NONNEGATIVITY AND ITERATIVE METHODS FOR ILL-POSED PROBLEMS

D. CALVETTI [*], G. LANDI [†], L. REICHEL [‡], AND F. SGALLARI [§]

Abstract. The Generalized Minimal RESidual (GMRES) method and Conjugate Gradient method applied to the normal equations (CGNR) are popular iterative schemes for the solution of large linear systems of equations. GMRES requires the matrix of the linear system to be square and nonsingular, while CGNR also can be applied to overdetermined or underdetermined linear systems of equations. When equipped with a suitable stopping rule, both GMRES and CGNR are regularization methods for the solution of linear ill-posed problems. Many linear ill-posed problems that arise in the sciences and engineering have nonnegative solutions. This paper describes iterative schemes, based on the GMRES or CGNR methods, for the computation of nonnegative solutions of linear ill-posed problems. The computations with these schemes are terminated as soon as a nonnegative approximate solution which satisfies the discrepancy principle has been found. Several computed examples illustrate that the schemes of the present paper are able to compute nonnegative approximate solutions of higher quality with less computational effort than several available numerical methods.

AMS subject classifications. 65F10, 65F22

Key words. GMRES, CGNR, nonnegativity, ill-posed problem, regularization, line search

1. Introduction.

The discretization of linear ill-posed problems gives rise to linear systems of equations

$$Ax = \hat{b}, \qquad A \in \mathbb{R}^{m\times n}, \quad \hat{b} \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \tag{1.1}$$

with a matrix A of ill-determined rank, i.e., A has many singular values of different orders of magnitude close to the origin. In particular, A is severely ill-conditioned and may be singular. We refer to linear systems of equations with matrices of this type as linear discrete ill-posed problems. The matrix A is allowed to be nonsymmetric and rectangular. The system (1.1) is assumed to be consistent.

We would like to determine the solution $\hat{x}$ of minimal Euclidean norm of (1.1). The matrix A is assumed to be large enough to make its factorization impossible or undesirable. A solution therefore has to be computed with an iterative method.

In many linear discrete ill-posed problems that arise in the sciences and engineering, the right-hand side vector is contaminated by measurement errors $d \in \mathbb{R}^m$, which we refer to as noise. Let $\hat{b}$ denote the noise-free, but unknown, right-hand side, and let

$$b = \hat{b} + d$$

be the available contaminated right-hand side vector. We assume that the norm

$$\delta = \|d\| \tag{1.2}$$

[*]Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106. E-mail: `dxc57@po.cwru.edu`. Research supported in part by NSF grant DMS-0107841 and NIH grant GM-66309-01.

[†]Dipartimento di Matematica, Universitá di Bologna, Piazza P.ta S. Donato 5, 40127 Bologna, Italy. E-mail: `landig@dm.unibo.it`.

[‡]Department of Mathematical Sciences, Kent State University, Kent, OH 44242. E-mail: `reichel@math.kent.edu`. Research supported in part by NSF grant DMS-0107858.

[§]Dipartimento di Matematica, Universitá di Bologna, Piazza P.ta S. Donato 5, 40127 Bologna, Italy. E-mail: `sgallari@dm.unibo.it`. Research supported by Italian MIUR-Cofin and GNCS-INDAM grants.

is known. Here and throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm.

Since the vector $\hat{b}$ is not available, we cannot solve the linear system (1.1). Instead, we seek to determine an approximation of the solution $\hat{x}$ of (1.1) by computing an approximate solution of the available linear system of equations

$$Ax = b. \tag{1.3}$$

The solution of (1.3), if it exists, generally is not a meaningful approximation of $\hat{x}$, because of the severe ill-conditioning of the matrix A and the presence of the noise d in the right-hand side vector b. Therefore, instead of solving the system (1.3), we replace it by a nearby linear system with a less ill-conditioned matrix, and solve the new linear system so obtained. This replacement is commonly referred to as regularization. We implicitly define and solve a nearby linear system of equations by applying a few steps of an iterative method. This approach to regularization is known as regularization by truncated iteration.

Until recently regularization by truncated iteration has only been applied to linear discrete ill-posed problems with a symmetric matrix. For instance, the Conjugate Gradient (CG) method, and the mathematically equivalent LSQR method by Paige and Saunders [14], have been applied to determine approximate solutions of the normal equations

$$A^T Ax = A^T b \tag{1.4}$$

associated with (1.3); see Björck [2, 3], Hanke [8], Hansen [11, Chapter 6], and references therein. We refer to this solution scheme as the CG Normal equations Residual-minimizing (CGNR) method, independently of the implementation used.

The Generalized Minimal RESidual (GMRES) method [17, 18] is a popular iterative scheme for the solution of large linear systems of equations with a nonsymmetric fairly well conditioned matrix. It has been shown in [5] that regularization by truncated iteration with the GMRES method is meaningful.

