ON THE CHOICE OF SUBSPACE FOR ITERATIVE METHODS FOR LINEAR DISCRETE ILL-POSED PROBLEMS

D. CALVETTI *, B. LEWIS †, AND L. REICHEL ‡

Abstract. Many iterative methods for the solution of linear discrete ill-posed problems with a large matrix require the computed approximate solutions to be orthogonal to the null space of the matrix. We show that when the desired solution is not smooth, it may be possible to determine meaningful approximate solutions with less computational work by not imposing this orthogonality condition.

Key words. Minimal residual method, conjugate gradient method, ill-posed problems.

1. Introduction. This paper is concerned with the design of iterative methods for the computation of approximate solutions of linear systems of equations

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

with a large matrix A of ill-determined rank. Thus, A has many "tiny" singular values of different orders of magnitude. In particular, A is severely ill-conditioned. Some of the singular values of A may be vanishing. We allow $m \geq n$ or $m < n$. The right-hand side vector $\tilde{b}$ is not required to be in the range of A.

Linear systems of equations of the form (1.1) with a matrix of ill-determined rank are obtained when discretizing linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. They also arise in image restoration. Following Hansen [18], we refer to linear systems of equations with a matrix of ill-determined rank as linear discrete ill-posed problems.

In many linear discrete ill-posed problems that arise in the applied sciences and engineering, the right-hand side vector $\tilde{b}$ is contaminated by measurement and discretization errors, i.e.,

$$\tilde{b} = b + e, \tag{1.2}$$

where $b \in \mathbb{R}^m$ denotes the unknown error-free right-hand side in the range of A and $e \in \mathbb{R}^m$ represents the error. We refer to e as the error vector or noise vector.

We would like to determine a solution of the consistent linear discrete ill-posed problem with the unknown error-free right-hand side vector b,

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

If the solution of (1.3) is not unique, then one typically is interested in computing the solution of minimal Euclidean norm. We denote the latter by x^*.

Since the right-hand side of (1.3) is not known, we seek to determine an approximation of x^* by computing an approximate solution of the available linear system of equations (1.1). We remark that due to the severe ill-conditioning of the matrix

*Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106. E-mail: dxc57@po.cwru.edu. Research supported in part by NSF grant DMS-9806702.

†Department of Mathematics and Computer Science, Kent State University, Kent, OH 44242. E-mail: blewis@mcs.kent.edu. Research supported in part by NSF grant DMS-9806413.

‡Department of Mathematics and Computer Science, Kent State University, Kent, OH 44242. E-mail: reichel@mcs.kent.edu. Research supported in part by NSF grant DMS-9806413.

A and due to the presence of the error e in the right-hand side vector $\tilde{b}$, the least-squares solution $\tilde{x}^*$ of minimal Euclidean norm of (1.1) generally does not constitute a meaningful approximation of the minimal-norm least-squares solution x^* of (1.3).

In order to be able to compute a meaningful approximation of x^*, it is necessary to replace the problem (1.1) by a problem that is less sensitive to perturbations in the right-hand side b. This replacement is commonly referred to as regularization. One of the most popular regularization methods is Tikhonov regularization, which replaces (1.1) by the minimization problem

$$\min_{x\in\mathbb{R}^n}(\|Ax-\tilde{b}\|^2+\lambda\|Lx\|^2). \tag{1.4}$$

Here λ is a nonnegative real scalar, often referred to as the regularization parameter, and L is a regularization operator. Throughout this paper $\|\cdot\|$ denotes the Euclidean norm or the associated induced matrix norm. The value of the regularization parameter determines the sensitivity of the problem (1.4) to perturbations, and how much the solution of the minimization problem differs the from the minimal-norm least-squares solution x^* of equation (1.3).

Recently, the computation of a meaningful approximation of x^* by applying a few steps of an iterative solution method to the linear discrete ill-posed problem with a contaminated right-hand side (1.1) has received considerable attention. This approach is discussed in [6, 9, 13, 16, 18] for linear discrete ill-posed problems (1.1) with a symmetric matrix and in [4, 5, 7] for linear discrete ill-posed problems with a nonsymmetric matrix. The performance of the Conjugate Gradient (CG) method applied to the normal equations associated with (1.1),

$$A^TA\tilde{x}=A^T\tilde{b} \tag{1.5}$$

is considered in [13, 14, 18]. Since the matrix A in (1.1) is assumed to be of ill-determined rank, so is the symmetric matrix A^TA of equation (1.5). Thus, the consistent linear system of equations (1.5) is a linear discrete ill-posed problem. In the iterative methods, the iteration number can be thought of as the regularization parameter; see below.

Iterative methods proposed in the literature for the solution of linear discrete ill-posed problems with a symmetric matrix A determine iterates in the range of the matrix to ensure that the iterates are orthogonal to the null space of the matrix; see [8, 9, 13] for discussions. For instance, the Minimal Residual (MR) method by Paige and Saunders [20] is a popular iterative method for the solution of linear systems of equations with a symmetric indefinite nonsingular matrix. When applied to the solution of the linear discrete ill-posed problem (1.1) with initial approximate solution $\tilde{x}_0^{(MR)}=0$, it generates a sequence of iterates $\tilde{x}_1^{(MR)},\tilde{x}_2^{(MR)},\tilde{x}_3^{(MR)},\ldots$, such that $\tilde{x}_\ell^{(MR)}$ is in the Krylov subspace

$$\mathbb{K}_\ell(A,\tilde{b})=\text{span}\{\tilde{b},A\tilde{b},\ldots,A^{\ell-1}\tilde{b}\}.$$

We remark that as the dimension ℓ of the Krylov subspace increases, the problem solved by the iterative method resembles more and more the original problem (1.1). Therefore the dimension of the Krylov subspace, or equivalently the iteration number, can be thought of as a regularization parameter.

Hanke [13] proposed that a variant of the MR method, which we refer to as the Range Restricted Minimal Residual (RRMR) method, be used for computing approximate solutions of linear discrete ill-posed problems. The ℓth iterate computed by the RRMR method, denoted by $\tilde{x}_\ell^{(RRMR)}$, lies in the Krylov subspace $\mathbb{K}_\ell(A, A\tilde{b})$, which is in the range of A. Since A is symmetric, $\tilde{x}_\ell^{(RRMR)}$ is in the range of A^T as well and, therefore, is orthogonal to the null space of A. Details of the MR and RRMR methods are discussed in Section 2. We remark that iterates $\tilde{x}_\ell^{(CG)}$ generated by the CG method when applied to the normal equations (1.5) with initial approximate solution $\tilde{x}_0^{(CG)} = 0$ are in the Krylov subspace $\mathbb{K}_\ell(A^TA, A^T\tilde{b})$ for $\ell \geq 1$. In particular, the iterates $\tilde{x}_\ell^{(CG)}$ are in the range of A^T, and therefore are orthogonal to the null space of A.

Let (1.1) be a linear discrete ill-posed problem with a noisy right-hand side $\tilde{b}$, such that the minimal-norm least-squares solution x^* of the associated discrete ill-posed problem (1.3) with the noise-free right-hand side b represents the discretization of a function that is not smooth. It is the purpose of the present paper to show that it can be advantageous to compute approximate solutions of (1.1) by iterative methods that do not force the computed iterates to be in the ranges of A or A^T. The advantages include faster convergence towards an acceptable approximation of x^* and, not seldom, a computed approximation of superior quality. The reason for this is that iterates in the ranges of A or A^T represent discretizations of smooth functions, which may not be well suited to approximate the discretization of the nonsmooth function x^*. On the other hand, when a linear discrete ill-posed problem has a minimal-norm least-squares solution x^* that is the discretization of a smooth function, iterative methods that require the iterates to lie in the ranges of A or A^T might perform well.

Section 2 discusses linear discrete ill-posed problems (1.1) with a symmetric matrix and presents numerical examples that illustrate the performance of the MR and RRMR methods. Both linear discrete ill-posed problems, whose minimal-norm least-squares solutions x^* represent smooth and non-smooth functions, are considered. Comparisons with the CG method applied to the normal equations (1.5) are also presented. Moreover, the section describes an implementation of the RRMR method that requires fewer inner product evaluations than implementations described in [11, 13].

Section 3 considers linear discrete ill-posed problems (1.1) with a nonsymmetric matrix. We compare the GMRES method, which determines iterates in the Krylov subspaces $\mathbb{K}_\ell(A, \tilde{b})$, $\ell = 1, 2, \ldots$, with a variant of this method, referred to as the Range Restricted GMRES (RRGMRES) method, which determines iterates in the Krylov subspaces $\mathbb{K}_\ell(A, A\tilde{b})$, $\ell = 1, 2, \ldots$. Thus, similarly as the RRMR method, the RRGMRES method determines iterates in the range of A. Moreover, we compare the GMRES and RRGMRES methods to the CG method applied to the normal equations (1.5) and find that, typically, the GMRES method determines meaningful approximate solutions of (1.3) with least computational effort when the minimal-norm least-squares solution x^* of (1.3) represents the discretization of a nonsmooth function. Section 3 also illustrates how the GMRES method can be applied to linear systems of equations with a nonsquare matrix. Section 4 contains concluding remarks.

