that  $\sigma_1 > \sqrt{3}\sigma_p$ , where  $p = \text{rank}(A)$ . Also assume that  $\lambda$  fulfills the condition  $0 < \epsilon \leq \lambda \leq 2/(\sigma_1^2 + \sigma_p^2)$ . Then

$$(3.24) \quad \hat{q} = \begin{cases} 1 - \lambda \sigma_p^2, & 0 < \lambda \leq \frac{2}{\sigma_1^2 + 2\sigma_p^2}, \\ \lambda(\sigma_1^2 + \sigma_p^2) - 1, & \frac{2}{\sigma_1^2 + 2\sigma_p^2} \leq \lambda < \frac{2}{\sigma_1^2 + \sigma_p^2}. \end{cases}$$

*Proof.* Using the SVD, we find

$$\hat{q} = \max(|\hat{a}|, |\hat{b}|, |\hat{c}|), \quad \hat{a} = 1 - \lambda \alpha^2 \mu - \lambda \sigma_p^2, \quad \hat{b} = 1 - \lambda \alpha^2 \mu - \lambda \sigma_1^2, \quad \hat{c} = 1 - \lambda \alpha^2 \mu.$$

We need to check the eight possible sign combinations of  $(\hat{a}, \hat{b}, \hat{c})$ .

- The case  $(+, +, +)$ . Then  $\hat{a} \geq 0$  gives  $\lambda \leq 1/(2\sigma_p^2)$ ,  $\hat{b} \geq 0$  gives  $\lambda \leq 1/(\sigma_1^2 + \sigma_p^2)$ , and  $\hat{c} \geq 0$  gives  $\lambda \leq 1/\sigma_p^2$ . We conclude that for this case  $0 < \lambda \leq 1/(\sigma_1^2 + \sigma_p^2)$ . Obviously  $q = \hat{c}$  holds.
- The case  $(-, -, -)$ . Then  $\lambda \geq 1/\sigma_p^2$ . Hence by the convergence assumptions we must have  $2/(\sigma_1^2 + \sigma_p^2) > 1/\sigma_p^2$ , which implies  $\sigma_1 < \sigma_p$ , which is a contradiction.

We next note that if  $\hat{b} > 0$ , then also  $\hat{a}, \hat{c} > 0$ , so this case is already treated. Thus three cases remain:  $(+, -, +)$ ,  $(+, -, -)$ , and  $(-, -, +)$ .

- The case  $(+, -, +)$ . Using the sign conditions on  $\hat{a}, \hat{b}, \hat{c}$  and the upper bound in the convergence conditions, one finds  $1/(\sigma_1^2 + \sigma_p^2) \leq \lambda < 2/(\sigma_1^2 + \sigma_p^2)$ . Now  $\hat{q} = \hat{c}$  or  $\hat{q} = -\hat{b}$ . In the first subcase  $\hat{c} \geq -\hat{b}$ , which in turn implies  $0 < \lambda \leq 2/(\sigma_1^2 + 2\sigma_p^2)$ . The second subcase  $\hat{q} = -\hat{b}$  implies  $-\hat{b} > \hat{c}$ , which in turn implies  $\lambda \geq 2/(\sigma_1^2 + 2\sigma_p^2)$ .
- The case  $(+, -, -)$ . Then we get  $\lambda < 1/(2\sigma_p^2), \lambda > 1/\sigma_p^2$ , which is a contradiction.
- The case  $(-, -, +)$ . The sign conditions lead to  $1/(2\sigma_p^2) \leq \lambda < 1/\sigma_p^2$ . Hence it must hold that  $1/(2\sigma_p^2) \leq 2/(\sigma_1^2 + \sigma_p^2)$ , leading to  $\sigma_1 \leq \sqrt{3}\sigma_p$ , which is a contradiction.  $\square$

Using Lemma 3.9, we can now repeat the steps in the semiconvergence analysis for the unregularized case, and we arrive at the *same* expressions for the noise- and iteration errors, respectively, when applying Algorithm P-SIRT to problem (2.3) and problem (3.22).