In some applications, the desired solution $\hat{x}$ of (1.1) is known to have only nonnegative entries. This is the case, for example, when the entries represent energies or concentrations of chemicals. It then is desirable that the computed approximate solution of (1.3) only has nonnegative entries. Thus, instead of computing an approximate solution of the linear system of equations (1.3), we would like to determine an approximate solution of the constrained minimization problem

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|, \qquad x \geq 0. \tag{1.5}$$

Different strategies for computing regularized nonnegative solutions to linear discrete ill-posed problems have been proposed in the literature. The projected Landweber method is discussed by Bertero and Boccacci [1, Section 6.3]. The use of a nonlinear transformation of the variables, which eliminates the need to explicitly impose the nonnegativity constraints, is described by Hanke et al. [9]. The resulting transformed system of equations is nonlinear and is solved by a quasi-Newton method. The application of the projected steepest descent method, together with a suitable preconditioner, for imposing nonnegativity is advocated by Nagy and Strakos [12], and Rojas and Steihaug [16] present a barrier method, which combines Tikhonov regularization with a constraint on the norm of the computed solution and the use of a logarithmic barrier function.

In this paper we propose new inner-outer iteration schemes for computing an approximate solution of (1.5). The schemes are based on the following general approach. A few steps of an iterative method are applied to the linear system (1.3), or the associated normal equations (1.4), and then the computed approximate solution is projected onto the set of vectors with nonnegative entries. If the projected approximate solution satisfies the discrepancy principle, i.e., if the norm of the associated residual vector is of the order δ, defined by (1.2), then we accept the projected approximate solution as an approximate solution of (1.5); otherwise we refine it by carrying out a few (more) steps with an iterative method. The iterative methods used are GMRES, the closely related Range Restricted GMRES (RRGMRES), and CGNR. For a recent discussion of the discrepancy principle and its justification, we refer to [11, Chapter 7].

This paper is organized as follows. Section 2 reviews the GMRES, RRGMRES, and CGNR iterative methods. A stopping criterion for the iterative methods, based on the discrepancy principle, is also introduced. Section 3 describes two types of schemes for computing nonnegative approximate solutions of (1.5). One type of schemes are projected restarted iterative methods, where before each restart the computed solution is projected onto the set of nonnegative vectors. The other type of schemes accepts the computed nonnegative approximate solution obtained after projection as an initial approximate solution for the next iteration only if the norm of the associated discrepancy is reduced. If this is not the case, a line search is applied to achieve a reduction. Computed examples are presented in Section 4. These examples show the schemes of the present paper to be competitive with recently proposed methods when applied to several standard test problems. Section 5 contains concluding remarks.

2. Iterative methods, discrepancy, and regularization. We review properties of the GMRES, RRGMRES, and CGNR iterative methods. Throughout this paper, we choose the initial approximate solution $x^{(0)} = 0$.

The GMRES and RRGMRES methods are suitable for the solution of linear systems of equations (1.3) with an $n \times n$ nonsingular matrix A. Introduce the Krylov subspaces

$$\mathcal{K}_j(A,b) = \operatorname{span}\{b, Ab, \ldots, A^{j-1}b\}, \qquad j = 1, 2, \ldots .$$

The jth iterate, $x^{(j)}$, determined by the GMRES method satisfies the minimization problem

$$\|b - Ax^{(j)}\| = \min_{x \in \mathcal{K}_j(A,b)} \|b - Ax\|, \qquad x^{(j)} \in \mathcal{K}_j(A,b).$$

Since the Krylov subspaces $\mathcal{K}_j(A,b)$ are nested, i.e., $\mathcal{K}_{j-1}(A,b) \subset \mathcal{K}_j(A,b)$, the norms of the residual vectors

$$r^{(j)} = b - Ax^{(j)}$$

form a nonincreasing sequence, i.e.,

$$\|r^{(j)}\| \leq \|r^{(j-1)}\|, \qquad j = 1, 2, \ldots .$$

The standard implementation of GMRES is based on the Arnoldi process for the determination of an orthonormal basis of $\mathcal{K}_j(A,b)$; see [17, Chapter 6]. Application of j steps of the Arnoldi process with initial vector $b/\|b\|$ yields the decomposition

$$AV_j = V_j H_j + f_j e_j^T, \qquad V_j \in \mathbb{R}^{n \times j}, \quad H_j \in \mathbb{R}^{j \times j}, \quad f_j \in \mathbb{R}^n, \tag{2.1}$$

where $V_j^T V_j = I_j$, $V_j e_1 = b/\|b\|$, $V_j^T f_j = 0$, and H_j is an upper Hessenberg matrix with nonvanishing subdiagonal elements. Here $I_j \in \mathbb{R}^{j\times j}$ denotes the identity matrix and e_ℓ the ℓth axis vector of appropriate dimension. The columns of V_j form an orthonormal basis of the Krylov subspace $\mathcal{K}_j(A,b)$. The computation of the Arnoldi decomposition (2.1) requires the evaluation of j matrix-vector products with the matrix A. The jth iterate $x^{(j)}$ of the GMRES method can easily be computed when the decomposition (2.1) is available; see [17, 18] for details.