2. Linear discrete ill-posed problems with a symmetric matrix.

Assume for the moment that the symmetric matrix $A \in \mathbb{R}^{m\times m}$ is of *well-determined* rank k and that the largest singular value is about one. Then the singular values $\sigma_j(A)$ of A can be ordered so that

$$\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_k(A) \gg \sigma_{k+1}(A) = \ldots = \sigma_m(A) = 0.$$

The least-squares solution of minimal Euclidean norm $\tilde{x}^*$ of (1.1) often is a meaningful approximation of the least-squares solution x^* of (1.3) of minimal Euclidean norm, because the smallest nonvanishing singular value $\sigma_k(A)$ is much larger than zero. Typically, iterative methods for the solution of linear systems of equations with a symmetric matrix of well-determined rank are designed so that the iterates converge to $\tilde{x}^*$ and the iterations are terminated when a good approximation of $\tilde{x}^*$ has been computed; see [10, 11, 15] and references therein.

We remark that often linear discrete ill-posed problems with a symmetric matrix arise as normal equations (1.5). Then multiplication of the right-hand side $\tilde{b}$ by the matrix A^T filters out the noise orthogonal to the range of A. However, in this section we assume that the matrix A is square and that the noise in the right-hand side contains components both in and orthogonal to the range of A.

The least-squares solution $\tilde{x}^*$ of minimal Euclidean norm of (1.1) is orthogonal to the $(m-k)$-dimensional null space of A. We would like that the iterates determined by the iterative method also be orthogonal to the null space of A in order to secure that they converge to $\tilde{x}^*$. Since the range and null space of a symmetric matrix are orthogonal, we achieve this by requiring that the iterates lie in Krylov subspaces of the form $\mathbb{K}_\ell(A, A\tilde{b})$. The RRMR method gives iterates in these Krylov subspaces and so do the methods described in [10, 11, 15].

Hanke [13] proposed the use of the RRMR method for the solution of linear discrete ill-posed problems (1.1) with a symmetric matrix. The RRMR method solves the following minimization problems. Let the initial approximate solution of (1.1) be $\tilde{x}_0^{(RRMR)} = 0$. Then the RRMR method determines iterates $\tilde{x}_\ell^{(RRMR)}$ for $\ell \geq 1$, such that

$$\|A\tilde{x}_\ell^{(RRMR)} - \tilde{b}\| = \min_{\tilde{x}\in\mathbb{K}_\ell(A,A\tilde{b})} \|A\tilde{x} - \tilde{b}\|, \qquad \tilde{x}_\ell^{(RRMR)} \in \mathbb{K}_\ell(A, A\tilde{b}). \tag{2.1}$$

We include an algorithm for the RRMR method below. This algorithm can be combined with an algorithm for the MR method. The iterates $\tilde{x}_\ell^{(MR)}$, $\ell \geq 1$, determined by the MR method with initial approximate solution $\tilde{x}_0^{(MR)} = 0$ satisfy

$$\|A\tilde{x}_\ell^{(MR)} - \tilde{b}\| = \min_{\tilde{x}\in\mathbb{K}_\ell(A,\tilde{b})} \|A\tilde{x} - \tilde{b}\|, \qquad \tilde{x}_\ell^{(MR)} \in \mathbb{K}_\ell(A, \tilde{b}). \tag{2.2}$$

The algorithm below computes iterates for the MR or RRMR methods. Its derivation is analogous to that of the MINRES algorithm of Paige and Saunders [20]. It uses the LQ-factorization of the symmetric tridiagonal matrices determined by the Lanczos process. An analogous implementation of the MR method, that is based on the QR-factorization instead of the LQ-factorization, has recently been described by Fischer [12]. The algorithm requires the matrix A to be symmetric. The matrix may be definite, indefinite or singular. We remark that if the matrix A is known to be positive definite, then the recursions can be simplified, see, e.g., Saad [22, p. 183] for an example of a simpler algorithm for the MR method.

Algorithm 2.1. *MR and RRMR methods*
Input: A, $\tilde{b}$, ℓ, *METHOD* $\in \{$*MR, RRMR*$\}$.
Output: Approximate solution $\tilde{x}_\ell^{(MR)}$ *or* $\tilde{x}_\ell^{(RRMR)}$ *of (1.1) determined by the MR or RRMR method, respectively.*
$\tilde{x}^{(MR)} := 0,\ \tilde{x}_0^{(RRMR)} := 0,\ z_{-1} := 0,\ z_0 := 0,\ v_0 := 0,$
$c := 1,\ s := 0,\ \mu_0 := 0,\ \mu_{-1} := 0,\ \epsilon_0 := 0,\ \delta_1 := 0,$

if METHOD=MR then
$v_1 := \tilde{b}$,
elseif METHOD=RRMR then
$v_1 := A\tilde{b}$,
endif
$\beta_0 := \|v_1\|$, $v_1 := v_1/\beta_0$,
for $j = 1, 2, \ldots, \ell$ *do*
$w := Av_j$,
$w := w - \beta_{j-1}v_{j-1}$,
$\alpha_j :=< w, v_j >$,
$w := w - \alpha_j v_j$,
$\beta_j := \|w\|$,
$v_{j+1} := w/\beta_j$,
if $j = 1$ *then*
$\bar{\delta} := \beta_1$,
$\bar{\gamma} := \alpha_1$,
$\gamma_1 := (\bar{\gamma}^2 + \bar{\delta}^2)^{1/2}$,
$\mu_1 := \beta_0/\gamma_1$, $\tau_1 := \beta_0\bar{\gamma}/\gamma_1$, $\sigma := 1$,
else
$c := \bar{\gamma}/\gamma_{j-1}$,
$s := \beta_{j-1}/\gamma_{j-1}$,
$\delta_j := c\bar{\delta} + s\alpha_j$,
$\epsilon_j := s\beta_j$,
$\bar{\gamma} := s\bar{\delta} - c\alpha_j$,
$\bar{\delta} := -c\beta_j$,
$\gamma_j := (\bar{\gamma}^2 + \beta_j^2)^{1/2}$,
$\mu_j := -(\delta_j\mu_{j-1} + \epsilon_{j-1}\mu_{j-2})/\gamma_j$,
$\sigma := s\sigma$, $\tau_j := \beta_0\sigma\bar{\gamma}/\gamma_j$,
endif
$z_j := (v_j - \delta_j z_{j-1} - \epsilon_{j-1}z_{j-2})/\gamma_j$,
if METHOD=MR then
$\tilde{x}_j^{(MR)} := \tilde{x}_{j-1}^{(MR)} + \tau_j z_j$,
elseif METHOD=RRMR then
$\tilde{x}_j^{(RRMR)} := \tilde{x}_{j-1}^{(RRMR)} + \mu_j z_j$,
endif
end □

For some problems, it may be desirable to compute the residual vectors $r_j^{(MR)} = b - A\tilde{x}_j^{(MR)}$ or $r_j^{(RRMR)} = b - A\tilde{x}_j^{(RRMR)}$ associated with the iterates $\tilde{x}_j^{(MR)}$ determined by the MR method or the iterates $\tilde{x}_j^{(RRMR)}$ computed by the RRMR method. The following recursion formulas may be attractive to use when the residual vectors are desired for all values of j,

$$r_j^{(MR)} = r_{j-1}^{(MR)} - \tau_j A z_j, \qquad r_j^{(RRMR)} = r_{j-1}^{(RRMR)} - \mu_j A z_j. \tag{2.3}$$

The residual vectors may be evaluated efficiently by using these formulas for increasing values of j, without requiring the evaluation of additional matrix vector products, by noting that $Az_1 = Av_1$ and exploiting the updating formula for z_j in Algorithm 2.1 for $j > 1$.

We remark that other implementations of the MR and RRMR methods are available. Implementations of the RRMR method based on explicit use of the three-term recurrence formula of the Lanczos process are presented in [11] and [13, p. 95]. These implementations require three inner product evaluations with m-vectors in each iteration; Algorithm 2.1 only requires two. Here we count the computation of the Euclidean norm of an m-vector as an inner product evaluation. Inner product evaluations can be bottle-necks on parallel computers.

The following example compares the MR and RRMR methods applied to the linear system (1.1) with the CG method applied to the associated normal equations (1.5). We use the implementation CGLS, discussed by Björck [2], of the CG method applied to the normal equations. We refer to this method as the CGLS method. It does not require the matrix $A^T A$ to be explicitly formed. All computations of the present paper were carried out in Octave on a workstation with machine epsilon $2 \cdot 10^{-16}$.

Example 2.1. This example is concerned with the restoration of a noisy and blurred image of a star cluster. The image is represented by a 256×256 array of grayscale values (pixels). The image formation process is defined by equations (1.1) and (1.2) with $m = n = 256^2$ as follows. Let the entries of vector x be row-ordered pixels of a noise- and blur-free model image and let the matrix A represent the blurring operator. Then the vector $b = Ax$ represents the blurred, but noise-free, image associated with the model image x. Let the noise vector e consist of normally distributed random entries with zero mean and with variance chosen so that we achieve the noise level $\|e\|/\|Ax\| = 5 \cdot 10^{-2}$. In the present example, the blurring matrix A models separable, spatially invariant Gaussian blur. We represent A by the Kronecker product $A = T \otimes T$, where $T = [t_{jk}]_{j,k=1}^{256}$ is a symmetric banded Toeplitz matrix with entries

$$t_{jk} = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(j-k)^2}{2\sigma^2}), & \text{if } |j-k| \leq \rho, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

The blurring radius in (2.4) is $\rho = 12$, and $\sigma = 1$. The parameter σ determines the "spread" of the blur. The model image x is displayed in Figure 2.1. It was obtained via ftp from the Space Telescope Science Institute.[1] The blurred and noisy image $\tilde{b} = Ax + e$ is shown in Figure 2.2. In the present example, we assume that the contaminated image $\tilde{b}$ and the blurring operator A are available, but that the blur- and noise-free model image x is not. We would like to determine an accurate approximation of x.