In practice, our computational experience with Algorithm P-SIRT (2.2) shows that for underdetermined and/or rank-deficient problems the unregularized algorithm gives results that are indistinguishable from those of the regularized algorithm. If one wants a guarantee that semiconvergence takes place, however, then one should use the regularized step from (3.23) in the projected algorithm.

**4. The scaled gradient projection method.** We will now extend the above analysis of the semiconvergence properties to the scaled gradient projection method. Following [4], we define  $y_k = S^{-1/2}x_k$  and  $\bar{A} = AS^{1/2}$  and get

$$x_k + \lambda_k SA^T M(b - Ax_k) = S^{1/2}(y_k + \lambda_k \bar{A}^T M(b - \bar{A}y_k)).$$

Assuming that

$$(4.1) \quad P_{\mathcal{C}}(Su) = SP_{\mathcal{C}}(u) \quad \text{for any } u \in \mathbb{R}^n$$

the iteration (2.4) becomes

$$(4.2) \quad y_{k+1} = P_{\mathcal{C}}(y_k + \lambda_k \bar{A}^T M(b - \bar{A}y_k)), \quad k = 0, 1, 2, \dots$$

Therefore, the convergence results for the unscaled gradient projection method apply to the sequence  $\{y_k\}$ . We next observe that if  $y_k \rightarrow y^*$ , then  $x_k \rightarrow S^{1/2}y^*$  for  $k \rightarrow \infty$ . Hence with  $x^* = S^{1/2}y^*$  it follows that  $\|\bar{A}y^* - b\|_M = \|Ax^* - b\|_M$ . This justifies the use of (2.4) for solving problem (2.3).

Assumption (4.1) is satisfied, e.g., when  $S$  is diagonal with positive elements (as in DROP) and  $\mathcal{C}$  represents nonnegativity constraints. If  $\mathcal{C}$  corresponds to box constraints, then the assumption is, in general, not fulfilled; in this case one can use the CAV method [11], which, for the unconstrained case, shows a behavior similar to that of DROP.

Our semiconvergence analysis can easily be extended to the iteration in (2.4). For the noise-error we get

$$\|x_k - \bar{x}_k\| \leq \|(I - \lambda_{k-1}SA^TMA)(x_{k-1} - \bar{x}_{k-1}) + \lambda_{k-1}SA^TM\delta b\|,$$

and then we should take

$$\begin{aligned} q_k &\equiv \|I - \lambda_kSA^TMA\| \\ &= \|S^{1/2}(I - \lambda_kS^{1/2}A^TMA S^{1/2})S^{-1/2}\| \\ &= \|I - \lambda_kS^{1/2}A^TMA S^{1/2}\|. \end{aligned}$$

Now we can use the SVD of  $M^{1/2}A S^{1/2}$  (rather than (3.1)), and Lemmas 3.1 and 3.9 will be valid.

**5. Choice of relaxation parameters.** Several strategies for picking the relaxation parameter (or step-length)  $\lambda_k$  have been proposed, and below we discuss the use of some of these strategies for Algorithm P-SIRT. Other choices, e.g., of adaptive type, can be found in [4] together with corresponding convergence results. A state-of-the-art study of step-length selection for nonnegativity constraints is given in [3], where scaled gradient projection methods, such as those in [6] and [23], are also considered.

**5.1. Optimal fixed parameter.** One strategy is the optimal choice strategy: this means finding that constant value of  $\lambda$  which gives rise to the fastest convergence to the smallest relative error in the solution. The value of  $\lambda$  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 [21] and [26].

**5.2. De Pierro and Dos Santos (DPDS) strategy.** This strategy is based on picking  $\lambda_k$  such that the error  $\|x^* - x_k\|$  is minimized in each iteration, where  $x^*$  is any solution to  $Ax = b$  which is assumed to be consistent. For the case  $S = I$  we obtain

$$(5.1) \quad \lambda_k = \frac{r_k^T M r_k}{\|A^T M r_k\|^2}, \quad r_k = b - Ax_k.$$

This case was investigated by De Pierro [14] for linear systems (using Cimmino iteration) and Dos Santos [15] for general convex objective functions, and we refer to (5.1) as the *DPDS strategy*. Let  $Q \neq \emptyset$  denote the set

$$(5.2) \quad Q = \{x \mid Ax = b, x \in \mathcal{C}\}.$$

It follows from [15, Theorem 1] (after some adaption to our case) that  $x_k \rightarrow x^* \in Q$  as  $k \rightarrow \infty$ . So in this case we are really solving a convex feasibility problem rather than a constrained optimization problem.