GMRES is said to break down at step $j+1$ if $\dim \mathcal{K}_{j+1}(A,b) = \dim \mathcal{K}_j(A,b)$. Then $f_j = 0$ in (2.1) and it follows readily that $x^{(j)}$ solves (1.3); see [17, 18] for a proof. We assume that j is small enough so that a decomposition of the form (2.1) exists.

The GMRES method requires that a basis of $\mathcal{K}_j(A,b)$, such as the matrix V_j in (2.1), be available when computing $x^{(j)}$. Hence, the memory requirement of the method increases linearly with j. To limit the memory requirement, GMRES often is restarted periodically, say every ℓ iterations. The resulting GMRES(ℓ) method is defined as follows. Having computed the approximate solution $x^{(\ell)}$ of (1.3) and the associated residual vector $r^{(\ell)}$, we determine a correction $w^{(\ell)}$ of $x^{(\ell)}$ by applying ℓ steps of the GMRES method to the linear system of equations $Aw = r^{(\ell)}$. Thus, $w^{(\ell)}$ satisfies

$$\|r^{(\ell)} - Aw^{(\ell)}\| = \min_{w\in\mathcal{K}_\ell(A,r^{(\ell)})} \|r^{(\ell)} - Aw\|, \qquad w^{(\ell)} \in \mathcal{K}_\ell(A, r^{(\ell)}).$$

We obtain the new approximate solution $x_2^{(\ell)} = x^{(\ell)} + w^{(\ell)}$ of (1.3). If $x_2^{(\ell)}$ is not satisfactory, then a correction $w_2^{(\ell)}$ of $x_2^{(\ell)}$ can be computed similarly as $w^{(\ell)}$. For further details on the GMRES(ℓ) method; see [17, 18].

The jth iterate, $x^{(j)}$, computed by the RRGMRES method satisfies

$$\|b - Ax^{(j)}\| = \min_{x\in\mathcal{K}_j(A,Ab)} \|b - Ax\|, \qquad x^{(j)} \in \mathcal{K}_j(A, Ab).$$

Thus, all approximate solutions $x^{(j)}$ determined by RRGMRES lie in the range of A. Applications of the RRGMRES method to ill-posed problems are discussed in [4].

The matrix A in linear discrete ill-posed problems often is a smoothing operator, such as a convolution with a smooth kernel. Therefore a linear combination of vectors in $\mathcal{K}_j(A,Ab)$ may be better suited to approximate a smooth solution than the approximate solution $x^{(j)} \in \mathcal{K}_j(A,b)$ determined by j steps of GMRES. In these situations, RRGMRES generally will give a better approximation of the solution $\hat{x}$ of (1.1) than GMRES. Conversely, a linear combination of vectors in $\mathcal{K}_j(A,b)$ often is better suited to approximate a nonsmooth solution $\hat{x}$ of (1.1) than any vector in $\mathcal{K}_j(A,Ab)$. This is illustrated by computed examples in Section 4; see also the numerical examples in [4]. The restarted RRGMRES(ℓ) method is defined analogously to the restarted GMRES(ℓ) method.

The $(j-1)$st iterate determined by RRGMRES lives in $\mathcal{K}_{j-1}(A,Ab)$ and the jth iterate computed by GMRES belongs to $\mathcal{K}_j(A,b)$. Since $\mathcal{K}_{j-1}(A,Ab) \subset \mathcal{K}_j(A,b)$, both iterates can be computed for essentially the same cost as computing one of them; see [4].

The vector $\tilde{x} \in \mathbb{R}^n$ is said to satisfy the discrepancy principle if the associated residual vector $\tilde{r} = b - A\tilde{x}$ is bounded by

$$\|\tilde{r}\| \leq \alpha\delta, \tag{2.2}$$

where δ is given by (1.2) and $\alpha \geq 1$ is a specified constant. The residual $\tilde{r}$ is sometimes referred to as the discrepancy associated with $\tilde{x}$. In the context of iterative methods for ill-posed problems, one is interested in studying the behavior of the first iterate $\tilde{x} = x^{(j)}$ that satisfies the discrepancy principle (2.2) as δ decreases to zero and α is held fixed; see, e.g., [5, 8]. Generally, j increases as δ is decreased.

It is shown in [5], that if (1.1) is an operator equation of infinite dimensions ($m = n = \infty$) and the solution $\hat{x}$ of (1.1) lives in a finite dimensional Krylov subspace $\mathcal{K}_j(A, b)$, then the GMRES method is a regularization method when the iteration process is terminated as soon as an iterate $x^{(j)}$ which satisfies the discrepancy principle has been determined.

We remark that when the norm δ of the error d is not known, the iterations with the GMRES method can be terminated by using the condition L-curve; see [6] for a discussion of this stopping criterion and several illustrations of its effectiveness.

The CGNR method is the Conjugate Gradient method applied to the normal equations (1.4) associated with the linear system (1.3). The jth iterate $x^{(j)}$ determined by the CGNR method lies in the Krylov subspace $\mathcal{K}_j(A^TA, A^Tb)$ and satisfies

$$\|b - Ax^{(j)}\| = \min_{x \in \mathcal{K}_j(A^TA, A^Tb)} \|b - Ax\|. \tag{2.3}$$

Thus, the norm of the residual vectors $r^{(j)} = b - Ax^{(j)}$ determined by the CGNR method decreases monotonically as j increases.

