BIT Numerical Mathematics
2003, Vol. 43, No. 1, pp. 001–018

0006-3835/03/4301-0001 $16.00

ITERATIVE REGULARIZATION WITH MINIMUM-RESIDUAL METHODS*

T. K. JENSEN and P. C. HANSEN

Informatics and Mathematical Modelling, Technical University of Denmark,
Building 321, DK-2800 Lyngby, Denmark.
emails: toke.jensen@gmail.dk, pch@imm.dtu.dk

Abstract.

We study the regularization properties of iterative minimum-residual methods applied to discrete ill-posed problems. In these methods, the projection onto the underlying Krylov subspace acts as a regularizer, and the emphasis of this work is on the role played by the basis vectors of these Krylov subspaces. We provide a combination of theory and numerical examples, and our analysis confirms the experience that MINRES and MR-II can work as general regularization methods. We also demonstrate theoretically and experimentally that the same is not true, in general, for GMRES and RRGMRES – their success as regularization methods is highly problem dependent.

AMS subject classification (2000): 65F22, 65F10.

Key words: Iterative regularization, discrete ill-posed problems, GMRES, RRGMRES, MINRES, MR-II, Krylov subspaces.

1 Introduction

We study iterative methods for solution of large-scale discrete ill-posed problems of the form $A\,x = b$ with $A \in \mathbb{R}^{n\times n}$ arising from discretization of an underlying linear ill-posed problem. Our focus is on iterative regularization, and in particular the minimum-residual methods GMRES and MINRES, for which the projection onto the underlying Krylov subspace may have a regularizing effect (and the dimension of the Krylov subspace therefore acts as a regularization parameter). Our goal is to study some of the mechanisms behind this behavior.

The singular value decomposition (SVD) $A = U\Sigma V^T = \sum_{i=1}^{n} u_i\sigma_i v_i^T$ provides a natural tool for analysis of discrete ill-posed problems, for which the singular values σ_i cluster at zero, and the right-hand side coefficients $u_i^T b$ satisfy the discrete Picard condition (DPC): on average, their absolute values decay faster than the singular values.

In the presence of noise in the right-hand side b, the "naive" solution $A^{-1}b$ is completely dominated by inverted noise. Regularized solutions can be computed by truncating or filtering the SVD expansion. For example, the truncated SVD (TSVD) method yields solutions $x_k = \sum_{i=1}^{k} \sigma_i^{-1}(u_i^T b)\, v_i$. Tikhonov regularization is another well-known method which, in its standard form, takes the

*Received xxx. Revised xxx. Communicated by xxx.

form $x_\lambda = \operatorname{argmin}_x \left\{ \|b - Ax\|_2^2 + \lambda^2 \|x\|_2^2 \right\}$, and the solution x_λ can be written in terms of the SVD of A as $x_\lambda = \sum_{i=1}^{n} \sigma_i (\sigma_i^2 + \lambda^2)^{-1} (u_i^T b)\, v_i$.

In general, a filtered SVD solution takes the form

$$x_{\text{filt}} = \sum_{i=1}^{n} \phi_i \frac{u_i^T b}{\sigma_i} v_i = V \Phi \Sigma^\dagger U^T b, \tag{1.1}$$

where $\Phi = \operatorname{diag}(\phi_i)$ is a diagonal filter matrix, and the filter factors are $\phi_i \in \{0, 1\}$ for TSVD and $\phi_i = \sigma_i^2/(\sigma_i^2 + \lambda^2)$ for Tikhonov regularization. Other regularization methods take a similar form, with different expressions for the filter factors. The effect of the filter is to remove the SVD components corresponding to the smaller singular values, and thereby to stabilize the solution.

The TSVD and Tikhonov methods are not always suited for large-scale problems. An alternative is to use *iterative regularization*, i.e., to apply an iterative method directly to $A\,x = b$ or $\min_x \|b - A\,x\|_2$ and obtain a regularized solution by early termination of the iterations. These methods exhibit semi-convergence, which means that the iterative solution improves during the first iterations, while at later stages the inverted noise starts to deteriorate the solution. For example, CGLS – which implicitly applies conjugate gradients to the normal equations $A^T Ax = A^T b$ – has this desired effect [8], [9], [14].

Other minimum-residual methods have also attained interest as iterative regularization methods. For some problems with a symmetric A, the algorithms MINRES [16] and MR-II [8] (which avoid the implicit cross-product $A^T A$ in CGLS) have favorable properties [8], [10], [14], [15]; in other situations they converge slower than CGLS.

If A is nonsymmetric and multiplication with A^T is difficult or impractical to compute, then CGLS is not applicable. GMRES [17] may seem as a natural candidate method for such problems, but only a few attempts have been made to investigate the regularization properties of this method and its variant RRGMRES [3], cf. [2], [4], [5].

The goal of this work is to perform a systematic study of the regularization properties of GMRES and related minimum-residual methods for discrete ill-posed problems, similar to Hanke's study [9] of the regularization properties of CGLS. Our focus is on the underlying mechanisms, and we seek to explain why – and when – such methods can be used for regularization. The hope is that our analysis will give better intuitive insight into the mechanisms of the regularization properties of these methods, which can aid the user in the choice of method.

In §2 we outline the theory for the minimum-residual methods considered here, and in §3 we take a closer look at the basis vectors for the underlying Krylov subspaces. In §4 we perform a theoretical and experimental study of the behavior of MINRES and the variant MR-II applied to symmetric indefinite problems, and in §5 we perform a similar analysis of GMRES and the variant RRGMRES applied to nonsymmetric problems.

Table 2.1: Minimum-residual methods and their Krylov subspaces.

Matrix	**Algorithm**	**Krylov subspace**	**Solution**
Symmetric	MINRES	$\mathcal{K}_k(A, b)$	$x^{(k)}$
	MR-II	$\mathcal{K}_k(A, Ab)$	$\bar{x}^{(k)}$
Nonsymmetric	GMRES	$\mathcal{K}_k(A, b)$	$x^{(k)}$
and square	RRGMRES	$\mathcal{K}_k(A, Ab)$	$\bar{x}^{(k)}$
Any	CGLS, LSQR	$\mathcal{K}_k(A^T A, A^T b)$	$\hat{x}^{(k)}$

2 Minimum-Residual Krylov Subspace Methods

In a projection method we seek an approximate solution to $A\,x = b$ for $x \in \mathcal{S}_k$, where $\mathcal{S}_k$ is some suitable k-dimensional subspace. Minimum-residual methods are special projection methods where the criterion for choosing x amounts to minimization of the 2-norm of the residual:

$$\min_x \|b - A\,x\|_2, \qquad \text{s.t.} \qquad x \in \mathcal{S}_k.$$

For example, the TSVD solution x_k minimizes the 2-norm of the residual over the subspace $\mathcal{S}_k = \text{span}\{v_1, v_2, \ldots, v_k\}$ spanned by the first k right singular vectors. For the solution subspace $\mathcal{S}_k$ we can also use a Krylov subspace, such as $\mathcal{K}_k(A, b) \equiv \text{span}\{b, Ab, \ldots, A^{k-1}b\}$. Table 2.1 lists several Krylov subspace methods, and we note that only CGLS allows a rectangular A matrix.