Next consider the case  $S \neq I$ . We first derive the relaxation parameter for the unconstrained case (which to our knowledge was not done previously). To do this we consider (cf. (4.2) without  $P_C$ ) the step

$$y_{k+1} = y_k + \mu_k p_k \quad \text{with} \quad p_k = \bar{A}^T M (b - \bar{A} y_k).$$

Let  $y^* = S^{-1/2}x^*$  such that  $\bar{A}y^* = b$ . Now apply the DPDS principle  $\min \|y_{k+1} - y^*\|$  again, giving

$$\mu_k = p_k^T (y^* - y_k) / \|p_k\|^2.$$

It follows after some calculations (and transforming back to  $x_k$ ) that

$$\mu_k = \frac{r_k^T M r_k}{\|A^T M r_k\|_S^2},$$

with  $r_k = b - A x_k$ . We may now use (2.4) with the above choice  $\lambda_k = \mu_k$ . Assuming that both (4.1) and (5.2) are satisfied, convergence toward a point in the set  $Q$  is ensured. To our knowledge no results on semiconvergence using the DPDS strategy are known.

**5.3. Diminishing stepsize strategy.** For the diminishing stepsize strategy, the relaxation parameter must fulfill

$$\lambda_k > 0 \quad \text{and} \quad \lambda_k \rightarrow 0 \quad \text{such that} \quad \sum_k \lambda_k = \infty.$$

This strategy is described together with convergence results in [4, pp. 43, 207–208] and [29, p. 140]. Here we consider two diminishing stepsize strategies that were recently proposed and analyzed in [18] for the unprojected SIRT algorithm. The idea behind these strategies is to monitor and control the noise-error, and we refer the reader to [18, pp. 328–329] for a motivation for this. Since we now have derived error bounds for Algorithm P-SIRT that are similar to those for Algorithm SIRT, it is natural to also use these in the constrained case.

We propose the following two rules for picking relaxation parameters in Algorithm P-SIRT, where  $\zeta_k$  are the roots of (3.14):

$$(5.3) \quad \Psi_1\text{-rule :} \quad \lambda_k = \begin{cases} \sqrt{2}\sigma_1^{-2} & \text{for } k = 0, 1, \\ 2\sigma_1^{-2}(1 - \zeta_k) & \text{for } k \geq 2, \end{cases}$$

$$(5.4) \quad \Psi_2\text{-rule :} \quad \lambda_k = \begin{cases} \sqrt{2}\sigma_1^{-2} & \text{for } k = 0, 1, \\ 2\sigma_1^{-2} \frac{1 - \zeta_k}{(1 - \zeta_k^k)^2} & \text{for } k \geq 2. \end{cases}$$

Both these strategies are of diminishing stepsize type [18, Propositions 3.1 and 3.3]. In our numerical tests we found that (5.4) usually gives faster convergence than (5.3).

We next bound the noise-error for the two strategies. First we check the assumption (3.17). Using the numerical values from Table 3.1, we find that (3.17) is satisfied when  $k \geq 3$  and  $k \geq 6$  for strategies  $\Psi_1$  and  $\Psi_2$ , respectively. Hence we get

$$(5.5) \quad \Psi_1\text{-rule :} \quad \|x_k - \bar{x}_k\| \leq \frac{\sigma_1^2 \lambda_0}{\sigma_n \sqrt{2}} \cdot \frac{1 - \zeta_k^k}{1 - \zeta_k} \|M^{1/2} \delta b\|, \quad k \geq 3,$$