Following Nagy and O'Leary [19], each pixel value v of the model image is replaced by $\min(\max(v, 50), 500)$ in order to avoid oversaturation in the displayed images. The pixel value zero corresponds to black and the value 500 to white.

We compute the relative error norms $\|x - \tilde{x}_\ell\|/\|x\|$, $\ell = 1, 2, \ldots$, for iterates $\tilde{x}_\ell$ determined by the MR and RRMR methods, applied to the solution of the linear system of equations (1.1), and for iterates determined by the CGLS method, applied to the solution of the normal equations (1.5). For all methods, we use the initial approximate solution $\tilde{x}_0 = 0$. These error norms are displayed in Figure 2.3.

The smallest relative error achieved by the MR and CGLS methods is $5.7 \cdot 10^{-1}$. The MR method requires 9 iterations to achieve this error and the CGLS method 73 iterations. The smallest relative error achieved by the RRMR method is $5.6 \cdot 10^{-1}$, which is obtained at iteration 18. Note that each iteration with the MR and RRMR

[1] Address ftp.stsci.edu, directory /pub/software/stsdas/testdata/restore/sims/star_cluster.

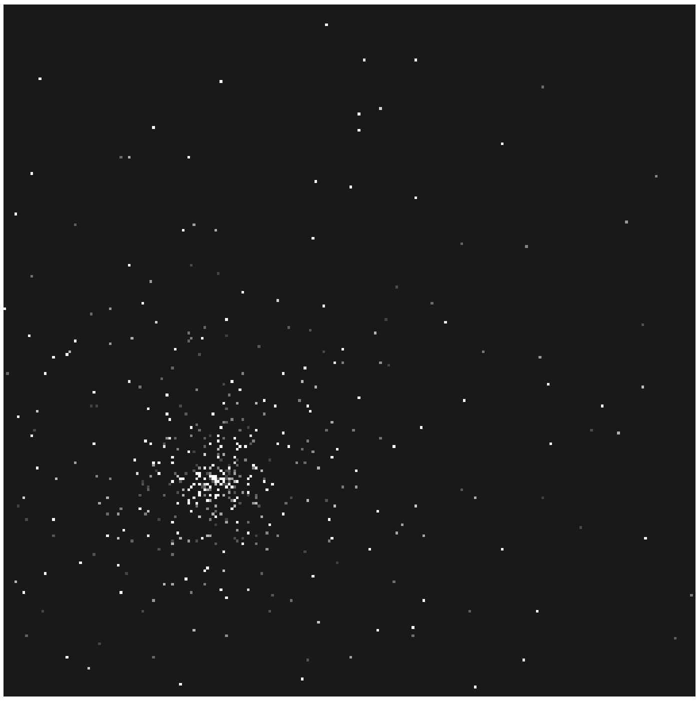

FIG. 2.1. *Example 2.1: Model image.*

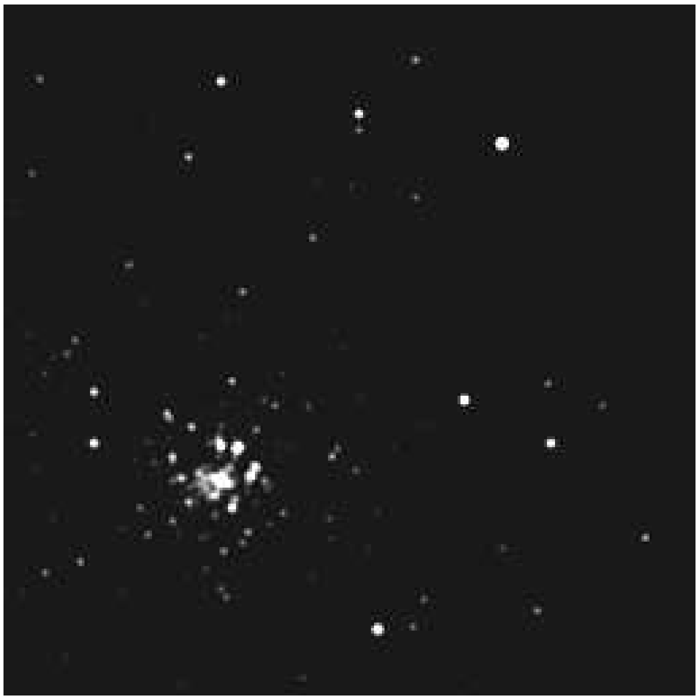

FIG. 2.2. *Example 2.1: Blurred and noisy image.*

methods requires the evaluation of one matrix-vector product with the matrix A, while each iteration with the CGLS method requires two matrix-vector product evaluations, one with the matrix A and one with A^T. Figure 2.4 displays the norm of the residual errors $\tilde{r}_\ell = \tilde{b} - A\tilde{x}_\ell$ associated with the iterates determined by the the MR, RRMR and CGLS methods.

Although the minimal relative error norms achieved by the MR, RRMR and CGLS methods are about the same, the restored images obtained by the different methods are qualitatively quite different; see Figures 2.5, 2.6 and 2.7. The restored image produced by the MR method exhibits less ringing or halo artifact around the brightest stars than the images computed by the other methods. Thus, using the Krylov subspaces $\mathbb{K}_\ell(A, \tilde{b})$, $\ell \geq 1$, makes it possible to determine a good approximation of the blur- and noise-free model image x with less computational work than when the Krylov subspaces $\mathbb{K}_\ell(A, A\tilde{b})$ or $\mathbb{K}_\ell(A^TA, A^T\tilde{b})$ are used. $\square$

We remark that in realistic applications of iterative methods to image restoration, the model image x is not known. The iterative methods therefore have to be equipped with a stopping criterion. Several stopping criteria are discussed in [4, 5, 6, 7, 13, 18]. However, in order to clearly show the importance of the choice of Krylov subspace, we have not used these criteria in the computed examples of the present paper.

Example 2.2. The linear discrete ill-posed problems of the present example has minimal-norm least-squares solution x^*, which is the discretization of a smooth function. The example is constructed from a Fredholm integral equation of the first kind discussed by Phillips [21]. Introduce the function

$$x(t) = \begin{cases} 1 + \cos(t\pi/3), & \text{if } |t| < 3, \\ 0, & \text{otherwise,} \end{cases}$$

and define the kernel $\kappa : \mathbb{R}^2 \to \mathbb{R}$ by $\kappa(s,t) = x(s-t)$. Then

$$\int_{-6}^{6} \kappa(s,t)x(t)dt = b(s), \qquad -6 \leq s \leq 6, \tag{2.5}$$

for $b(s) = (6-|s|)(1+\frac{1}{2}\cos(s\pi/3))+\frac{9}{2\pi}\sin(|s|\pi/3)$. Given the kernel $\kappa(s,t)$ and a perturbation of the right-hand side function $b(s)$, we seek to compute an approximation of the solution $x(t)$ of (2.5).

We discretize the integral equation (2.5) by the Galerkin method with the orthonormal basis functions

$$v_j(s) = \begin{cases} h^{-(1/2)}, & \text{if } s \in [(j-1)h, jh], \\ 0, & \text{otherwise,} \end{cases}$$

where $h = 12/500$, using software written in Matlab by Hansen [17]. This yields a linear system of equations with the matrix $A = [\alpha_{jk}] \in \mathbb{R}^{500\times 500}$ and the right-hand side vector $\hat{b} = [\hat{\beta}_j] \in \mathbb{R}^{500}$ with entries

$$\alpha_{jk} = \int_{-6}^{6} K(s,t)v_j(s)v_k(t)dsdt, \quad \hat{\beta}_j = \int_{-6}^{6} b(s)v_j(s)ds, \quad 1 \leq j,k \leq 500. \tag{2.6}$$

The matrix A so obtained has condition number $\sigma_1(A)/\sigma_{500}(A) = 1.7 \cdot 10^9$; thus, it is quite ill-conditioned.

The discretization of the solution $x(t)$ of (2.5) gives the vector $x = [\xi_j] \in \mathbb{R}^{500}$ with entries

$$\xi_j = \int_{-6}^{6} x(t)v_j(t)dt, \qquad 1 \leq j \leq 500.$$

Note that due to discretization errors, the vector x does not satisfy the linear systems of equations determined by the matrix A and right-hand side vector $\hat{b}$; we have $\|Ax-\hat{b}\| = 2.4\cdot 10^{-4}$.

Let the vector $\hat{e}\in\mathbb{R}^{500}$ have normally distributed random entries with zero mean and with the variance chosen so that $\|\hat{e}\| = 1\cdot 10^{-2}$. Introduce the right-hand side vector

$$\tilde{b} := \hat{b} + \hat{e}. \tag{2.7}$$

In the present example, we assume that the discretization x of the solution $x(t)$ of the integral equation (2.5) is the desired approximate solution of the linear system of equations (1.1), with the matrix A defined by (2.6) and the right-hand side vector $\tilde{b}$ given by (2.7). The norm of the error e in the vector $\tilde{b}$ is $\|Ax-\tilde{b}\| = 1\cdot 10^{-2}$, cf. (1.2).