The variants MINRES and MR-II for symmetric matrices are based on three-term recurrence schemes for generating the desired Krylov subspaces, while GMRES and RRGMRES need to carry along the entire set of basis vectors for their Krylov subspaces. Throughout this paper, we always use the symmetric variants when A is symmetric.

The methods RRGMRES and MR-II based on $\mathcal{K}_k(A, Ab)$ were originally designed for solving singular and inconsistent systems, and they restrict the Krylov subspace to be a subspace of $\mathcal{R}(A)$, the range of A. When $A = A^T$ this has the effect that MR-II computes the minimum-norm least squares solution [7]; for a general matrix A, RRGMRES computes the minimum-norm least squares solution when $\mathcal{R}(A) = \mathcal{R}(A^T)$ [3]. This is not the case for MINRES and GMRES.

When solving discrete ill-posed problems, we are not interested in the final convergence to the minimum-norm least squares solution, but rather in a good regularized solution. The Krylov subspace $\mathcal{K}_k(A, A\,b)$ may still be favorable to $\mathcal{K}_k(A, b)$, because the noise in the initial Krylov vector of the former subspace is damped by multiplication with A. A similar effect is automatically achieved in the CGLS subspace $\mathcal{K}_k(A^T A, A^T b)$ due to the starting vector $A^T b$. The subspace $\mathcal{K}_k(A, b)$ includes directly the noise component present in b, which can have a dramatic and undesirable influence on the early iteration vectors. For

this reason, RRGMRES and MR-II may provide better regularized solutions than GMRES and MINRES.

Our analysis of the algorithms is based on the fact that any solution in a Krylov subspace can be written in polynomial form. For example, for GMRES or MINRES we can write the kth iterate as

$$x^{(k)} = \mathcal{P}_k(A)\, b,$$

where $\mathcal{P}_k$ is a polynomial of degree $\leq k-1$. The corresponding residual $b - A\, x^{(k)}$ is therefore given by

$$b - A\, \mathcal{P}_k(A)\, b = \big(I - A\, \mathcal{P}_k(A)\big)\, b = \mathcal{Q}_k(A)\, b,$$

where $\mathcal{Q}_k(A) = I - A\, \mathcal{P}_k(A)$ is a polynomial of degree $\leq k$ with $\mathcal{Q}_k(0) = 1$. There are similar expressions for the RRGMRES/MR-II and CGLS solutions:

$$\bar{x}^{(k)} = \overline{\mathcal{P}}_{k+1}(A)\, b, \qquad \hat{x}^{(k)} = \widehat{\mathcal{P}}_k(A^T A)\, A^T b.$$

The RRGMRES/MR-II polynomial $\overline{\mathcal{P}}_{k+1}$ has degree $\leq k$ (instead of $k-1$) and the constant term is zero by definition.

The SVD of A allows us to carry out a more careful study of the Krylov subspaces. For the GMRES and RRGMRES Krylov subspaces we obtain

$$\begin{aligned} \mathcal{K}_k(A, b) &= \text{span}\{b, U\Sigma V^T b, \ldots, (U\Sigma V^T)^{k-1} b\}, \\ \mathcal{K}_k(A, Ab) &= \text{span}\{U\Sigma V^T b, (U\Sigma V^T)^2 b, \ldots, (U\Sigma V^T)^k b\}, \end{aligned}$$

respectively. If we define the orthogonal matrix C as well as the vector β by

$$C = V^T U, \qquad \beta = U^T b, \tag{2.1}$$

then the GMRES iterates $x^{(k)}$ and the RRGMRES iterates $\bar{x}^{(k)}$ satisfy

$$V^T x^{(k)} \in \mathcal{K}_k(C\Sigma, C\beta), \qquad V^T \bar{x}^{(k)} \in \mathcal{K}_k(C\Sigma, C\Sigma\, C\beta). \tag{2.2}$$

It follows that we can write the GMRES and RRGMRES solutions as

$$x^{(k)} = V\, \Phi_k\, \Sigma^\dagger \beta, \qquad \Phi_k = \mathcal{P}_k(C\,\Sigma) C\,\Sigma, \tag{2.3}$$

$$\bar{x}^{(k)} = V\, \overline{\Phi}_k\, \Sigma^\dagger \beta, \qquad \overline{\Phi}_k = \overline{\mathcal{P}}_{k+1}(C\,\Sigma) C\,\Sigma. \tag{2.4}$$

Due to the presence of the matrix C, the "filter matrices" Φ_k and $\overline{\Phi}_k$ are full, in general. Hence, neither the GMRES nor the RRGMRES iterates have a filtered SVD expansion of the form (1.1) (the SVD components are "mixed" in each iteration), and therefore we cannot expect that these iterates resemble the TSVD or Tikhonov solutions.

When A is symmetric we can write $A = V\, \Omega\Sigma\, V^T$, where $\Omega = \text{diag}(\pm 1)$ is a signature matrix and $\Omega\Sigma$ contains the eigenvalues of A. Hence $C = \Omega$, and the Krylov subspaces in (2.2) simplify to

$$V^T x^{(k)} \in \mathcal{K}_k\big(\Omega\Sigma, \Omega\beta\big), \qquad V^T \bar{x}^{(k)} \in \mathcal{K}_k(\Omega\Sigma, \Sigma\beta). \tag{2.5}$$

In this case the "filter matrices" $\Phi_k = \mathcal{P}_k(\Omega\Sigma)\Omega\Sigma$ and $\overline{\Phi}_k = \overline{\mathcal{P}}_{k+1}(\Omega\Sigma)\Omega\Sigma$ are diagonal (possibly with some negative elements), and therefore the MINRES and MR-II iterates have simple expressions in the SVD basis.

For the CGLS algorithm we have $A^TA = V\Sigma^2V^T$, and it follows that

$$\mathcal{K}_k(A^TA, A^Tb) = \text{span}\{V\Sigma U^Tb, V\Sigma^3U^Tb, \ldots, V\Sigma^{2k-1}U^Tb\}, \tag{2.6}$$

and that the CGLS iterates $\hat{x}^{(k)}$ satisfy

$$\hat{x}^{(k)} = V\,\widehat{\Phi}_k\,\Sigma^\dagger\beta, \qquad \widehat{\Phi}_k = \widehat{\mathcal{P}}_k(\Sigma^2)\Sigma^2, \tag{2.7}$$

where $\widehat{\Phi}_k$ is a diagonal matrix and $\widehat{\mathcal{P}}_k$ is the CGLS polynomial introduced above. This relation shows that the CGLS iterates $\hat{x}^{(k)}$ also have a simple expression in the SVD basis, namely, as a filtered SVD expansion of the form (1.1) with nonnegative filter factors given in terms of the polynomial $\widehat{\mathcal{P}}_k$.