and

$$(5.6) \quad \Psi_2\text{-rule :} \quad \|x_k - \bar{x}_k\| \leq \frac{\sigma_1^2 \lambda_0}{\sigma_n \sqrt{2}} \cdot \frac{(1 - \zeta_k^k)^2}{1 - \zeta_k} \|M^{1/2} \delta b\|, \quad k \geq 6,$$

where we have used  $1/\sqrt{\lambda_{k-1}} \leq 1/\sqrt{\lambda_k}$  to get simpler formulas. Since

$$\frac{(1 - \zeta_k^k)^2}{1 - \zeta_k} \leq \frac{1 - \zeta_k^k}{1 - \zeta_k} = 1 + \zeta_k + \zeta_k^2 + \cdots + \zeta_k^{k-1} \leq k,$$

we see that strategy  $\Psi_2$  has a lower upper bound for the noise-error than strategy  $\Psi_1$ , and both can (rather crudely) be bounded by  $k$ .

In [18] we also proposed accelerating the above two strategies by replacing  $\lambda_k$  with  $\sqrt{2}\sigma_1^{-1}$  for  $k < k_0$  and  $\tau\lambda_k$  for  $k \geq k_0$ , where  $\tau$  and  $k_0$  are parameters to be chosen. Specifically, we have

$$(5.7) \quad \text{modified } \Psi_1\text{-rule:} \quad \lambda_k = \tau \frac{2}{\sigma_1^2} (1 - \zeta_k), \quad k \geq k_0,$$

with  $0 < \epsilon_1 \leq \tau < (1 - \zeta_k)^{-1}$  for  $k \geq k_0$ , and

$$(5.8) \quad \text{modified } \Psi_2\text{-rule:} \quad \lambda_k = \tau \frac{2}{\sigma_1^2} \frac{1 - \zeta_k}{(1 - \zeta_k^k)^2}, \quad k \geq k_0,$$

with  $0 < \epsilon_2 \leq \tau < (1 - \zeta_k^k)^2 / (1 - \zeta_k)$  for  $k \geq k_0$ . The reason for the bounds on  $\tau$  is that they will guarantee (as is easily seen) that the corresponding modified relaxation parameters are in the appropriate interval  $(0, 2/\sigma_1^2)$ . In this paper we use  $k_0 = 2$  and  $\tau = 2$  for strategy  $\Psi_1$  and  $\tau = 1.5$  for strategy  $\Psi_2$ . The modified strategies also belong to the diminishing step-length type; for a proof see [18].

**6. Computational results.** We report on some numerical tests with an example taken from the field of tomographic image reconstruction from projections, using the SNARK93 software package [8] and the standard head phantom from Herman [22]. This software generates a sparse matrix  $A$ , an exact solution  $\bar{x}$  (which represents the “phantom”—see Figure 6.2), and a right-hand side  $\bar{b}$  that represents the noise-free data.

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 inverse problems.

Our phantom has size  $63 \times 63$ , leading to an exact solution of length  $n = 3969$ , and we add Gaussian noise  $\delta b$  to the exact data scaled such that  $\eta = \|\delta b\|/\|\bar{b}\|$  equals 0.01, 0.05, and 0.08. As the set  $\mathcal{C}$  we use the positive orthant such that we enforce nonnegativity on the reconstructions. We use Algorithm P-SIRT and the projected scaled gradient projection method (2.4), and for both methods we compute and plot the relative errors  $\|\bar{x} - x_k\|/\|\bar{x}\|$  for  $k = 1, 2, \dots, 100$ .

All computations are done with the AIR Tools software package [21]. Note that the parameter-choice strategies need an estimate of the spectral radius, and we use a few Lanczos iterations to estimate  $\rho(A^T M_C A)$  and  $\rho(S A^T M_C A)$ , as implemented in [21]. This approach is fast because for our matrices with nonnegative elements, by experience, the largest eigenvalue is well separated from the others.