We use the implementation of CGNR described in, e.g., [3, pp. 288 ff.]. This implementation computes the jth iterate from the previous iterate according to

$$x^{(j)} = x^{(j-1)} + \widetilde{\alpha}_{j-1} p^{(j-1)},$$

where $\widetilde{\alpha}_{j-1} \in \mathbb{R}$ is the solution of the minimization problem

$$\min_{\widetilde{\alpha} \in \mathbb{R}} \|b - A(x^{(j-1)} + \widetilde{\alpha} p^{(j-1)})\|. \tag{2.4}$$

The equivalence of the minimization problems (2.3) and (2.4) is due to the fact that the search directions $p^{(0)}, p^{(1)}, \ldots, p^{(j-1)}$ are A^TA-conjugate and span $\mathcal{K}_j(A^TA, A^Tb)$. The vector $p^{(j-1)}$ is computed as a linear combination of $p^{(j-2)}$ and the residual $\breve{r}^{(j-1)} = A^Tb - A^TAx^{(j-1)}$ associated with the normal equations (1.4). We remark that the matrix A^TA does not have to be explicitly formed, and bases for the Krylov subspaces $\mathcal{K}_j(A^TA, A^Tb)$, $j = 1, 2, \ldots$, do not have to be stored. The evaluation of $x^{(j)}$ requires the computation of $2j$ matrix-vector products with the matrices A or A^T.

The restarted version of the CGNR method, denoted by CGNR(ℓ), is defined analogously to the GMRES(ℓ) method. When the CGNR iterations are terminated according to the discrepancy principle, the CGNR method is a regularizing method; see [8, 11] for discussions.

3. Restarted iteration and nonnegativity. We describe two types of inner-outer iterative schemes for the computation of nonnegative approximate solutions of the constrained minimization problem (1.5). The first type of schemes are based on projected restarted iterative methods, where before each restart the computed solution is projected onto the set $\mathbb{S}$ of nonnegative vectors,

$$\mathbb{S} = \{x \in \mathbb{R}^n : x \geq 0\}. \tag{3.1}$$

Let $x^{(0)} := 0$ be the initial iterate and let $x^{(k)}$ be the kth nonnegative approximate solution of (1.5) determined by one of our schemes. Assume that $x^{(k)}$ does not satisfy the discrepancy principle, i.e., $\|b - Ax^{(k)}\| > \alpha\delta$, where α and δ are the same as in (2.2). We then apply GMRES, RRGMRES, CGNR, or a restarted variant of these methods, to compute an approximate solution $w^{(k)}$ of the linear system of equations

$$Aw = -r^{(k)}, \tag{3.2}$$

where $r^{(k)} = b - Ax^{(k)}$. The iterations are terminated as soon as an iterate that satisfies the discrepancy principle has been determined or, in case a restarted iterative method is used, when the maximum number of allowed consecutive iterations have been carried out. We refer to these iterations as inner iterations. Denote the computed iterate by $w^{(k)}$. The new approximate solution $x^{(k+1)}$ of (1.5) is the orthogonal projection of $x^{(k)} + w^{(k)}$ onto the set $\mathbb{S}$, i.e.,

$$x^{(k+1)} = \left(x^{(k)} + w^{(k)}\right)_+,$$

where $(\cdot)_+$ is defined by

$$((s)_+)_i = \max\{s_i, 0\}, \quad 1 \le i \le n, \quad s \in \mathbb{R}^n.$$

We refer to the computation of $x^{(k+1)}$ from $x^{(k)}$ as an outer iteration. If $x^{(k+1)}$ satisfies the discrepancy principle, then we accept $x^{(k+1)}$ as an approximate solution of (1.5); otherwise we compute a correction of $x^{(k+1)}$ similarly as we determined a correction of $x^{(k)}$.

The outer iterations are terminated as soon as a nonnegative approximate solution $x^{(j)}$ that satisfies the discrepancy principle has been determined, or the maximum number of outer iterations, specified by the parameter $maxit$, have been carried out.

When $m = n$, we may apply GMRES, RRGMRES, CGNR, or restarted variants, for the inner iterations, and refer to the corresponding iterative schemes as projected restarted GMRES, projected restarted RRGMRES, or projected restarted CGNR, respectively. If $m \ne n$, then the inner iterations are carried out by CGNR or restarted CGNR. The projected restarted iterative schemes are summarized by the following algorithm.

ALGORITHM 1: *Projected Restarted Iteration*

Input: A, b, α, δ, ℓ, $maxit$.

Output: Approximate nonnegative solution to (1.5).

Step 0: *Initialization.* Let $x^{(0)} = 0$ and $k = 0$.

Step 1: Let $r^{(k)} = b - Ax^{(k)}$. Apply an iterative method to compute an approximate solution $w^{(k)}$ of the linear system $Aw = -r^{(k)}$. Terminate the iterations when either the discrepancy principle is satisfied or the maximum number of consecutive iterations, ℓ, has been carried out.