The Krylov vectors for CGLS, MINRES and MR-II, respectively, have the form

$$V\Sigma^{2k-1}\beta, \qquad V(\Omega\Sigma)^{k-1}\Omega\beta \qquad \text{and} \qquad V(\Omega\Sigma)^{k-1}\Sigma\beta, \tag{2.8}$$

for $k = 1, \ldots, n$. The diagonal elements of Σ decay, and so do the coefficients in β, on average, due to the DPC. Therefore the first SVD components will, in general, be more strongly represented in these Krylov subspaces than SVD components corresponding to smaller singular values. This indicates a correspondence of the CGLS, MINRES and MR-II solutions with both TSVD and Tikhonov solutions which are also, primarily, spanned by the first right singular vectors. A similar argument cannot be used to demonstrate that the GMRES and RRGMRES solutions are regularized solutions, due to the mixing by the full matrix C in each iteration.

3 A Closer Look at the Solution Subspaces

Depending on the method used, the iterates lie in one of the three subspaces $\mathcal{R}(A)$, $\mathcal{R}([A, b])$ and $\mathcal{R}(A^T)$, as listed in Table 3.1, and this has an effect on the regularized solutions produced by the three methods. For this study, it is important to note that discrete ill-posed problems *in practise* – due to the decaying singular values and the effects of finite-precision arithmetic – behave as if the matrix A is *rank deficient*.

From Table 3.1 it is clear why CGLS and MR-II are successful as iterative regularization methods: they produce solutions in subspaces of $\mathcal{R}(A^T)$, similar to TSVD and Tikhonov. MINRES can also be used, but we get a subspace of $\mathcal{R}(A^T)$ only for consistent systems – and unfortunately discrete ill-posed problems with noisy data behave as inconsistent problems. Neither GMRES nor RRGMRES produce solutions in the desired subspace $\mathcal{R}(A^T)$.

For symmetric matrices, the DPC implies that all the Krylov vectors in Eq. (2.8) have elements which, on average, decay for increasing index i. However, due to the different powers of Σ, the damping imposed by the multiplication with the

Table 3.1: Overview of the fundamental subspaces in which the iterates lie, with $\mathcal{R}(A) = \text{span}\{u_1, \ldots, u_r\}$, $\mathcal{R}([A, b]) = \text{span}\{b, u_1, \ldots, u_r\}$, $\mathcal{R}(A^T) = \text{span}\{v_1, \ldots, v_r\}$ and r = rank of A.

Subspace	**Method**	**Krylov subspace**	**Kind of system**
$\mathcal{R}(A)$	GMRES	$\mathcal{K}_k(A, b)$	consistent systems
	RRGMRES	$\mathcal{K}_k(A, Ab)$	all systems
$\mathcal{R}([A, b])$	GMRES	$\mathcal{K}_k(A, b)$	inconsistent systems
	MINRES	$\mathcal{K}_k(A, b)$	inconsistent systems
$\mathcal{R}(A^T)$	MINRES	$\mathcal{K}_k(A, b)$	consistent systems
	MR-II	$\mathcal{K}_k(A, Ab)$	all systems
	CGLS	$\mathcal{K}_k(A^T A, A^T b)$	all systems

singular values is different for these methods. For example, the kth CGLS Krylov vector is equal to the $2k$th Krylov vector of MINRES and the $(2k-1)$st Krylov vector of MR-II. Moreover, the vector b appears undamped in the MINRES basis, while in the CGLS and MR-II bases it is always damped by A^T and A, respectively. In the presence of noise, the fact that b appears undamped in the MINRES Krylov subspace $\mathcal{R}([A, b]) = \mathcal{R}([A^T, b])$ can have a dramatic impact, as we illustrate below.

For nonsymmetric matrices the behavior of CGLS is identical to the symmetric case. On the other hand, the Krylov vectors for GMRES and RRGMRES are different, cf. (2.2); even if the DPC is fulfilled, we cannot be sure that the coefficients $|v_i^T b|$ decay, on average, for increasing index i. Furthermore, due to the presence of the non-diagonal matrix $C\Sigma$, no structured damping of these coefficients is obtained because the SVD components are "mixed." This means that GMRES and RRGMRES, in general, cannot be assumed to produce a solution subspace that resembles that spanned by the first right singular vectors.

Below we illustrate these issues with two numerical examples, in which the $n \times k$ matrices W_k, $\overline{W}_k$ and $\widehat{W}_k$ have orthonormal columns that span the Krylov subspaces for GMRES/MINRES, RRGMRES/MR-II and CGLS, respectively.

3.1 Example: Krylov Subspaces for a Symmetric Matrix

We use the symmetric problem deriv2(100,3) from Regularization Tools [11] with a 100×100 coefficient matrix, and add white Gaussian noise e to the right-hand side such that $b = A\, x_{\text{exact}} + e$ with $\|e\|_2 / \|A x_{\text{exact}}\|_2 = 5 \cdot 10^{-4}$. Figure 3.1 shows the relative errors for a series of CGLS, MINRES and MR-II iterates, as well as the relative errors of similar TSVD solutions.

MINRES does not reduce the relative error as much as the other methods due to the noise component in the initial Krylov vector. MR-II and CGLS reduce the

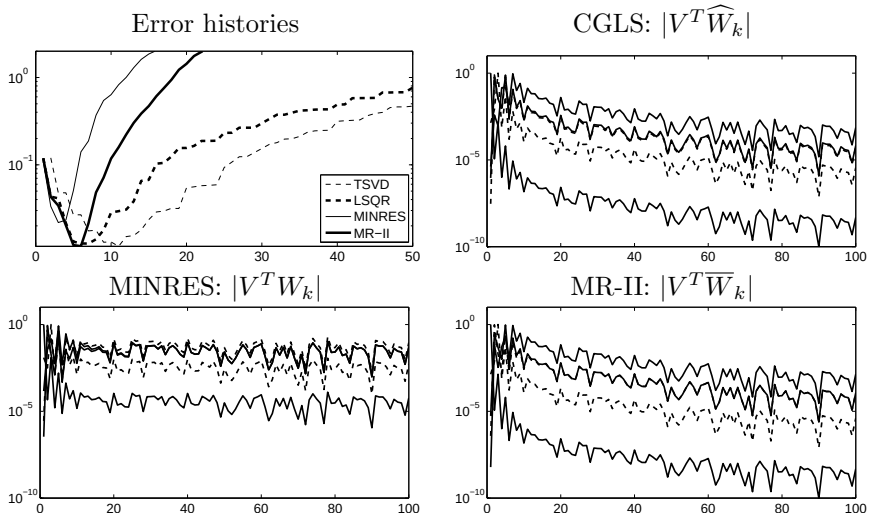


Figure 3.1: Symmetric test problem deriv2(100,3). Top left: the relative error $\|x_{\text{exact}} - x^{(k)}\|_2/\|x_{\text{exact}}\|_2$ for CGLS, MINRES and MR-II, and $\|x_{\text{exact}} - x_k\|_2/\|x_{\text{exact}}\|_2$ for TSVD. Remaining plots: the first five orthonormal Krylov vectors of the CGLS, MINRES and MR-II subspaces in the SVD basis.

relative error to about the same level, 0.0117 for MR-II and 0.0125 for CGLS, in 5–6 iterations. This makes MR-II favorable for this problem, because the number of matrix-vector multiplications is halved compared to CGLS. The best TSVD solution includes 11 SVD components, which indicates that the CGLS and MR-II solution subspaces are superior compared to the TSVD solution subspace of equal dimensions. This fact was originally noticed by Hanke [9].