In our first experiment we use an overdetermined problem in which the number of nonzero rows of the matrix is  $m = 5430$ , and we checked that this matrix has full rank. Figure 6.1 shows the relative error histories for two projected methods, P-Cimmino and P-DROP, using the optimal fixed relaxation parameter as well as the DPDS strategy (5.1) and the modified  $\Psi_1$  and  $\Psi_2$  strategies (5.7)–(5.8). These curves clearly show the semiconvergence of the iterations, as expected from our analysis. The zigzagging behavior of the DPDS strategy (for the unconstrained case) was also noted in [13, pp. 497 and 504] and in [18]; the reason seems to be that this strategy assumes consistent data.

For the fixed- $\lambda$  case we clearly see the semiconvergence of the iterations, as expected from our analysis. For the modified  $\Psi_1$  and  $\Psi_2$  strategies we also have semiconvergence to the same minimum error, but the strict minimum is attained after a much larger number of iterations; see Table 6.1. The phantom and the reconstructions are shown in Figure 6.2.

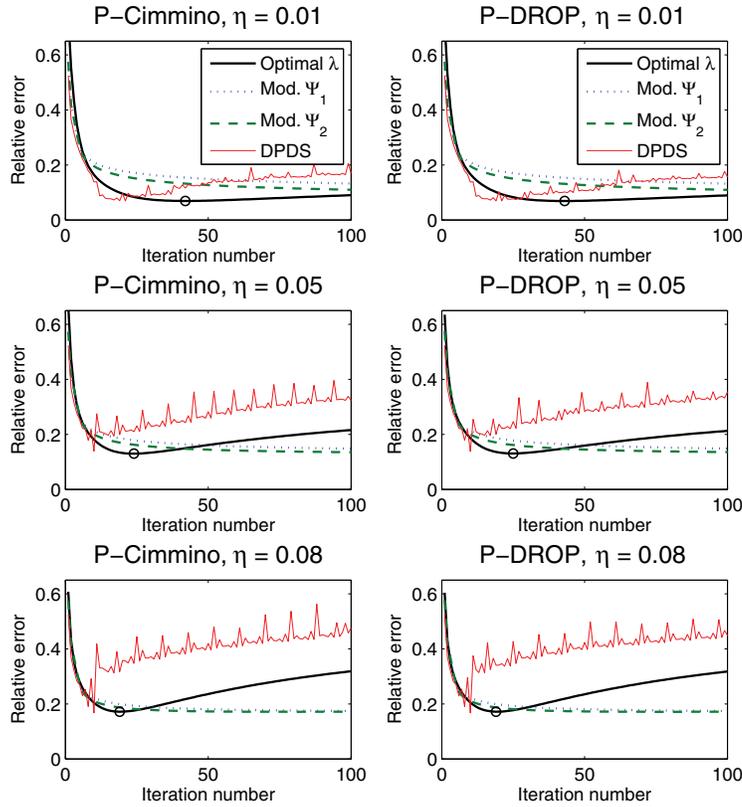


FIG. 6.1. Relative error histories  $\|\bar{x} - x_k\|_2 / \|\bar{x}\|_2$  for the overdetermined system, for three noise levels. The black circles denote the minimum error.

TABLE 6.1

The number of iterations needed to reach the minimum reconstruction error, and the minimum relative error at that point for the overdetermined problem.

	$\eta = 0.01$		$\eta = 0.05$		$\eta = 0.08$	
	iterations	min	iterations	min	iterations	min
P-Cimmino						
optimal $\lambda$	42	0.0692	22	0.1306	16	0.1709
mod. $\Psi_1$	>50,000	—	2322	0.1317	264	0.1726
mod. $\Psi_2$	$\approx 48,500$	0.0693	358	0.1316	81	0.1723
DPDS	21	0.0691	10	0.1387	10	0.1674
P-DROP						
optimal $\lambda$	43	0.0990	22	0.1308	16	0.1710
mod. $\Psi_1$	>50,000	—	2308	0.1316	262	0.1726
mod. $\Psi_2$	$\approx 47,000$	0.0691	357	0.1316	81	0.1724
DPDS	21	0.0691	10	0.1384	10	0.1673