Similarly to Example 2.1, we compute the relative error norms $\|x-\tilde{x}_\ell\|/\|x\|$, $\ell = 1, 2, \ldots$, for iterates $\tilde{x}_\ell$ determined by the MR and RRMR methods, applied to the solution of the linear system of equations (1.1), and for iterates determined by the CGLS method, applied to the solution of the normal equations (1.5). For all methods, we use the initial approximate solution $\tilde{x}_0 = 0$. These error norms are displayed in Figure 2.8. In this example, the CGLS and RRMR methods achieved slightly lower minimum relative error norms than the MR method. Figure 2.9 displays the norm of the residual errors $\tilde{r}_\ell = \tilde{b} - A\tilde{x}_\ell$ associated with the iterates determined by the the MR, RRMR and CGLS methods. □

3. Discrete ill-posed problems with a nonsymmetric matrix. The Generalized Minimal Residual (GMRES) method by Saad and Schultz [23] is a popular iterative method for the solution of linear systems of equations (1.1) with a (square) nonsymmetric nonsingular matrix $A\in\mathbb{R}^{m\times m}$. Let the initial approximate solution be $\tilde{x}_0^{(GMRES)} = 0$. Then the iterates $\tilde{x}_\ell^{(GMRES)}$, $\ell\geq 1$, determined by the GMRES method satisfy

$$\|A\tilde{x}_\ell^{(GMRES)} - \tilde{b}\| = \min_{\tilde{x}\in\mathbb{K}_\ell(A,\tilde{b})} \|A\tilde{x}-\tilde{b}\|, \qquad \tilde{x}_\ell^{(GMRES)} \in \mathbb{K}_\ell(A,\tilde{b}). \tag{3.1}$$

The minimization problem (3.1) is the same as (2.2). However, while the minimization problem (2.2) with a symmetric matrix can be solved by an algorithm with short recurrence relations for increasing values of ℓ (Algorithm 2.1), the minimization problem (3.1) with a nonsymmetric matrix in general cannot; the computation of the iterate $\tilde{x}_\ell^{(GMRES)}$ typically requires that a basis of the Krylov subspace $\mathbb{K}_{\ell-1}(A,\tilde{b})$ be available. Algorithm 3.1 below describes an implementation of the GMRES method. The computation of the iterate $\tilde{x}_\ell^{(GMRES)}$ by Algorithm 3.1 requires ℓ evaluations of matrix-vector products with the matrix A. Results on the performance of the GMRES method when applied to discrete ill-posed problems (1.1) with a nonsymmetric (square) matrix are presented in [4, 5, 7].

The Range Restricted GMRES (RRGMRES) method, proposed in [3], is a variant of the GMRES method. Let the initial iterate be given by $\tilde{x}_0^{(RRGMRES)} = 0$. Then the iterates $\tilde{x}_\ell^{(RRGMRES)}$, $\ell\geq 1$, determined by the RRGMRES method satisfy

$$\|A\tilde{x}_\ell^{(RRGMRES)} - \tilde{b}\| = \min_{\tilde{x}\in\mathbb{K}_\ell(A,A\tilde{b})} \|A\tilde{x}-\tilde{b}\|, \qquad \tilde{x}_\ell^{(RRGMRES)} \in \mathbb{K}_\ell(A,A\tilde{b}). \tag{3.2}$$

The minimization problem (3.2) is the same as the minimization problem (2.1) solved by the RRMR method. The RRGMRES method was developed for the solution of inconsistent linear systems of equations with a nonsymmetric matrix. The RRGMRES method requires one more matrix-vector product evaluation with the matrix A than the GMRES method and the same amount of computer storage to evaluate the ℓth iterate.

The following algorithm presents implementations based on the Arnoldi process of both the GMRES and RRGMRES methods. The GMRES implementation is due to Saad and Schultz, and we refer to [23] for further discussion on the GMRES algorithm.

ALGORITHM 3.1. *GMRES and RRGMRES methods*

Input: A, $\tilde{b}$, ℓ, *METHOD* $\in$ {*RRGMRES, GMRES*}.
Output: Approximate solution $\tilde{x}_\ell$ *of (1.1).*
if METHOD=GMRES
  $v_1 := \tilde{b}/\|\tilde{b}\|$,
elseif METHOD=RRGMRES
  $v_1 := A\tilde{b}/\|A\tilde{b}\|$,
endif
Define $\bar{H}_\ell := [h_{ij}]$, $h_{ij} := 0$, $1 \le i \le \ell+1$, $1 \le j \le \ell$,
for $j = 1, 2, \ldots, \ell$ *do*
  $w := Av_j$,
  for $i = 1, 2, \ldots, j$ *do*
    $h_{ij} := v_i^T w$, $w := w - h_{ij} v_i$,
  end
  $h_{j+1,j} := \|w\|$, $v_{j+1} := w/h_{j+1,j}$,
end
$V_\ell := [v_1, v_2, \ldots, v_\ell]$, $V_{\ell+1} := [v_1, v_2, \ldots, v_{\ell+1}]$,
$\tilde{y}_\ell := \bar{H}_\ell^\dagger V_{\ell+1}^T \tilde{b}$,
$\tilde{x}_\ell := V_\ell \tilde{y}_\ell$ □

Algorithm 3.1 assumes that the matrix A has the same number of rows as columns. We describe in Example 3.3 below how this restriction can be removed.

Example 3.1. This example is concerned with the restoration of an image that is contaminated by noise and spatially variant blur. The matrix A that models the blur is nonsymmetric. Specifically, partition the discretized image into 4 square non-overlapping subregions of equal size enumerated clockwise starting from the upper left corner. Assume that the blur in each subregion is spatially invariant. This gives rise to a blurring matrix A of the form

$$A = D_1 A_1 + D_2 A_2 + D_3 A_3 + A_4 D_4, \tag{3.3}$$

where D_i is a diagonal matrix, whose jth diagonal entry is 1 if the jth pixel is in subregion i and 0 otherwise. The blurring matrix A_i models spatially invariant Gaussian blur in the ith subregion. Each blurring matrix A_i is the Kronecker product of a banded symmetric Toeplitz matrix $T_i = [t_{jk}^{(i)}]_{j,k=1}^{256}$ with itself, i.e., $A_i = T_i \otimes T_i$. The entries of T_i are given by

$$t_{jk}^{(i)} = \begin{cases} \frac{1}{\sigma_i\sqrt{2\pi}} \exp(-\frac{(j-k)^2}{2\sigma_i^2}), & \text{if } |j-k| \le \rho_i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

The blurring radii are $\rho_i = \lceil 12\sigma_i \rceil$, $1 \le i \le 4$, and the parameters σ_i determine the "spread" of the blur in each subregion. We chose $\sigma_1 = 8 \cdot 10^{-1}$, $\sigma_2 = 9 \cdot 10^{-1}$,

$\sigma_3 = 9 \cdot 10^{-1}$ and $\sigma_4 = 1 \cdot 10^0$. The blur- and noise-free model image x is shown in Figure 2.1. The error vector e in (1.2) has normally distributed random entries with zero mean and variance such that $\|e\|/\|Ax\| = 5 \cdot 10^{-2}$. The degraded image looks like the image shown in Figure 2.2, even though the blurring operator differs from the one used for the construction of the image in the figure.

Figure 3.1 displays the relative error norms $\|x - \tilde{x}_\ell\|$, $\ell = 1, 2, \dots$, for the iterates $\tilde{x}_\ell$ determined by the GMRES, RRGMRES and CGLS methods, and Figure 3.2 shows the norm of the residual errors $\tilde{r}_\ell = \tilde{b} - A\tilde{x}_\ell$ associated with the computed iterates. The initial approximate solution $\tilde{x}_0 = 0$ is used for all three iterative methods.

Figures 3.3, 3.4 and 3.5 show the images associated with the iterates of smallest relative error norms determined by the GMRES, RRGMRES and CGLS methods. Thus, Figure 3.3 and 3.4 display the restored images determined by 8 iterations with the GMRES method and by 16 iterations with the RRGMRES method, respectively. Figure 3.5 shows the restored image computed by 59 iterations with the CGLS method. We see that while the relative error norms for the displayed images are about the same, the restored images determined by the three methods look quite different. The restored image produced by the GMRES method exhibits less ringing or halo artifact around the brightest stars than the images computed by the other methods, and its computation requires the fewest matrix-vector product evaluations with the matrices A or A^T. Thus, the Krylov subspaces $\mathbb{K}_\ell(A, \tilde{b})$, $\ell \geq 1$, give a good restoration with less computational work than the Krylov subspaces $\mathbb{K}_\ell(A, A\tilde{b})$ and $\mathbb{K}_\ell(A^TA, A^T\tilde{b})$. $\Box$

Example 3.2. We consider the Fredholm integral equation of the first kind,

$$\int_0^{\pi/2} \kappa(s,t)x(s)ds = b(t), \qquad 0 \leq t \leq \pi, \tag{3.5}$$

where $\kappa(s,t) = \exp(s\cos(t))$ and $b(t) = 2\sinh(t)/t$, discussed by Baart [1]. The solution is given by $x(t) = \sin(t)$. We use Matlab code from [17] to discretize (3.5) by the Galerkin method with 100 orthonormal box functions as test and trial functions. This gives a linear system of equations with a matrix $A \in \mathbb{R}^{100\times100}$ and right-hand side $\hat{b} \in \mathbb{R}^{100}$. The singular values of A "cluster" at the origin; the matrix therefore is of ill-determined rank. We determine a noise vector $e \in \mathbb{R}^{100}$ consisting of normally distributed random entries of zero mean and variance such that $\|e\| = 5 \cdot 10^{-2}$. We obtain the noisy right-hand side by $\tilde{b} = \hat{b} + e$, similarly as in Example 2.2. The desired solution x is the discretization of the solution $x(t)$ of the integral equation (3.5).