Figure 3.1 also shows the first five orthonormal Krylov vectors expressed in terms of the right singular vectors v_i, i.e., the first five columns of $|V^TW_k|$, $|V^T\overline{W}_k|$ and $|V^T\widehat{W}_k|$. We see that the CGLS and MR-II vectors are mainly spanned by the first right singular vectors as expected, and that the contribution from the latter singular vectors is damped. We also see that the contribution from the latter right singular vectors is much more pronounced for MINRES due to the direct inclusion of the noise.

3.2 Example: Krylov Subspaces for a Nonsymmetric Matrix

Here we use the nonsymmetric problem ilaplace(100) from Regularization Tools [11] with a coefficient matrix of size 100×100 and additive white Gaussian noise e with $\|e\|_2/\|Ax_{\text{exact}}\|_2 = 5 \cdot 10^{-4}$. Figure 3.2 shows plots similar to those for the symmetric case. Only TSVD and CGLS are able to reduce the relative error considerably; neither GMRES nor RRGMRES produce iterates with small relative errors.

We note that the CGLS Krylov vectors behave similar to the symmetric case, i.e., the components that correspond to small singular values are effectively

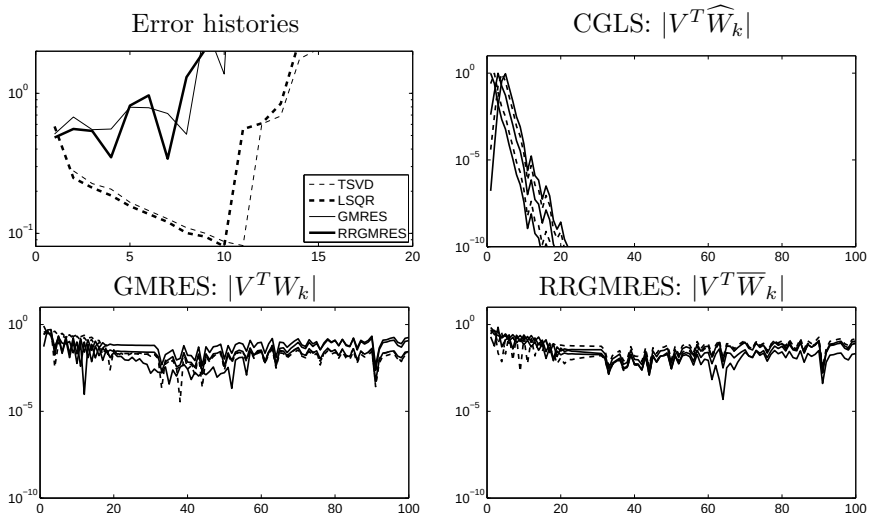


Figure 3.2: Nonsymmetric test problem ilaplace(100). Top left: the relative errors $\|x_{\text{exact}} - x^{(k)}\|_2/\|x_{\text{exact}}\|_2$ for CGLS, GMRES and RRGMRES, and $\|x_{\text{exact}} - x_k\|_2/\|x_{\text{exact}}\|_2$ for TSVD. Remaining plots: the first five orthonormal Krylov vectors of the CGLS, GMRES and RRGMRES subspaces in the SVD basis.

damped. Furthermore, we see that the GMRES and RRGMRES Krylov vectors do not exhibit any particular damping. In fact, all Krylov vectors contain significant components along all right singular vectors – including those that correspond to small singular values. Therefore the GMRES and RRGMRES iterates are composed not only of the first right singular vectors, they also include significant components in the direction of the last right singular vectors.

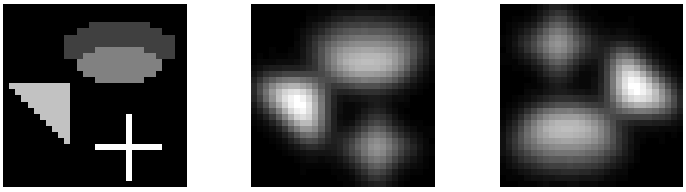
4 Iterative Regularization with MINRES and MR-II

We can express the MINRES and MR-II residual norms as

$$\|b - A\,x^{(k)}\|_2 = \|\mathcal{Q}_k(\Omega\Sigma)\,\beta\|_2, \qquad \|b - A\,\bar{x}^{(k)}\|_2 = \|\overline{\mathcal{Q}}_{k+1}(\Omega\Sigma)\,\beta\|_2,$$

where $\mathcal{Q}_k$ is the MINRES residual polynomial defined in Section 2, and $\overline{\mathcal{Q}}_{k+1} = I - A\,\overline{\mathcal{P}}_{k+1}(A)$ is the MR-II residual polynomial. Since these methods minimize the residual's 2-norm in each iteration, the effect of the residual polynomial is to "kill" the large components of $|\beta|$. Hence the residual polynomials must be small at those eigenvalues for which $|\beta|$ has large elements. (On the other hand, if a component $|\beta|$ is small, then the corresponding value of the residual polynomial need not be small.) Thus, MINRES and MR-II have the same intrinsic regularization property as the CGLS algorithm.

The main difference is that CGLS only has to "kill" components for positive eigenvalues of A^TA, while – for indefinite matrices – MINRES and MR-II must "kill" components corresponding to both positive and negative eigenvalues of A. The latter is more difficult, due to the polynomial constraints



$x = \mathrm{vec}(X)$ $b = \mathrm{vec}(B)$ Pb

Figure 4.1: True and blurred images for the symmetric two-dimensional problem.

$\mathcal{Q}_k(0) = \overline{\mathcal{Q}}_{k+1}(0) = 1$. These issues have been studied in more detail for MINRES [14], [15], and also for general symmetric minimum-residual methods, see [6] and the references therein.

Below we present two examples which illustrate the above observations and support the results obtained by Kilmer and Stewart [15]. Our examples also illustrate that the definiteness of the coefficient matrix affects the convergence and the iterates of the MINRES and MR-II, while both methods still produce regularized solutions. Furthermore, we demonstrate that MINRES and MR-II concentrate on the components that are significant for reducing the residual, i.e., the large right-hand side components in the eigenvector basis. A similar observation was done by Hanke [9] for CGLS.