Step 2: *Update the iterate.* Let

$$x^{(k+1)} = \left(x^{(k)} + w^{(k)}\right)_+. \tag{3.3}$$

Step 3: *Check the stopping criteria.* If a stopping criterion is satisfied, then exit with approximate solution $x^{(k+1)}$ of (1.5); otherwise, let $k = k + 1$ and go to Step 1.

As pointed out in Section 2, the memory requirement of GMRES grows linearly with the number of consecutive iterations. In order to keep the memory requirement from becoming very large, the restarted GMRES(ℓ) method is often used instead of the (unrestarted) GMRES method. The projected restarted GMRES(ℓ) method differs from restarted GMRES(ℓ) in that GMRES is restarted with the orthogonal projection $(x^{(k)} + w^{(k)})_+$ as initial iterate and the iterations are terminated as soon as an iterate that satisfies the discrepancy principle has been computed or when ℓ consecutive iterations have been carried out.

In our numerical experiments, we have observed that in the beginning of the computations, the parameter ℓ may limit the number of consecutive iterations carried out by the projected GMRES(ℓ) method, while later during the computations the number of consecutive iterations carried out often is determined by the discrepancy principle. Computed examples in Section 4 illustrate that restarted schemes may be beneficial to overcome the tendency of the discrepancy principle to yield underregularized approximate solutions, in particular when the error in the right-hand side is small.

Our second type of schemes complements the projected restarted iterative schemes outlined above with a line search. The motivation for applying a line search is that the norm of the discrepancy associated with the nonnegative iterates $x^{(k)}$, $k = 1, 2, \ldots$, determined by Algorithm 1 is not guaranteed to decrease monotonically with k. In order to generate a sequence of nonnegative iterates $x^{(k)}$ with monotonically decreasing norm of the associated discrepancies, we require an Armijo condition to be satisfied. Armijo conditions are commonly imposed in globally convergent quasi-Newton methods; see, e.g., [7, 13].

The computations proceed as follows with our second type of schemes. Let $x^{(k)}$ be the kth computed nonnegative approximate solution of (1.5). If $x^{(k)}$ satisfies the discrepancy principle, then, just like in the schemes described above, $x^{(k)}$ is accepted as an approximate solution of (1.5); otherwise an approximate solution $w^{(k)}$ of (3.2) is computed similarly as described above. The vector $w^{(k)}$ is multiplied by a factor 2^{-m}, where m is the smallest nonnegative integer such that the residual of the projected approximation $\left(x^{(k)} + 2^{-m} w^{(k)}\right)_+$ satisfies the Armijo condition

$$\|b - A\left(x^{(k)} + 2^{-m} w^{(k)}\right)_+\| < (1 - 2^{-m}\beta)\|b - Ax^{(k)}\|, \tag{3.4}$$

where β is a small positive constant. The new approximate solution of (1.5) is given by

$$x^{(k+1)} = \left(x^{(k)} + 2^{-m} w^{(k)}\right)_+ .$$

Thus, our second type of schemes are restricted step projected restarted iterative methods. As in Algorithm 1, the outer iterations are terminated when either a nonnegative approximate solution of (1.5) that satisfies the discrepancy principle has been found, or the maximum number of outer iterations, specified by the parameter $maxit$, have been carried out. We refer to this scheme as a restricted step projected GMRES(ℓ) method when the inner iterations are carried out by GMRES(ℓ). The restricted step projected RRGMRES(ℓ) and CGNR(ℓ) methods are defined similarly. The following algorithm summarizes the computations for the schemes of the second type.

ALGORITHM 2: *Restricted Step Projected Restarted Iteration*

Input: A, b, α, δ, ℓ, $maxit$, β.
Output: Approximate nonnegative solution to (1.5).
Step 0: *Initialization.* Let $x^{(0)} = 0$ and $k = 0$.
Step 1: Let $r^{(k)} = b - Ax^{(k)}$. Apply an iterative method to compute an approximate solution $w^{(k)}$ of the linear system $Aw = -r^{(k)}$. Terminate the iterations when either the discrepancy principle is satisfied or the maximum number of consecutive iterations, ℓ, has been carried out.
Step 2: *Line search.* Find the smallest integer $m \geq 0$ that satisfies the condition (3.4).
Step 3: *Update the iterate.* Let

$$x^{(k+1)} = \left(x^{(k)} + 2^{-m} w^{(k)}\right)_+ . \tag{3.5}$$

Step 4: *Check the stopping criteria.* If a stopping criterion is satisfied, then exit with the approximate solution $x^{(k+1)}$ of (1.5); otherwise, let $k = k + 1$ and go to Step 1.

In our computed examples, we let $\beta = 1 \cdot 10^{-4}$. We require that $m \leq 8$ in (3.5) in order to avoid that the corrections become too small.