Figure 3.6 displays the relative error norms $\|x - \tilde{x}_\ell\|/\|x\|$, $\ell = 1, 2, \dots$, for iterates $\tilde{x}_\ell$ determined by the GMRES, RRGMRES and CGLS methods. The initial approximate solution $\tilde{x}_0 = 0$ is used for all three iterative methods. The norms of residual errors associated with the computed iterates are shown in Figure 3.7. In this example the RRGMRES achieves the smallest minimum relative error norm and the GMRES method the largest minimum relative error norm. $\Box$

Example 3.3. This example illustrates how the GMRES and RRGMRES methods can be applied to linear systems of equations with a nonsquare matrix. We wish to determine a least-squares solution of an underdetermined linear system (1.1) with $m = 50$ and $n = 150$, obtained by discretizing the inverse Laplace transform

$$\int_0^{\infty} \exp(-st)f(t)dt = g(s), \qquad s \geq 0, \tag{3.6}$$

where the right-hand side function g is given by

$$g(s) = \frac{1}{s+1/2}.$$

The solution $f(t)$ of (3.6) is

$$f(t) = \exp(-t/2). \tag{3.7}$$

We discretize the integral equation (3.6) by software written in Matlab by Hansen [17]. The integral is replaced by a 150-point Gauss-Laguerre quadrature rule and the equation so obtained is required to be satisfied at the collocation points $s_j = j/10$, $1 \le j \le 150$. This determines the matrix $\hat{A} \in \mathbb{R}^{150\times 150}$ and the right-hand side vector $\hat{b} = [\hat{b}_j]_{j=1}^{150}$ with entries $\hat{b}_j := g(s_j)$. We obtain an underdetermined linear system of equations by discarding the last 100 rows of the matrix $\hat{A}$ and vector $\hat{b}$. Denote the matrix so obtained by A. This is the matrix in the linear system of equations (1.1) for which we would like to compute an approximate solution. The condition number of A, measured by the ratio of the largest and smallest positive singular values of the matrix, is $2.0 \cdot 10^{18}$. Thus, A is numerically singular. We refer to the vector $x = [x_j]_{j=1}^{150}$, where $x_j = f(s_j)$, as the discretized exact solution of the error-free equation (3.6), or briefly as the exact solution.

We determine the right-hand side $\tilde{b}$ of (1.1) as follows. Denote the vector obtained by discarding the last 100 entries of $\hat{b}$ by $\breve{b}$. Let $e \in \mathbb{R}^{50}$ be a noise vector with normally distributed random entries with zero mean and with variance chosen so that $\|e\|/\|Ax\| = 1 \cdot 10^{-5}$. The right-hand side vector in (1.1) is given by $\tilde{b} = \breve{b} + e$. Note that the vector Ax may differ from the vector $\breve{b}$ due to discretization errors.

The GMRES and RRGMRES methods can be applied to the underdetermined linear system of equations (1.1) constructed as described above by padding the matrix A and right-hand side vector $\tilde{b}$ with 100 rows with zero entries. Of course, these zero entries do not have to be stored. We remark that linear systems of equations (1.1) with more rows than columns can be handled in a similar fashion.

Figure 3.8 displays the relative error norms $\|x - \tilde{x}_\ell\|/\|x\|$, $\ell = 1, 2, \ldots$, for iterates $\tilde{x}_\ell$ determined by the GMRES, RRGMRES and CGLS methods. The norms of residual errors associated with the computed iterates are shown in Figure 3.9. In this example the GMRES and RRGMRES method require about the same number of iterations to give the minimal relative error norm. The CGLS method requires considerably more iterations. □

4. Conclusion. The computed examples presented in Sections 2 and 3, as well as experience from numerous other examples, suggest that for many discrete ill-posed problems with a solution that is not the discretization of a smooth function, the MR and GMRES methods often give meaningful approximate solutions of as good or higher quality as the RRMR, RRGMRES and CGLS methods with less computational work. The latter is measured in terms of the number of matrix-vector product evaluations with the matrices A or A^T required. When the desired solution is smooth, the MR, RRMR, GMRES and RRGMRES methods generally all perform well, and for some problems the range restricted methods RRMR and RRGMRES may yield better approximations of the minimal-norm least-squares solution x^* of (1.3). For many discrete ill-posed problems, the CGLS method requires considerable more computational work than the other methods considered.

Acknowledgement. We would like to thank Mario Bertero for valuable comments.

REFERENCES

[1] M. L. Baart, The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems, IMA J. Numer. Anal., 2 (1982), pp. 241–247.

[2] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.

[3] D. Calvetti, B. Lewis and L. Reichel, GMRES-type methods for inconsistent systems, Linear Algebra Appl., 316 (2000), pp. 157–169.

[4] D. Calvetti, B. Lewis and L. Reichel, GMRES, L-curves, and discrete ill-posed problems, BIT, to appear.

[5] D. Calvetti, B. Lewis and L. Reichel, Restoration of images with spatially variant blur by the GMRES method, in Advanced Signal Processing Algorithms, Architectures, and Implementations X, ed. F. T. Luk, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), vol. 4116, The International Society for Optical Engineering, Bellingham, WA, 2000, pp. 364–374.

[6] D. Calvetti, B. Lewis and L. Reichel, An L-curve for the MINRES method, in Advanced Signal Processing Algorithms, Architectures, and Implementations X, ed. F. T. Luk, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), vol. 4116, The International Society for Optical Engineering, Bellingham, WA, 2000, pp. 385–395.

[7] D. Calvetti, B. Lewis and L. Reichel, On the regularizing properties of the GMRES method, Numer. Math., to appear.

[8] D. Calvetti, L. Reichel and Q. Zhang, Iterative exponential filtering for large discrete ill-posed problems, Numer. Math., 83 (1999), pp. 535–556.

[9] D. Calvetti, L. Reichel and Q. Zhang, Iterative solution methods for large linear discrete ill-posed problems, Applied and Computational Control, Signals and Circuits, 1 (1999), pp. 313–367.

[10] D. Calvetti, L. Reichel and Q. Zhang, An adaptive semiiterative method for symmetric semidefinite linear systems, in Approximation and Computation, ed. R. V. M. Zahar, Int'l Series of Numer. Math. # 119, Birkhäuser, Basel, 1994, pp. 77–96.

[11] D. Calvetti, L. Reichel and Q. Zhang, Conjugate gradient algorithms for symmetric inconsistent linear systems, in Proceedings of the Cornelius Lanczos International Centenary Conference, eds. J. D. Brown, M. T. Chu, D. C. Ellison and R. J. Plemmons, SIAM, Philadelphia, 1994, pp. 267–272.

[12] B. Fischer, Polynomial Based Iteration Methods for Symmetric Linear Systems, Wiley-Teubner, New York, 1996.

[13] M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems, Longman, Harlow, 1995.

[14] M. Hanke and P. C. Hansen, Regularization methods for large-scale problems, Surv. Math. Ind., 3 (1993), pp. 253–315.

[15] M. Hanke and M. Hochbruck, A Chebyshev-like semiiteration for inconsistent linear systems, Elec. Trans. Numer. Anal., 1 (1993), pp. 89–103.

[16] M. Hanke and J. G. Nagy, Restoring images degraded by spatially-variant blur, Inverse Problems, 19 (1998), pp. 1063–1082.

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

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

[19] J. G. Nagy and D. P. O'Leary, Restoring images degraded by spatially-variant blur, SIAM J. Sci. Comput., 19 (1998), pp. 1063–1082.

[20] C. C. Paige and M. A. Saunders, Solution of sparse indefinite systems of linear equations, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

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

[22] Y. Saad, Iterative Methods for Sparse Linear Systems, PWS, Boston, 1996.

[23] 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.