4.1 Example: Image Deblurring with a Symmetric Matrix

Let $X \in \mathbb{R}^{30\times 30}$ be the sharp image seen in Fig. 4.1 (left), and let the matrix A be a discretization of a two-dimensional Gaussian blurring [12] with zero boundary conditions. The blurred image $B \in \mathbb{R}^{30\times 30}$ is shown in Fig. 4.1 (middle). The discrete ill-posed problem takes the form $A\,x = b$ where $x = \mathrm{vec}(X)$ is the column-wise stacked image X, and $b = \mathrm{vec}(B)$ is the stacked image B. We added white Gaussian noise to b with $\|e\|_2/\|A\,x_{\mathrm{exact}}\|_2 = 10^{-2}$.

The coefficient matrix A is both symmetric and persymmetric, i.e., $A = A^T$ and $PA = (PA)^T$, where P is the reversal matrix. Moreover, A is positive definite and PA is indefinite. Since the 2-norm is invariant under multiplication with P it follows that $\|b - A\,x\|_2 = \|P\,b - PA\,x\|_2$ where the permuted vector Pb represents a 180° rotation of B as shown in Fig. 4.1 (right). We apply CGLS, MINRES and MR-II to the two problems

$$A\,x = b \qquad \text{and} \qquad PA\,x = P\,b. \tag{4.1}$$

Obviously, the CGLS Krylov subspaces for the two problems are identical because $\mathcal{K}_k(A^TP^TPA, A^TP^TPb) = \mathcal{K}_k(A^TA, A^TA)$. However, this is not the case for MINRES and MR-II because $\mathcal{K}_k(A, b) \neq \mathcal{K}_k(PA, Pb)$ and $\mathcal{K}_k(A, Ab) \neq \mathcal{K}_k(PA, PAPb)$.

Figure 4.2 shows the reduction of the residual norms for the three methods applied to both versions of the problem. No reorthogonalization of the Krylov vectors is performed. Obviously, the convergence of CGLS is the same for the

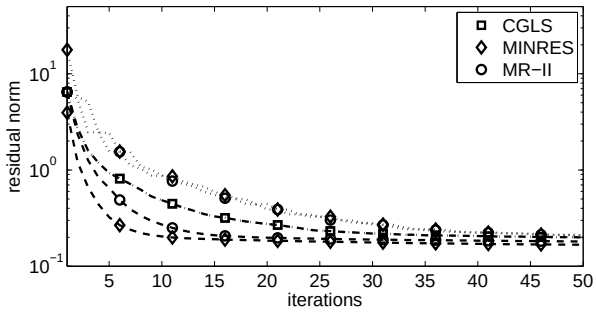


Figure 4.2: Reduction of the residual for CGLS, MINRES and MR-II, for the original problem (dashed lines) and the permuted problem (dotted lines) in (4.1).

Table 4.1: Number of iterations and relative error for best iterates, for the deblurring problems in §4.1 and §4.2.

	CGLS		MINRES		MR-II	
Problem	its	rel. err	its	rel. err	its	rel. err
$A\,x = b$	59	0.4680	6	0.4916	20	0.4678
$PA\,x = P\,b$	59	0.4680	81	0.4682	79	0.4681
$A\,x' = b'$	84	0.4823	9	0.5166	24	0.4817
$PA\,x' = P\,b'$	83	0.4823	95	0.4864	93	0.4834

two problems, whereas the convergence of MINRES and MR-II is *faster* than the convergence of CGLS when applied to $Ax = b$, and *slower* when applied to the permuted problem.

Figure 4.3 shows the eigenvalues for the two problems, together with the residual polynomials for the first four iterations of MINRES and MR-II. Both residual polynomials satisfy $\mathcal{Q}_k(0) = 1$ and $\overline{\mathcal{Q}}_{k+1}(0) = 1$, and in addition $\overline{\mathcal{Q}}'_{k+1}(0) = 0$. We see that the residual polynomials behave better – i.e., they are small for a greater range of eigenvalues – when all eigenvalues are positive. This explains why the convergence is faster for the problem with the positive definite matrix A.

Table 4.1 gives more information about the convergence of the two methods for the two problems, namely, the number of iterations and the relative error of the iterates with the smallest relative errors (compared to the exact solution). CGLS performs identically for the two problems. For the original problem, CGLS and MR-II produce slightly better solutions than MINRES, and MR-II needs much fewer iterations than CGLS; moreover MINRES produces a slightly inferior solution in only six iterations. For the permuted problem all three methods produce solutions of the same quality. MR-II and MINRES are comparable in speed and faster than CGLS (because CGLS needs two matrix-vector products per iteration).

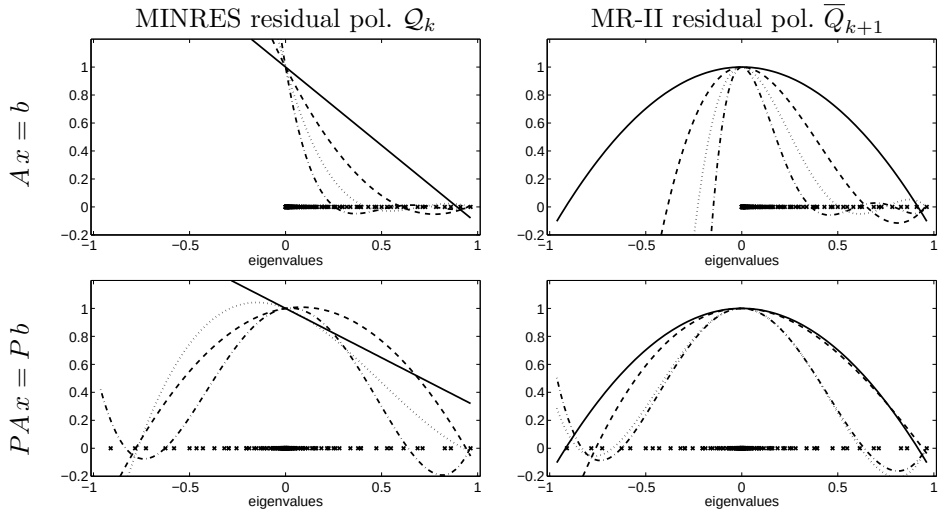
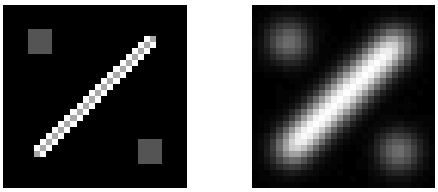


Figure 4.3: The first four residual polynomials (solid, dashed, dotted, and dash-dotted lines) for MINRES and MR-II applied to the original positive definite problem $A\,x = b$ and the permuted indefinite problem $PA\,x = P\,b$. The eigenvalues of A and PA, respectively, are shown by the small crosses.

4.2 Example: The Role of the Solution Coefficients

The residual polynomials of the methods also depend on the coefficients of the solution (and the right-hand side) in the eigenvector basis; not just the eigenvalues. To illustrate this, we create another sharp image X' which is invariant to a 180° rotation, i.e., the column-wise stacked image satisfies $Px' = x'$. The symmetries of A imply that the blurred image B' also satisfies $Pb' = b'$, and again the noise level in b is such that $\|e\|_2/\|Ax_{\text{exact}}\|_2 = 10^{-2}$. The sharp and blurred images are shown in Fig. 4.4.