4. Computed examples. This section presents several computed examples which illustrate the performance of the schemes for computing nonnegative approximate solutions of (1.5) described in the previous section. In order to be able to compare these schemes to methods discussed in the literature, we present several computed examples already used by other authors. When possible, we report the computational cost and accuracy achieved with other methods.

Example 4.1. Consider the Fredholm integral equation of the first kind

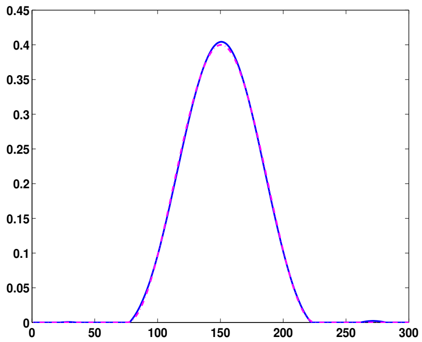
$$\int_{-6}^{6} \kappa(\tau, \sigma) x(\sigma) d\sigma = b(\tau), \qquad -6 \leq \tau \leq 6, \tag{4.1}$$

discussed by Phillips [15]. Its solution, kernel, and right-hand side are given by

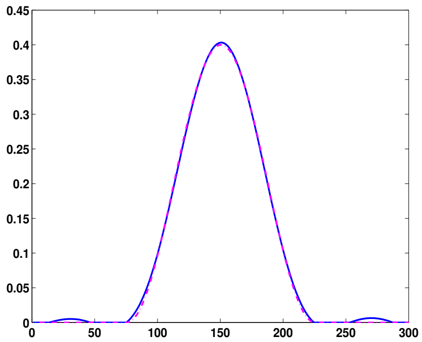
$$\begin{aligned} x(\sigma) &= \begin{cases} 1 + \cos(\frac{\pi}{3}\sigma), & \text{if } -3 < \sigma < 3, \\ 0, & \text{otherwise}, \end{cases} \\ \kappa(\tau, \sigma) &= x(\tau - \sigma), \\ b(\tau) &= (6 - |\tau|)(1 + \frac{1}{2}\cos(\frac{\pi}{3}\tau)) + \frac{9}{2\pi}\sin(\frac{\pi}{3}|\tau|). \end{aligned}$$

We use the Matlab code phillips from [10] to discretize (4.1) by a Galerkin method with orthonormal box functions as test and trial functions to obtain the symmetric indefinite matrix $A \in \mathbb{R}^{300 \times 300}$ and right-hand side $b \in \mathbb{R}^{300}$. The matrix A has many singular values in the vicinity of the origin; it is of ill-determined rank with condition number $\kappa(A) = \|A\|\|A^{-1}\|$ of size $4.2 \cdot 10^7$. Let the entries of the error vector $d \in \mathbb{R}^{300}$ be normally distributed with zero mean, and be normalized so that $\|d\|/\|\hat{b}\| = 1 \cdot 10^{-2}$. Thus, $\delta = 1 \cdot 10^{-2}\|\hat{b}\|$, cf. (1.2).

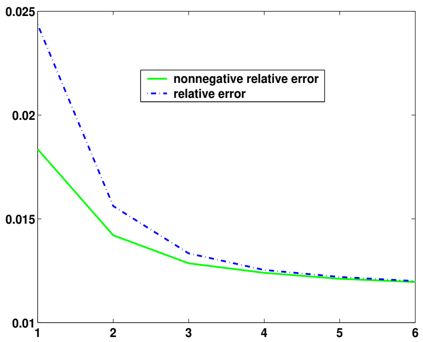
We apply Algorithm 1. Since the solution is a smooth function, it is not surprising that for this example the projected restarted RRGMRES and CGNR methods give more accurate approximations of the solution $\hat{x}$ of (1.1) than projected restarted GMRES. We terminated the iterations as soon as a nonnegative approximate solution that satisfied the discrepancy principle was found. The smallest relative error was



(a) Exact solution (dashdot) and solution computed by projected RRGMRES(15) (solid).



(b) Computed solution after one outer iteration (solid) and exact solution (dashdot).