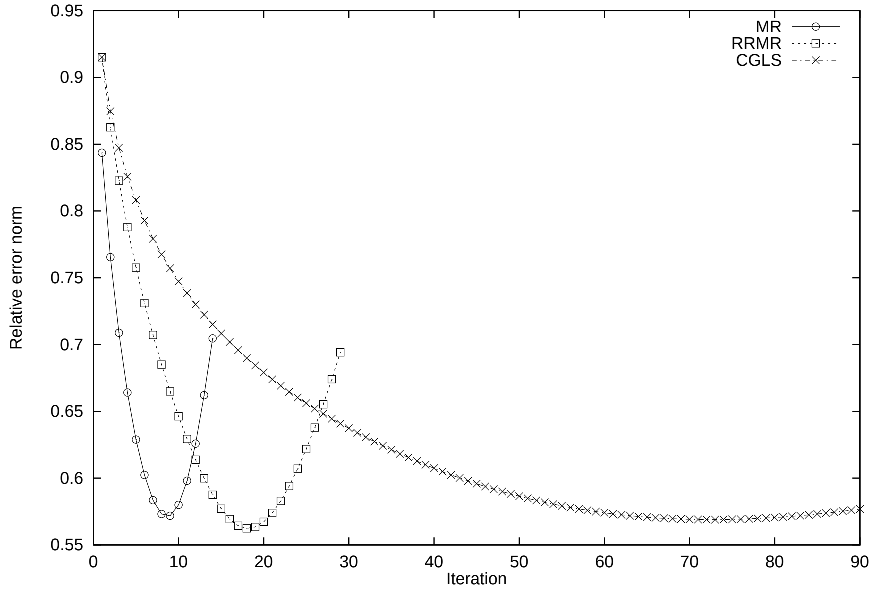

Fig. 2.3. *Example 2.1: Norm of relative errors in iterates generated by the MR, RRMR and CGLS methods.*

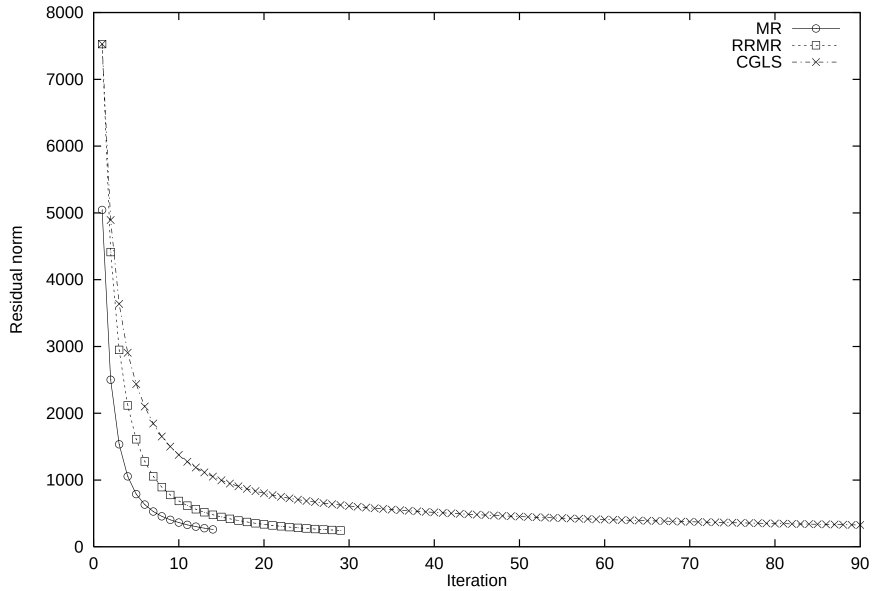

Fig. 2.4. *Example 2.1: Norm of residual errors associated with iterates generated by the MR, RRMR and CGLS methods.*

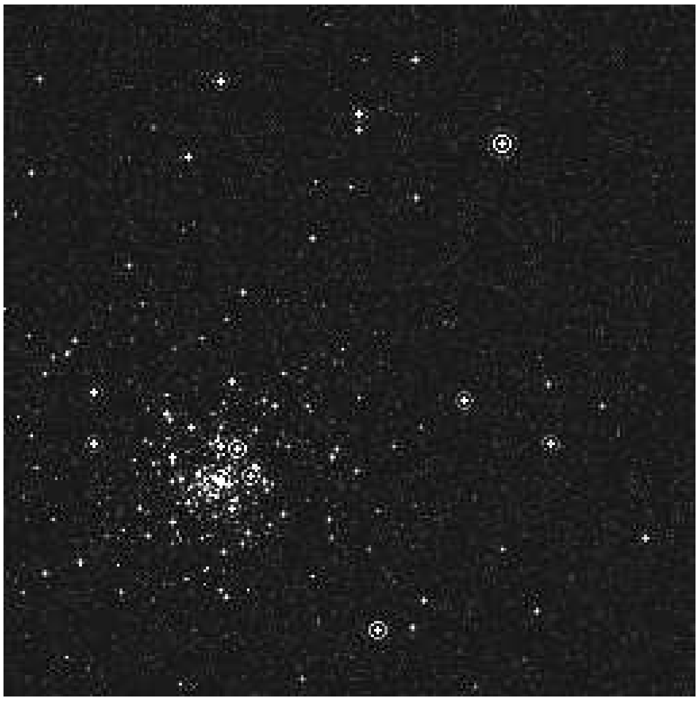

FIG. 2.5. *Example 2.1: Image restored by* 9 *iterations with the MR method.*

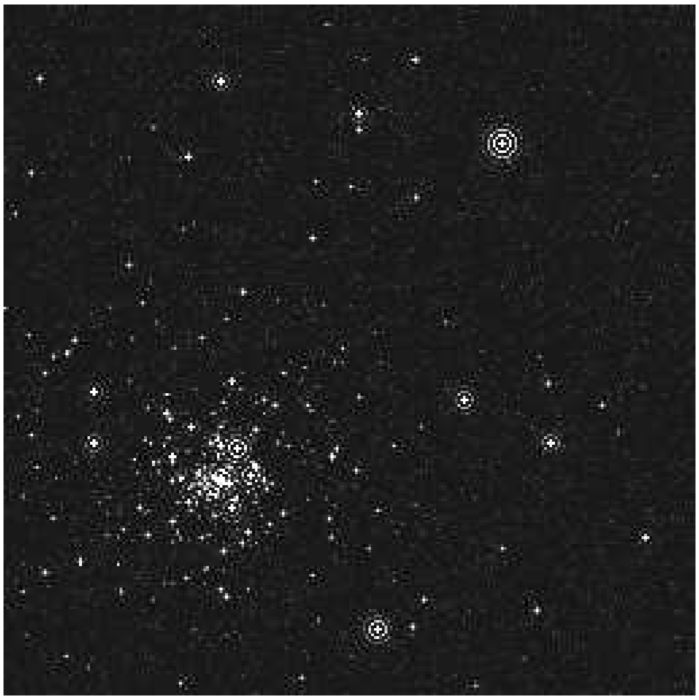

FIG. 2.6. *Example 2.1: Image restored by* 18 *iterations with the RRMR method.*

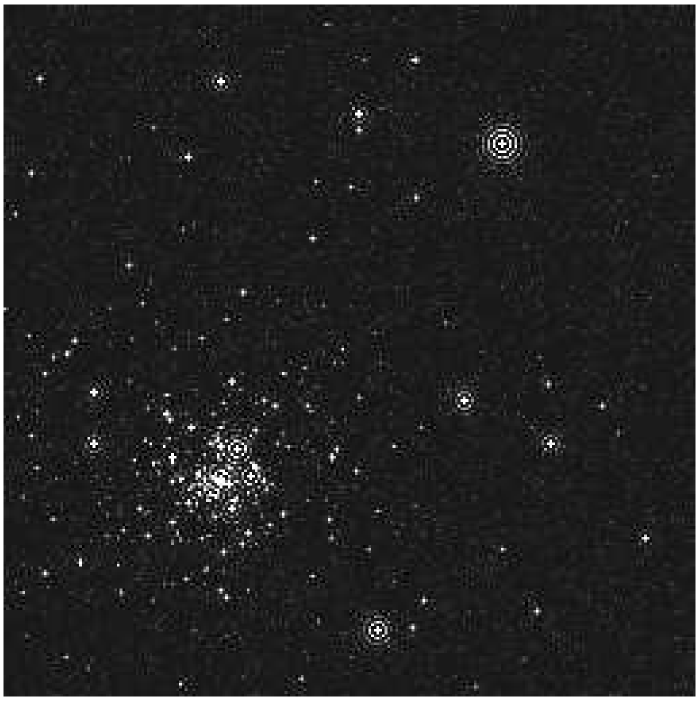

Fig. 2.7. *Example 2.1: Image restored by* 73 *iterations with the CGLS method.*

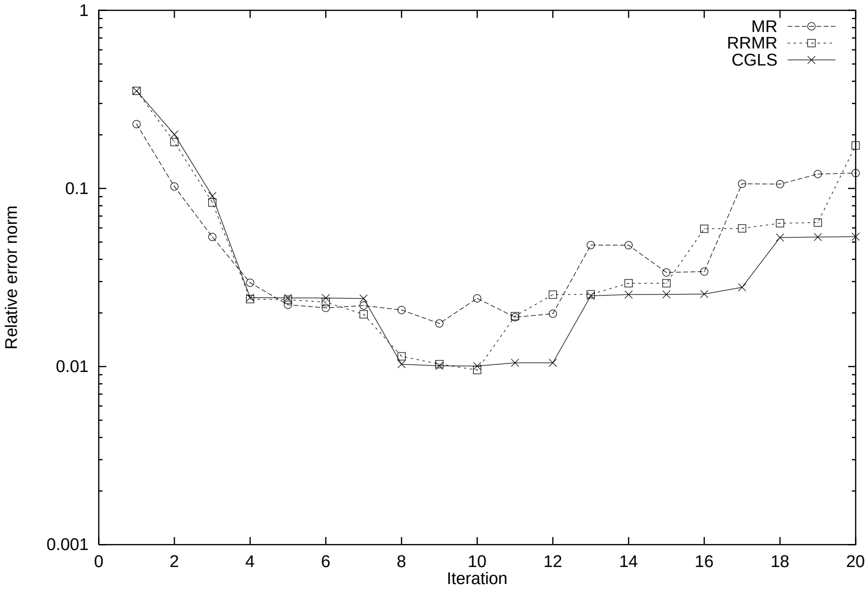

Fig. 2.8. *Example 2.2: Norm of relative errors in iterates generated by the MR, RRMR and CGLS methods.*