For this particular right-hand side, the components in the eigenvector basis that correspond to negative eigenvalues of PA are small (they are nonzero



$x' = \text{vec}(X')$ $\quad$ $b' = \text{vec}(B')$

Figure 4.4: True and blurred images for the modified problem $A\,x' = b'$.

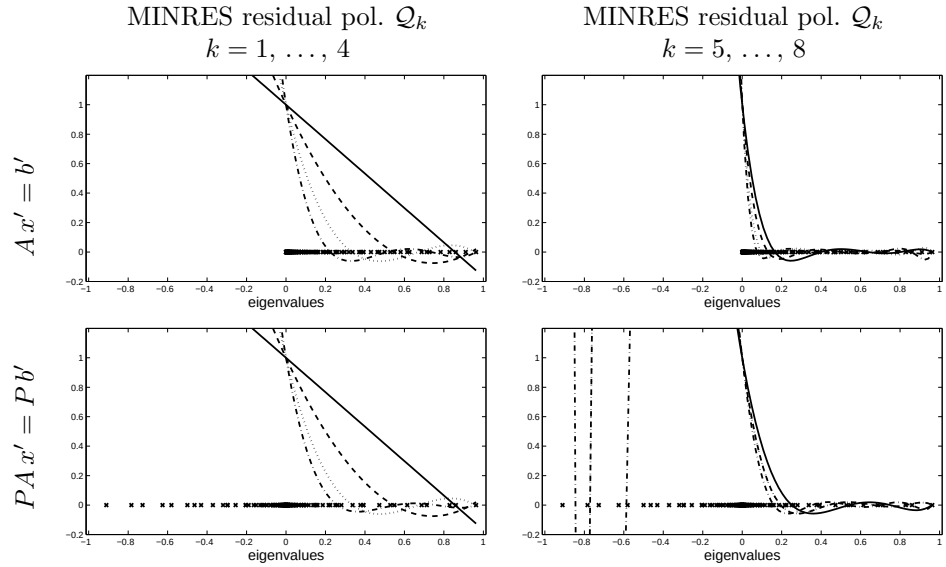


Figure 4.5: MINRES residual polynomials for the modified problems $A\,x' = b'$ and $PA\,x' = P\,b'$ (where A is positive definite and PA is indefinite). The eigenvalues are shown by the small crosses. The first four residual polynomials (bottom left plot) are not affected by the negative eigenvalues of PA, while the next four polynomials (bottom right plots) are.

solely due to the added noise). Therefore, in the initial iterations the residual polynomials need not pay as much attention to the negative eigenvalues. Figure 4.5 shows the MINRES residual polynomials corresponding to the first eight iterations, for both problems $A\,x' = b'$ and $PA\,x' = P\,b'$. Note that the first four polynomials are practically identical for the two problems, showing that MINRES is not affected by the small components corresponding to the negative eigenvalues. For the next four iterations, the small noise components for the negative eigenvalues start to affect the residual polynomials, thus slowing down the convergence. The situation is similar for the MR-II polynomials and not shown here.

This example illustrates that – at least in theory – the convergence for an indefinite problem can be similar to that of a definite problem. In practise, however, when noise is present in the right-hand side the convergence is always slower for the indefinite problem. Table 4.1 shows the convergence results which, in essence, are very similar to those for the previous example: both MINRES and MR-II produce regularized solutions, and in terms of computational work they are both favorable compared to CGLS.

5 Iterative Regularization with GMRES and RRGMRES

We now consider systems with nonsymmetric coefficient matrices and the Krylov methods GMRES and RRGMRES. Saad and Schultz [17, §3.4] showed that for any nonsingular matrix A the GMRES iterations do not break down until the exact solution is found. On the other hand, as noted, e.g., by Brown and Walker [1], anything may happen when A is singular. Our interest is in numerically singular systems (i.e., systems with tiny singular values), and we distinguish between rank-deficient problems and ill-posed problems.

5.1 Rank-Deficient Problems

Rank-deficient problems are characterized by having a distinct gap between "large" and "small" singular values. If the underlying operator has a null space, then the matrix A has small singular values whose size reflects the discretization scheme and the machine precision. Therefore, it makes sense to define the numerical subspaces $\mathcal{R}(A)$, $\mathcal{N}(A^T)$, $\mathcal{R}(A^T)$ and $\mathcal{N}(A)$, where the null spaces are spanned by the singular vectors corresponding to the small singular values.

If A is rank-deficient and the noise in the right-hand side is so small that it primarily affects the SVD components outside $\mathcal{R}(A)$, then the minimum-norm least squares solution $A^\dagger b$ (which is really a TSVD solution) is a good regularized solution. On the other hand, if the noise in the right-hand side is larger such that it also affects the components of b in $\mathcal{R}(A)$, then the problem effectively acts like a discrete ill-posed problem, and the solution must be further regularized to minimize the effect of the noise.

It was shown by Brown and Walker [1, Thm. 2.4] that GMRES computes the minimum-norm least squares solution if the system fulfills $\mathcal{N}(A) = \mathcal{N}(A^T)$ and if it is consistent. In this case GMRES constructs a solution in $\mathcal{R}(A) = \mathcal{R}(A^T)$, and it is obvious that if no solution components in $\mathcal{R}(A)$ are too affected by the noise, then GMRES will eventually produce a suitable regularized solution.

5.2 Example: Numerically Rank-Deficient Problem

When the noise level is small, then the test problem `heat` from Regularization Tools [11] behaves as a numerically rank-deficient problem. For $n = 100$ the first 97 singular values lie between 10^{-7} and 10^{-3}, while the last three are of the order 10^{-15}. We add white Gaussian noise to b with $\|e\|_2/\|Ax_{\text{exact}}\|_2 = 10^{-8}$, such that the first 97 SVD components of b are almost unaffected by the noise.

Figure 5.1 shows the relative errors compared to the exact solution for GMRES, RRGMRES and CGLS, and we see that only CGLS converges. The reason is that for this problem we have $\mathcal{N}(A) = \text{span}\{v_{98}, v_{99}, v_{100}\} \neq \mathcal{N}(A^T) = \text{span}\{u_{98}, u_{99}, u_{100}\}$ (see Fig. 5.2), such that neither GMRES nor RRGMRES are guaranteed to produce the desired minimum-norm least squares solution.