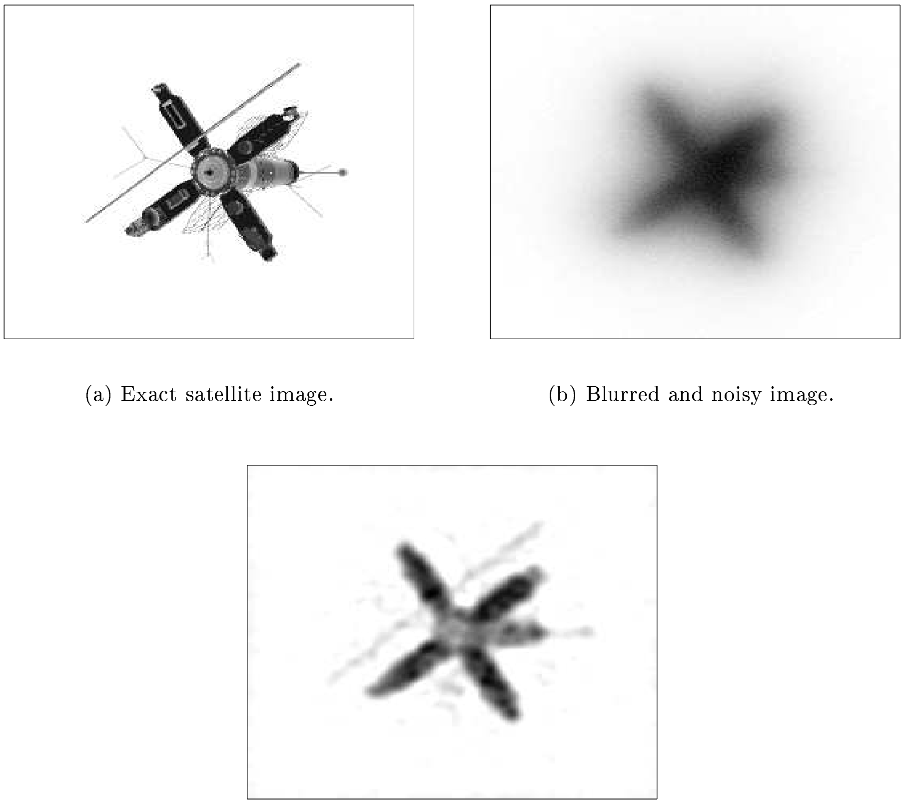
(c) Relative error norms in unconstrained (dashdot) and nonnegative (solid) computed solutions at each outer iteration.

FIG. 4.1. *Example 4.1*

achieved with projected RRGMRES(15) after 6 outer iterations, which required 17 matrix-vector product evaluations. The vector $\hat{x}$, referred to as the exact solution, and the nonnegative approximate solution $\tilde{x}_+$ computed by projected RRGMRES(15) are displayed in figure 4.1(a). Thus, $\tilde{x}_+$ satisfies the discrepancy principle and has relative error $\|\tilde{x}_+ - \hat{x}\|/\|\hat{x}\| = 0.012$. Figure 4.1(b) shows the nonnegative solution obtained after one outer iteration; it is clear that repeated iteration and projection reduces the oscillations in the computed solution significantly. Figure 4.1(c) displays a graph of the relative errors at the end of each outer iteration before and after projecting the computed solution onto the set (3.1). Notice how the distance between the curves decreases with the number of outer iterations.

When RRGMRES(15) is replaced by CGNR(15) in Algorithm 1, an approximate solution $\tilde{x}_+$ is determined after 6 outer iterations and a total of 18 iterations with CGNR, which required the evaluation of 36 matrix-vector products. The relative error of $\tilde{x}_+$ is 0.012. Thus, CGNR requires more matrix-vector product evaluations than

RRGMRES(15) to determine an approximate solution of the same quality. □



(a) Exact satellite image. (b) Blurred and noisy image.

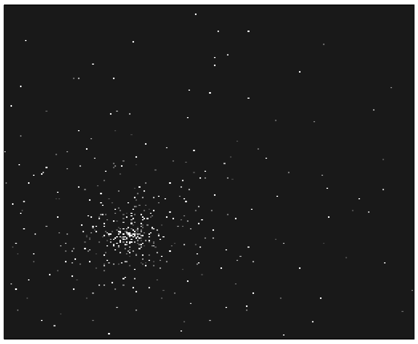
(c) Image restored by projected RRGMRES(15).

FIG. 4.2. *Example 4.2*

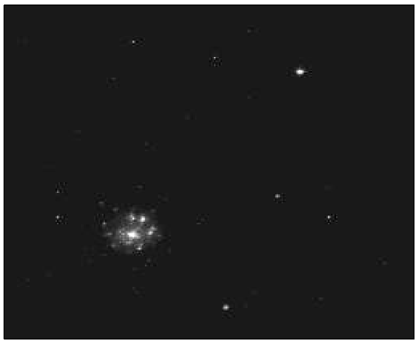
Example 4.2. The data for this example was developed at the US Air Force Phillips Laboratory and has been used to test the performance of several available algorithms for computing regularized nonnegative solutions. The data consists of the noise- and blur-free image of the satellite shown in figure 4.2(a), the point spread function, and the available blurred and noisy image of the satellite shown in figure 4.2(b). The images are represented by 256×256 pixels. The vector $\hat{x}$ represents the noise- and blur-free image, thus $\hat{x}$ is of dimension $256^2 = 65536$. We refer to $\hat{x}$ as the exact image. The matrix $A \in \mathbb{R}^{256^2 \times 256^2}$ represents the blurring operator which is determined by the point spread function, and the vector $\hat{b} = A\hat{x}$ represents the blurred noise-free image. Finally, b represents the blurred and noisy image. We compute $\delta = \|b - \hat{b}\|$.