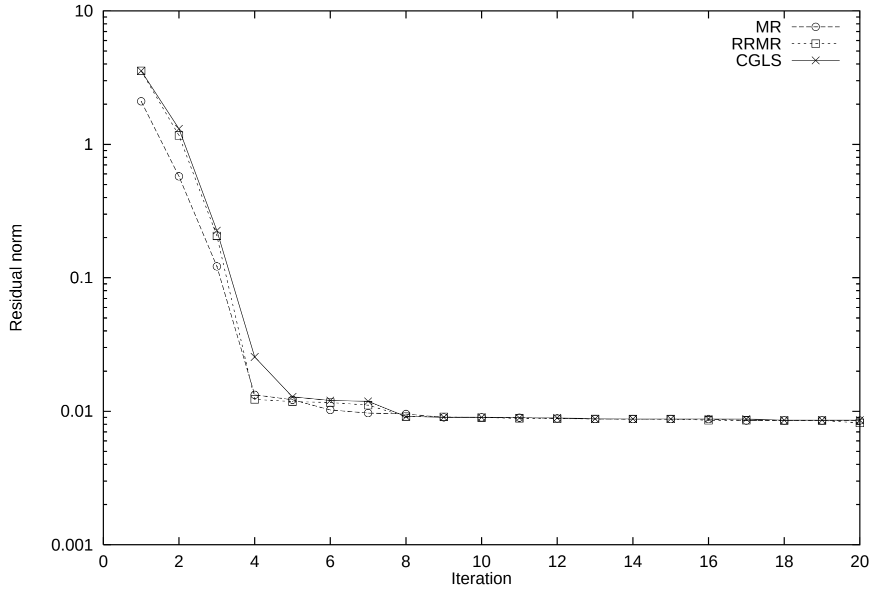

FIG. 2.9. *Example 2.2: Norm of residual errors associated with iterates generated by the MR, RRMR and CGLS methods.*

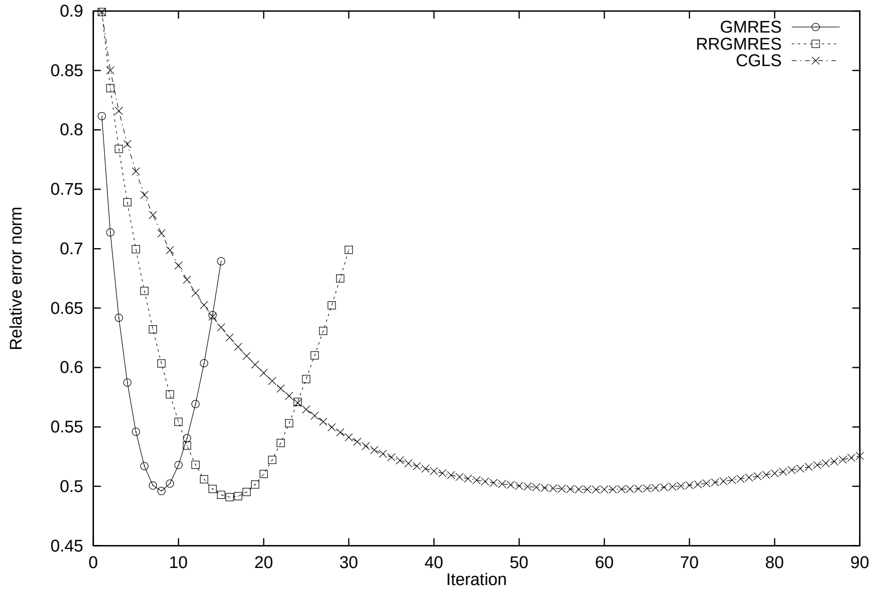

FIG. 3.1. *Example 3.1: Norm of relative errors in iterates generated by the GMRES, RRGMRES and CGLS methods.*

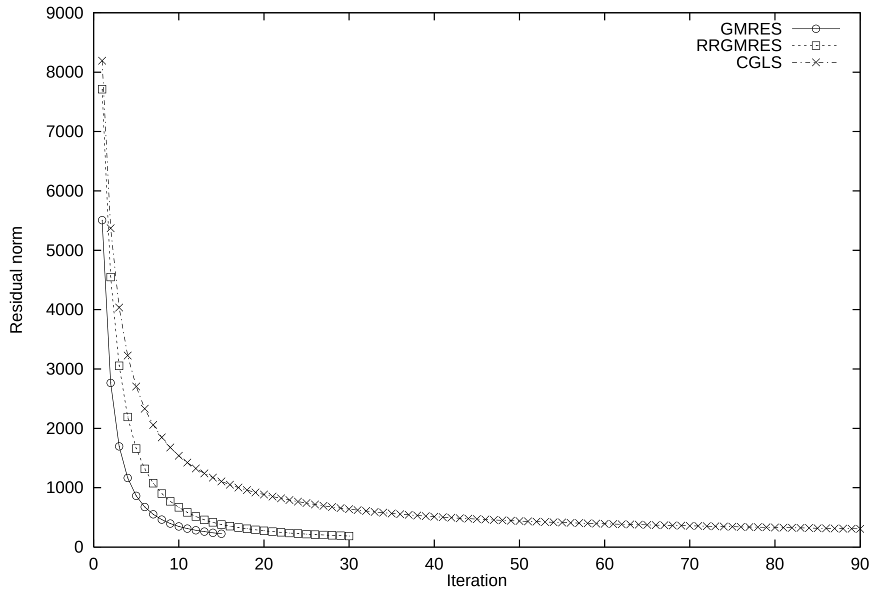

FIG. 3.2. *Example 3.1: Norm of residual errors associated with iterates generated by the GMRES, RRGMRES and CGLS methods.*

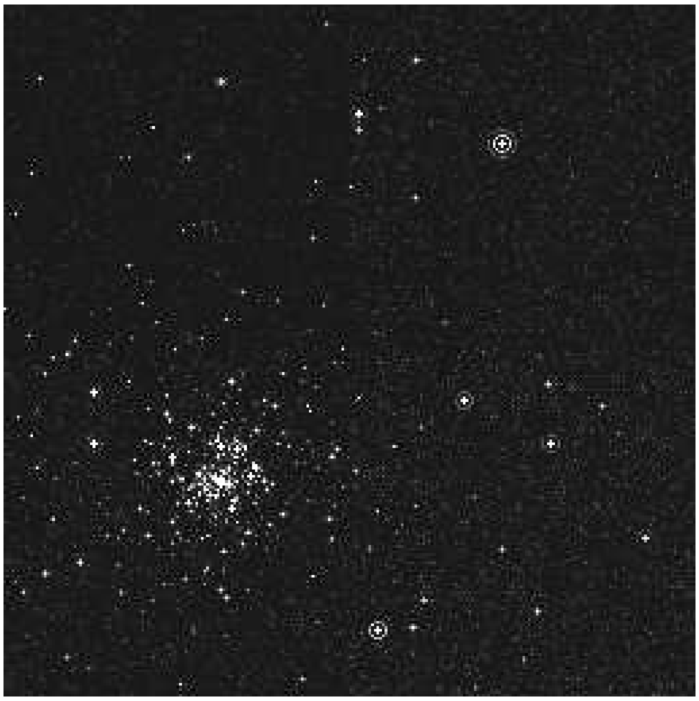

FIG. 3.3. *Example 3.1: Image restored by* 8 *iterations with the GMRES method.*

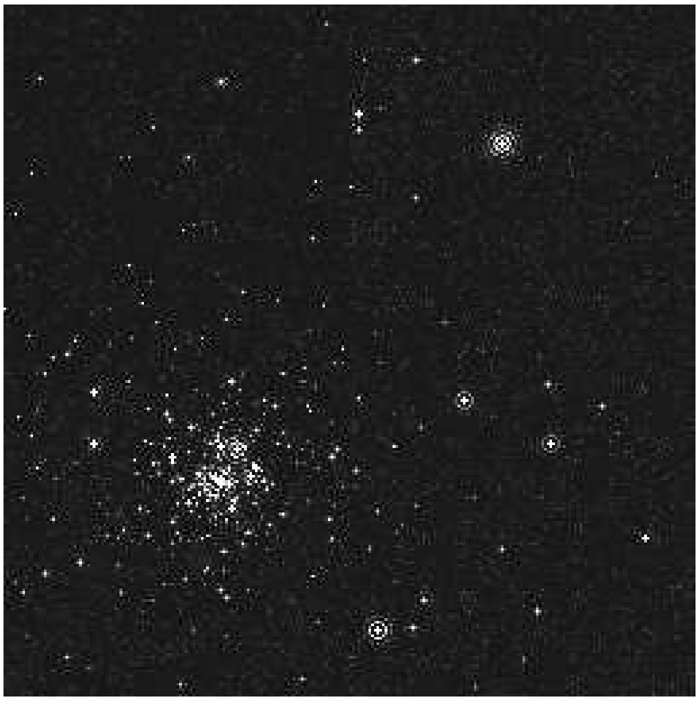

FIG. 3.4. *Example 3.1: Image restored by* 16 *iterations with the RRGMRES method.*