5.3 Discrete Ill-Posed Problems

For discrete ill-posed problems, the notion of numerical subspaces is not well defined due to decaying singular values with no gap in the spectrum. Moreover,

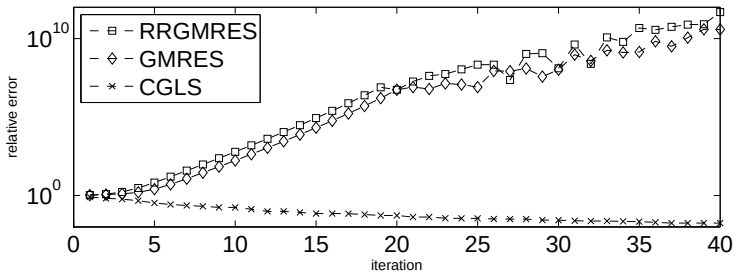


Figure 5.1: Relative errors for the first 40 iterations of GMRES, RRGMRES and CGLS applied to the rank-deficient inverse heat problem.

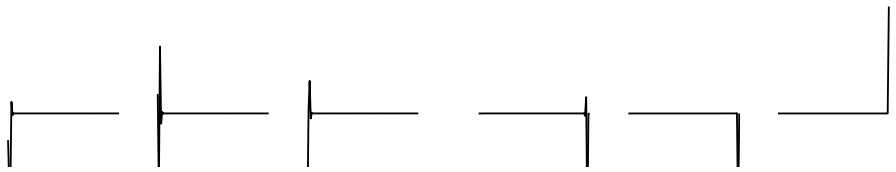


Figure 5.2: The vectors spanning $\mathcal{N}(A)$ and $\mathcal{N}(A^T)$ for the inverse heat problem. From left to right, the left null vectors u_{98}, u_{99}, u_{100} and the right null vectors v_{98}, v_{99}, v_{100}.

for GMRES and RRGMRES a *mixing* of the SVD components occurs in each iteration, due to the presence of the matrix $C = V^T U$ in the expression for the Krylov vectors, cf. (2.2). The mixing takes place right from the start:

$$v_i^T b = \sum_{j=1}^{n} c_{ij}\, \beta_j, \qquad v_i^T A\, b = \sum_{j=1}^{n} \left(\sum_{\ell=1}^{n} \sigma_\ell\, c_{i\ell}\, c_{\ell j} \right) \beta_j \ .$$

An important observation here is that the noisy components of b are also mixed by C, and therefore the non-diagonal filters of GMRES and RRGMRES not only change the solution subspaces, but also mix the contributions from the noisy components. This is contrary to spectral filtering methods such as TSVD, Tikhonov regularization, CGLS and MINRES/MR-II. In particular, the first iterations of GMRES and RRGMRES need not favor the less noisy SVD components in the solution, and the later iterations need not favor the more noisy components.

5.4 Example: Image Deblurring – GMRES and RRGMRES Work

GMRES and RRGMRES were proposed for image deblurring in [2] and [4]; here we consider a test problem similar to Example 4.3 in [4] with spatially

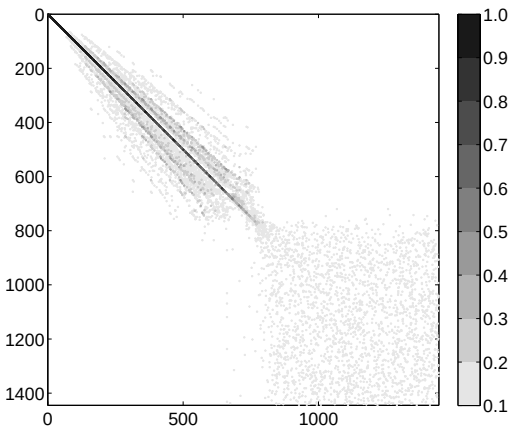


Figure 5.3: Illustration of the "structure" of the matrix $C = V^T U$ for the image deblurring problem; only matrix elements $|c_{ij}| \geq 0.1$ are shown.

variant Gaussian blur, in which the nonsymmetric coefficient matrix is given by

$$A = \begin{pmatrix} I_o & 0 \\ 0 & 0 \end{pmatrix} (T_1 \otimes T_1) + \begin{pmatrix} 0 & 0 \\ 0 & I_o \end{pmatrix} (T_2 \otimes T_2),$$

where I_o is the $\frac{n}{2} \times \frac{n}{2}$ identity matrix, and T_1 and T_2 are $N \times N$ Toeplitz matrices:

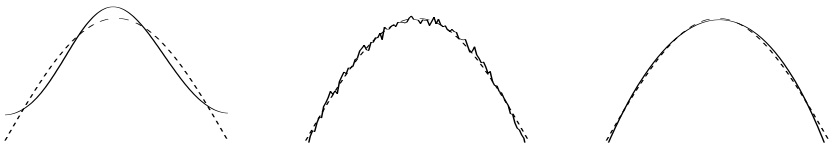
$$(T_\ell)_{ij} = \frac{1}{\sigma_\ell \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{i-j}{\sigma_\ell} \right)^2 \right), \qquad \ell = 1, 2$$

with $\sigma_1 = 4$ and $\sigma_2 = 4.5$. This models a situation where the left and right halves of the $N \times N$ image are degraded by two different point spread functions.

Here we use $N = 38$ (such that $n = N^2 = 1444$). For a problem of this size we can explicitly calculate the SVD of A. The singular values (not show here) decay gradually from $\sigma_1 = 1$ to $\sigma_{800} \approx 10^{-16}$, while the remaining singular values stay at this level. The "structure" of $C = V^T U$ is illustrated in Fig. 5.3 which shows all elements $|c_{ij}| \geq 0.1$. Although A is nonsymmetric, the matrix C is close to diagonal in the upper left corner (which corresponds to the numerically nonzero singular values). Therefore, GMRES and RRGMRES will only introduce a limited amount of "mixing" between the SVD components of the Krylov subspaces, and moreover the SVD components corresponding to the numerically zero singular values will not enter the Krylov subspaces. Hence GMRES and RRGMRES can be expected to compute regularized solutions in the SVD basis, confirming the experimental results in [4].

5.5 Example: Test Problem baart *– GMRES and RRGMRES Work*

Here we use the test problem baart(100) from Regularization Tools [11], and we added white Gaussian noise e to the right-hand side such that $\|e\|_2 / \|A\, x_{\text{exact}}\|_2 =$



CGLS $k = 4$ GMRES $k = 3$ RRGMRES $k = 3$

Figure 5.4: Best iterates of CGLS, GMRES and RRGMRES applied to the baart(100) test problem (the exact solution is shown by the dotted lines).

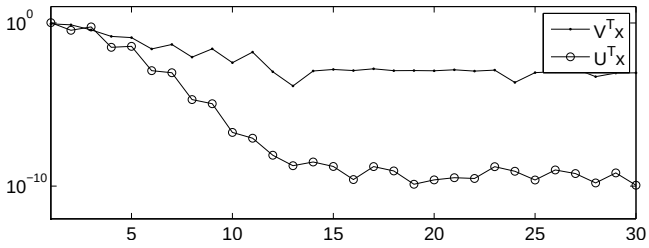


Figure 5.5: The coefficients of exact solution x^{exact} to the baart(100) test problem in the bases of the left and right singular vectors.