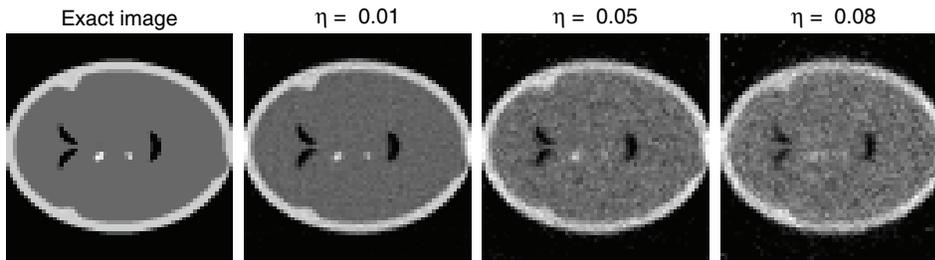


FIG. 6.2. The exact image (the phantom) and the three optimal reconstructions for the overdetermined problem (the difference between the reconstructions from the different methods is indistinguishable).

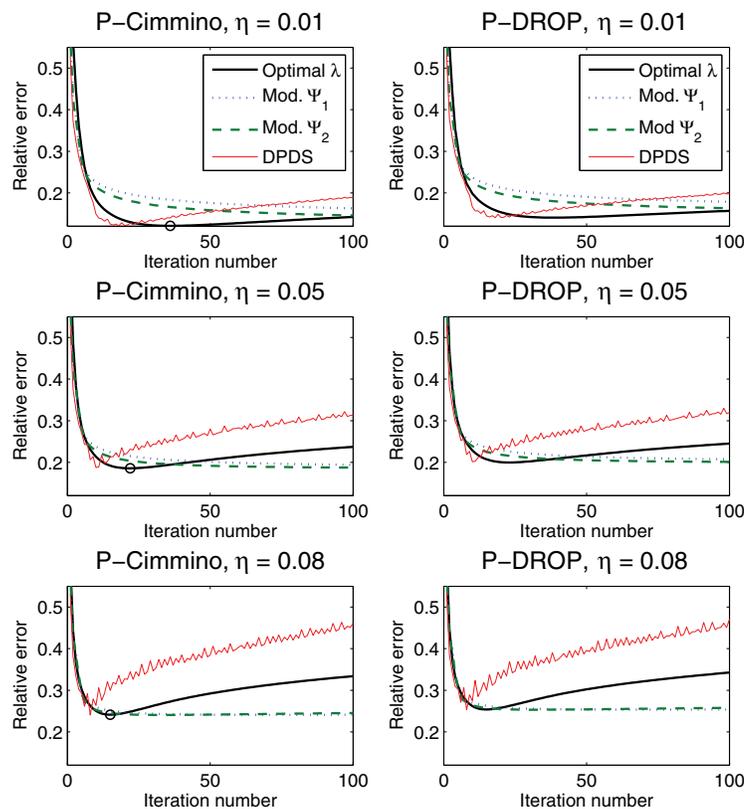


FIG. 6.3. Relative error histories for the underdetermined system.

The important observation here is that for the larger noise levels the errors for the modified  $\Psi_1$  and  $\Psi_2$  strategies exhibit a fast-converging initial phase following the errors associated with the optimal  $\lambda$ . When they reach an error level close to the minimum, they enter the second phase, where the convergence slows down. This demonstrates that our diminishing stepsize strategy is able to control the iterations as desired in such a way that the initial convergence is as fast while convergence slows down approximately when the minimum error is reached.

In our second experiment we use a highly underdetermined system with the same

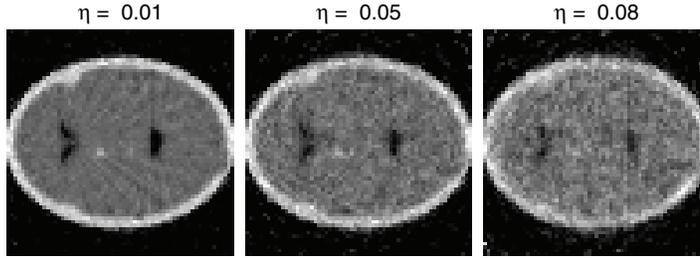


FIG. 6.4. The three optimal reconstructions for the underdetermined problem.