Given A, b and δ, we would like to determine an approximation of the (assumed unavailable) exact image $\hat{x}$ and first apply Algorithm 1. We refer to the computed approximation of $\hat{x}$ as the restored image. RRGMRES gave a smaller error in the restored image than GMRES. We therefore report results obtained with the projected restarted RRGMRES(15) method. Figure 4.2(c) shows the computed nonnegative solution. A total of 24 matrix-vector products with the matrix A were evaluated to achieve a relative error in the computed solution of 0.3406. We remark that already at the end of the second outer iteration, the norm of the relative error in the computed solution is 0.3422; this error is smaller than the errors reported in [9] and [16]; the computation of this approximate solution required the evaluation of only 18 matrix-vector products, which are fewer than what is reported in [9, 16]. In this example, the discrepancy principle was satisfied after 5 outer iterations; the first inner cycle of iterations was terminated because the maximum number of iterations, 15, was reached, while the other inner iterations were stopped because the discrepancy principle was satisfied.

We turn to Algorithm 2. The projected restarted RRGMRES(15) method with line search performed only one outer iteration and stopped because the Armijo condition (3.4) was violated. Nonetheless, the relative error in the computed solution is 0.3447, which is smaller than the smallest error achieved by the algorithms proposed in [9]. Only 16 matrix-vector product evaluations were required. □



(a) Exact star cluster image.



(b) Blurred and noisy image.

(c) Image restored by projected GMRES(15).

(d) Image restored by GMRES with negative entries set to zero.

FIG. 4.3. *Example 4.3*

Example 4.3. The data used for this example can be obtained from the Telescope Science Institute and is intended to simulate a star cluster image taken by the Hubble space telescope before its defective mirror was replaced. The desired solution $\hat{x}$ is displayed in figure 4.3(a), a star cluster image represented by 256×256 pixels. We refer to $\hat{x}$ as the exact star cluster. The point spread function used is a combination of 25 different point spread functions, each one acting on a separate part of the sky to yield a space-variant point spread function. The matrix which models this blurring a nonsymmetric. Noise was added to the blurred image to produce the "observed" image displayed in figure 4.3(b). The latter is represented by the vector b.

We first apply Algorithm 1. The large difference in the light intensity between the stars and the black sky makes the desired solution pronouncedly nonsmooth. Therefore the projected restarted GMRES method determined a better restoration than the projected restarted RRGMRES method. The image obtained with the projected GMRES(15) method is shown in figure 4.3(c); its relative error norm is only 0.0677. It was computed in 11 outer iterations, which required 16 matrix-vector product evaluations. The outer iterations were terminated because the maximum number of outer iterations was reached. When we allowed completion of 31 outer iterations, the norm of the relative error in the computed solution decreased to 0.0500 and the number of matrix-vector products increased to 36. Only six GMRES steps were required in the first inner iteration to determine an approximate solution $x^{(6)}$ of (1.3), which satisfies the discrepancy principle. Orthogonal projection onto the set (3.1) gave the nonnegative approximate solution $x_+^{(6)}$, which does not satisfy the discrepancy principle. We will comment on $x_+^{(6)}$ below. All subsequent inner iterations only required one GMRES step. Note the clarity of the restored image and, in particular, the lack of "ringing" around the brightest stars.

Figure 4.3(d) displays the computed approximate solution $x_+^{(6)}$ of (1.5) defined above. This approximate solution is inexpensive to compute but does not satisfy the discrepancy principle. Its relative error is 0.1316. The expectation-maximization algorithm, a method frequently used by astronomers, requires 1000 iterations to determine a solution with a relative error of 0.15769; see [12]. □

5. Conclusion. This paper presents simple iterative methods for the computation of nonnegative approximate solutions of large-scale ill-posed problems. The computed examples show that the new methods are competitive with several available methods for some standard test problems.

REFERENCES

[1] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, 1998.
[2] Å. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659–670.
[3] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[4] D. Calvetti, B. Lewis, and L. Reichel, *On the choice of subspace for iterative methods for linear discrete ill-posed problems*, Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.
[5] D. Calvetti, B. Lewis, and L. Reichel, *On the regularizing properties of the GMRES method*, Numer. Math., 91 (2002), pp. 20–35.
[6] D. Calvetti, B. Lewis, and L. Reichel, *GMRES, L-curves and discrete ill-posed problems*, BIT, 42 (2002), pp. 44–65.
[7] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.
[8] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman, Harlow, 1995.
[9] M. Hanke, J. Nagy, and C. Vogel, *Quasi-Newton approach to nonnegative image restorations*, Linear Algebra Appl., 316 (2000), pp. 223–236.

[10] P. C. Hansen, *Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.

[11] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.

[12] J. Nagy and Z. Strakos, *Enforcing nonnegativity in image reconstruction algorithms*, in Mathematical Modeling, Estimation and Imaging, ed. D. C. Wilson et al., of the Society of Photo-Optical Instrumentation Engineers (SPIE), vol. 4121, The International Society for Optical Engineering, Bellingham, WA, 2000, pp. 182–190.

[13] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[14] C. C. Paige and M. A. Saunders, *LSQR: an algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.

[15] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.

[16] M. Rojas and T. Steihaug, *An interior point trust-region based method for large, nonnegative regularization*, Inverse Problems, 18 (2002), pp. 1291–1307.

[17] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003.

[18] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual method for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.