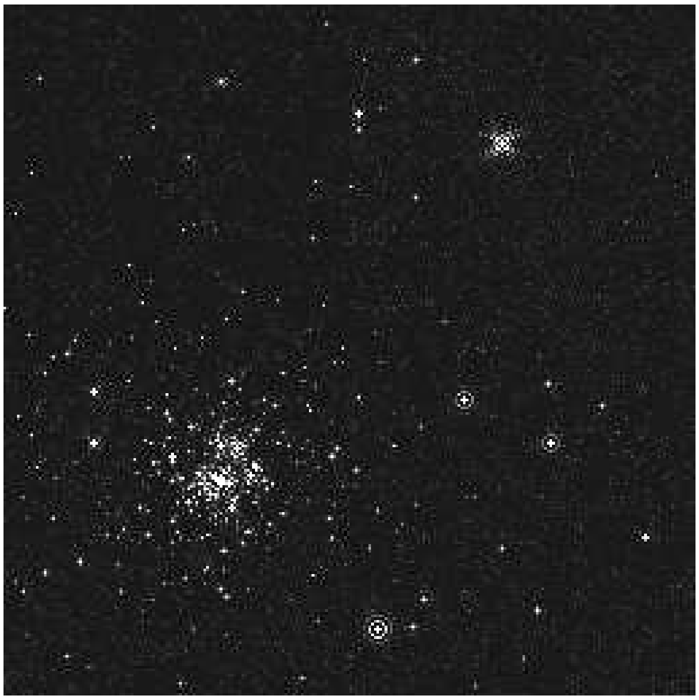

FIG. 3.5. *Example 3.1: Image restored by* 59 *iterations by the CGLS method.*

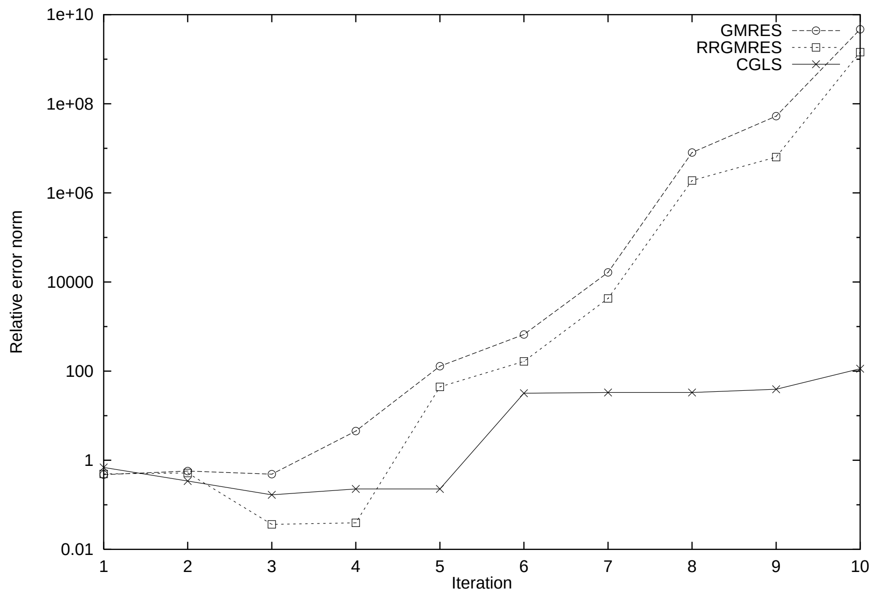

FIG. 3.6. *Example 3.2: Norm of relative errors in iterates generated by the GMRES, RRGMRES and CGLS methods.*

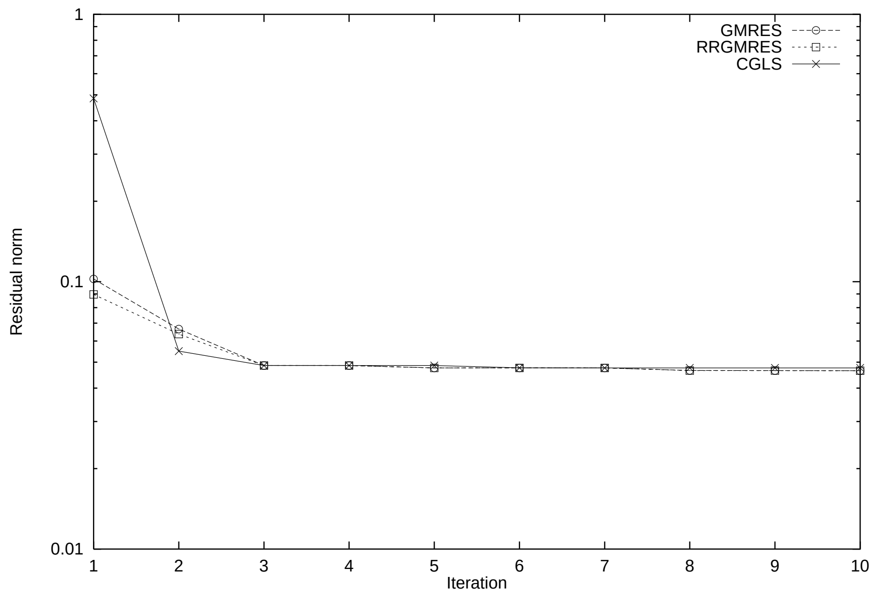

FIG. 3.7. *Example 3.2: Norm of residual errors associated with iterates generated by the GMRES, RRGMRES and CGLS methods.*

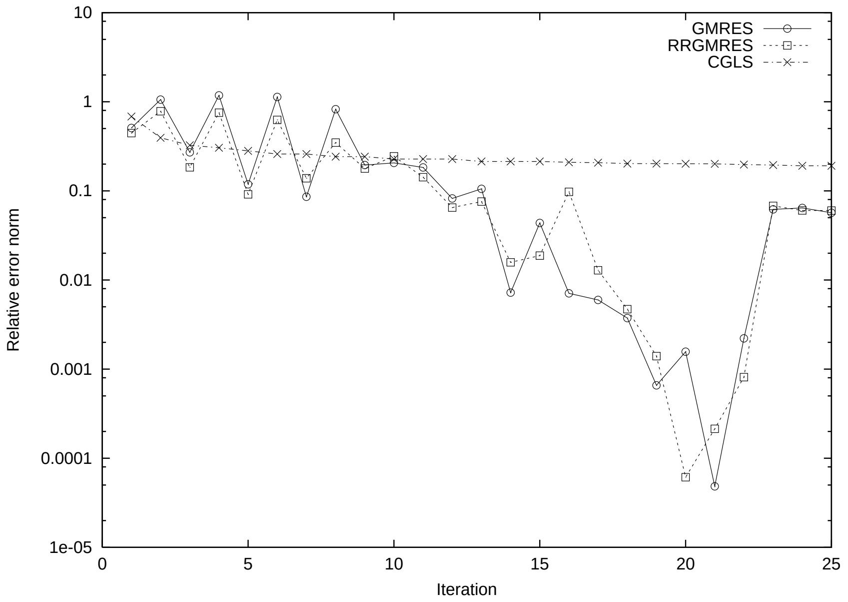

FIG. 3.8. *Example 3.3: Norm of relative errors in iterates generated by the GMRES, RRGMRES and CGLS methods.*

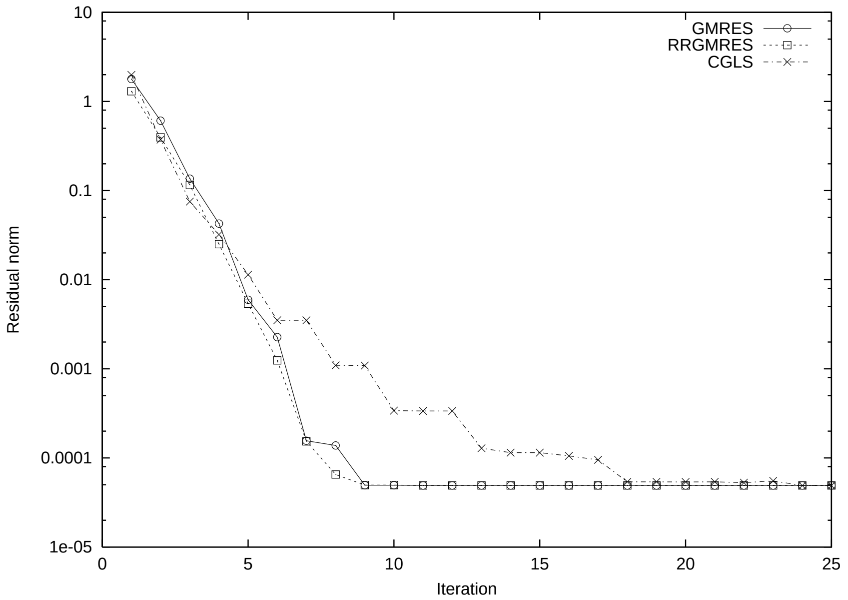

FIG. 3.9. *Example 3.3: Norm of residual errors associated with iterates generated by the GMRES, RRGMRES and CGLS methods.*