TABLE 6.2  
Same as Table 6.1 for the underdetermined problem.

	$\eta = 0.01$		$\eta = 0.05$		$\eta = 0.08$	
	iterations	min	iterations	min	iterations	min
P-Cimmino						
optimal $\lambda$	36	0.121	19	0.185	12	0.235
mod. $\Psi_1$	>30,000	—	1098	0.187	72	0.239
mod. $\Psi_2$	$\approx 16,000$	0.141	211	0.187	33	0.239
DPDS	18	0.121	10	0.188	8	0.240
P-DROP						
optimal $\lambda$	38	0.141	19	0.200	12	0.248
mod. $\Psi_1$	>30,000	—	1169	0.202	69	0.252
mod. $\Psi_2$	$\approx 23,000$	0.141	220	0.202	32	0.252
DPDS	18	0.141	10	0.201	8	0.253

$63 \times 63$  phantom and  $m = 1376$  nonzero rows in  $A$ , corresponding to measurements with fewer rays in the projections. This examples allows us to demonstrate experimentally that the projected algorithms still show semiconvergence and that our parameter-choice algorithms also work for these systems—in spite of the lack of a full theory. Figure 6.3 shows the same relative error histories as before, and the conclusion is the same as above—even though our theory does not cover this case. The optimal reconstructions are shown in Figure 6.4, and Table 6.2 lists the minimum errors and the number of iterations needed to reach them.

Finally, we return to the underdetermined system, and this time we solve the regularized system (3.22) by means of Cimmino's method for which  $\mu = \|M\| = 0.31$ . As long as the regularization parameter  $\alpha$  is small, we compute essentially the same solutions as those for the unregularized problem. For example, if we choose  $\alpha = \sigma_p/\sqrt{\mu} = 2.5 \cdot 10^{-4}/\sqrt{0.31} = 4.3 \cdot 10^{-4}$  according to Lemma 3.9, then the 2-norm of the difference between the iterates is always less than  $3 \cdot 10^{-8}$ , and if  $\alpha = 10^{-2}$ , then the norm of the difference is less than  $2 \cdot 10^{-5}$ . We conclude that while incorporating regularization into the problem is a safeguard which ensures that the theory holds, it may not be needed in practice.

**Conclusion.** Using a descending relaxation parameter sequence in Algorithm P-SIRT, we derived a formula for the influence of noise in the right-hand side for two cases. The first case is when  $\text{rank}(A) = n$ , and the second is for a regularized problem with no rank restrictions on  $A$ . In both cases we show that the algorithm exhibits semiconvergence. Further we showed that relaxation strategies, previously defined for unconstrained problems, can also be applied to constrained problems. We also demonstrated the performance of our strategies by examples taken from tomographic imaging.

**Acknowledgment.** We thank the referees of an earlier version of this paper for several comments that led to improvements of the theory as well as the presentation.

## REFERENCES

- [1] M. BERTERO AND P. BOCCACCI, *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, UK, 1998.
- [2] M. BERTERO, D. BINDI, P. BOCCACCI, M. CATTANEO, C. EVA, AND V. LANZA, *Application of the projected Landweber method to the estimation of the source time function in seismology*, *Inverse Problems*, 13 (1997), pp. 465–486.
- [3] F. BENVENUTO, R. ZANELLA, L. ZANNI, AND M. BERTERO, *Nonnegative least-squares image deblurring: Improved gradient projection approaches*, *Inverse Problems*, 26 (2010), 025004.
- [4] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
- [5] D. P. BERTSEKAS AND J. N. TSITSIKLIS, *Parallel and Distributed Computation*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [6] S. BONETTINI, R. ZANELLA, AND L. ZANNI, *A scaled gradient projection method for constrained image deblurring*, *Inverse Problems*, 25 (2009), 015002.
- [7] P. BRIANZI, F. DI BENEDETTO, AND C. ESTATICO, *Improvement of space-invariant image deblurring by preconditioned Landweber iterations*, *SIAM J. Sci. Comput.*, 30 (2008), pp. 1430–1458.
- [8] J. A. BROWNE, G. T. HERMAN, AND D. ODHNER, *SNARK93: A Programming System for Image Reconstruction from Projections*, Technical report MIPG198, The Medical Imaging Processing Group (MIPG), Department of Radiology, University of Pennsylvania, Philadelphia, PA, 1993.
- [9] C. BYRNE, *Iterative oblique projection onto convex sets and the split feasibility problem*, *Inverse Problems*, 18 (2002), pp. 441–453.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] P. L. COMBETTES, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, *IEEE Trans. Image Process.*, 6 (1997), pp. 493–506.
- [14] A. R. DE PIERRO, *Methodos de projeção para a resolução de sistemas gerais de equações algébricas lineares*, Thesis (tese de Doutorado), Instituto de Matemática da UFRJ, Cidade Universitária, Rio de Janeiro, Brasil, 1981.
- [15] L. T. DOS SANTOS, *A parallel subgradient projections method for the convex feasibility problem*, *J. Comput. Appl. Math.*, 18 (1987), pp. 307–320.
- [16] B. EICKE, *Iteration methods for convexly constrained ill-posed problems in Hilbert space*, *Numer. Funct. Anal. Optim.*, 13 (1992), pp. 413–429.
- [17] T. ELFVING, T. NIKAZAD, AND C. POPA, *A class of iterative methods: Semi-convergence, stopping rules, inconsistency, and constraining*, in *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning, and Inverse Problems*, Y. Censor, M. Jiang, and G. Wang, eds., Medical Physics Publishing, Madison, WI, 2010, pp. 157–183.
- [18] T. ELFVING, T. NIKAZAD, AND P. C. HANSEN, *Semi-convergence and relaxation parameters for a class of SIRT algorithms*, *Electron. Trans. Numer. Anal.*, 37 (2010), pp. 321–336.
- [19] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, The Netherlands, 1996.
- [20] P. C. HANSEN, *Discrete Inverse Problems: Insight and Algorithms*, *Fundam. Algorithms* 7, SIAM, Philadelphia, 2010.
- [21] P. C. HANSEN AND M. SAXILD-HANSEN, *AIR Tools: A MATLAB package of algebraic iterative reconstruction methods*, *J. Comput. Appl. Math.*, 236 (2011), pp. 2167–2178.
- [22] G. T. HERMAN, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2nd ed., Springer, New York, 2009.
- [23] B. JOHANSSON, T. ELFVING, V. KOZLOV, Y. CENSOR, P.-E. FORSSÉN, AND G. GRANLUND, *The application of an oblique-projected Landweber method to a model of supervised learning*, *Math. Comput. Modelling*, 43 (2006), pp. 892–909.
- [24] L. LANDWEBER, *An iterative formula for Fredholm integral equations of the first kind*, *Amer. J. Math.*, 73 (1951), pp. 615–624.

- [25] Z.-Q. LUO AND P. TSENG, *Error bound and reduced-gradient projection algorithms for convex minimization over a polyhedral set*, SIAM J. Optim., 3 (1993), pp. 43–59.
- [26] 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 Trans. Image Process., 13 (2004), pp. 549–561.
- [27] F. NATTERER, *The Mathematics of Computerized Tomography*, Classics Appl. Math. 32, SIAM, Philadelphia, 2001.
- [28] M. PIANA AND M. BERTERO, *Projected Landweber method and preconditioning*, Inverse Problems, 13 (1997), pp. 441–464.
- [29] B. POLYAK, *Introduction to Optimization*, Optimization Software, Inc., New York, 1987.
- [30] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
- [31] R. ZDUNEK AND A. CICHOCKI, *Fast nonnegative matrix factorization algorithms using projected gradient approaches for large-scale problems*, Comput. Intell. Neurosci., 2008 (2008), 939567.
- [32] H. ZHANG AND L. Z. CHENG, *Projected Landweber iteration for matrix completion*, J. Comput. Appl. Math., 235 (2010), pp. 593–601.