10^{-3}. Figure 5.4 shows the best CGLS, GMRES and RRGMRES iterates. Note that especially RRGMRES approximates the exact solution very well; the best GMRES solution is somewhat more noisy, but it also approximates the exact solution quite well. The best CGLS solution is not noisy, but neither is it a good approximation to the exact solution.

For this problem, the matrix $C = V^T U$ (not shown here) is far from a diagonal matrix, and hence the SVD components are considerably mixed in the Krylov subspaces of GMRES and RRGMRES. The reason why these methods are able to approximate the exact solution so well is that the solution subspaces (spanned by the left singular vectors u_i) are favorable for approximating this particular solution.

Figure 5.5 shows the coefficients of the exact solution x^{exact} in the left and right singular vector bases. The singular values σ_i for $i > 15$ are at the rounding error level, and therefore the coefficients $u_i^T x^{\text{exact}}$ and $v_i^T x^{\text{exact}}$ are unreliable for $i > 15$. We see that x^{exact} is indeed well expressed by a few left singular vectors u_i, while more right singular vectors v_i (the "usual" basis vectors for regularized solutions) are needed. I.e., the solution we seek is better represented in a small subspace of $\mathcal{R}(A)$ than in a subspace of $\mathcal{R}(A^T)$, cf. Table 2.1. For this particular problem, both CGLS, TSVD and Tikhonov regularization are not particularly well suited, and GMRES and RRGMRES happen to produce better solution subspaces and better regularized solutions.

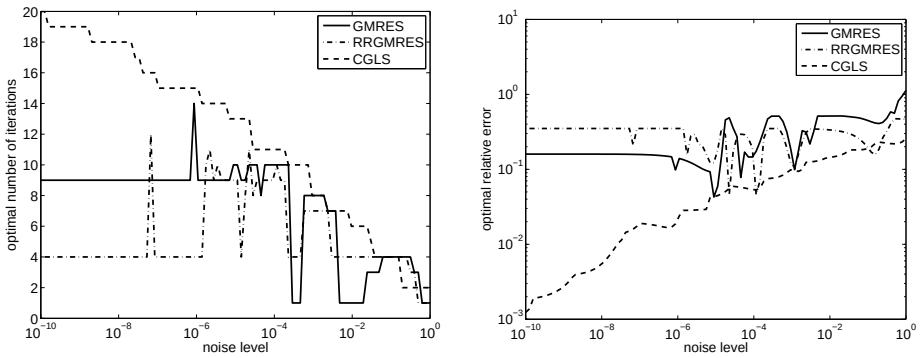


Figure 5.6: Results for the ilaplace(100) test problem. Left: the iteration count k for optimal CGLS, GMRES and RRGMRES solutions as a function of the noise level. Right: the corresponding relative errors $\|x^{\text{exact}} - x^{(k)}\|_2 / \|x^{\text{exact}}\|_2$ versus the noise level.

5.6 Example: Test Problem ilaplace – GMRES and RRGMRES Fail

Finally, we consider the nonsymmetric problem ilaplace(100) from Regularization Tools [11]. The singular values decay gradually and hit the machine precision around index 30. The noise vector e contains white Gaussian noise, and we use 100 different scalings of e such that $\|e\|_2/\|Ax_{\text{exact}}\|_2$ is logarithmically distributed between 10^{-10} and 10^0.

For each noise level, we perform 30 iterates of CGLS, GMRES and RRGMRES, and the iterates with smallest relative errors are found for all methods and all noise levels. Figure 5.6 shows the iteration count for the best solution as a function of the noise level, as well as the corresponding relative errors versus the noise level.

For CGLS there is a clear correspondence between the noise level and the number of iterations – for a high noise level only few iterations can be performed before the noise distorts the solutions, while more iterations can be performed for a lower noise level. The plot of the relative errors confirms that, overall, the solutions get better as the noise level decreases.

The same behavior is *not* observed for GMRES and RRGMRES. For small noise levels, the solutions do not improved beyond a certain level, and for higher noise levels the optimal number of iterations is unpredictable. For example, for the noise level 10^{-10}, the 4th RRGMRES iterate is the optimal solution, with a relative error of 0.3500, but for the higher noise level $1.15 \cdot 10^{-4}$ the best RRGMRES iterate is the 10th, with a relative error of only 0.04615.

As mentioned in §5.3 the sensitivity to the noise is indeed very different for GMRES and RRGMRES than for CGLS. The behavior observed for this example indicates that it may be very difficult to find suitable stopping rules for GMRES and RRGMRES.

5.7 Normal Matrices

If A is normal then we can write $A = V D V^T$, where V is orthogonal and D is block diagonal with 1×1 blocks d_i and 2×2 blocks D_i (see, e.g., [13, §2.5]). Here, $d_i = g_i \, \sigma_i$ with $g_i = \text{sign}(d_i)$ and $\sigma_i = |d_i|$, while

$$D_i = \begin{pmatrix} a_i & b_i \\ -b_i & a_i \end{pmatrix} = \sigma_i \, G_i, \qquad \sigma_i = \sqrt{a_i^2 + b_i^2} \qquad G_i = \begin{pmatrix} c_i & s_i \\ -s_i & c_i \end{pmatrix},$$

i.e., the 2×2 blocks are scaled Givens rotations with $c_i = a_i/\sigma_i$ and $s_i = b_i/\sigma_i$, and thus $D_i = G_i \, \Sigma_i$ with $\Sigma_i = \text{diag}(\sigma_i, \sigma_i)$.

Now collect all g_i and G_i in the orthogonal block diagonal matrix G, and all scaling factors σ_i and Σ_i in the diagonal matrix Σ. Then an SVD of A is given by $A = U \Sigma V^T$ with $U = V \, G$, and it follows that $C = V^T U = G$. Therefore, for a normal matrix A, the mixing of the SVD components is limited to a mixing of subspaces of dimension two, and hence the Krylov subspaces for GMRES and RRGMRES are likely to be well behaved.

6 Conclusion

MINRES and MR-II have regularization properties for the same reason as CGLS does: by "killing" the large SVD components of the residual – in order to reduce its norm as much as possible – they capture the desired SVD components and produce a regularized solution. Negative eigenvalues do not inhibit the regularizing effect of MINRES and MR-II, but they influence the convergence rate.

GMRES and RRGMRES mix the SVD components in each iteration and thus do not provide a filtered SVD solution. For some problems GMRES and RRGMRES produce regularized solutions, either because the mixing is weak (see §5.4) or because the Krylov vectors are well suited for the problem (see §5.5). For other problems neither GMRES nor RRGMRES produce regularized solutions, either due to an unfavorable null space (see §5.2) or due to a severe and undesired mixing of the SVD components (see §5.6).

Our bottom-line conclusion is that while CGLS, MINRES and MR-II have general regularization properties, one should be very careful using GMRES and RRGMRES as general-purpose regularization methods for practical problems.

REFERENCES

1. P. N. Brown and H. F. Walker, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.
2. D. Calvetti, G. Landi, L. Reichel, and F. Sgallari, *Non-negativity and iterative methods for ill-posed problems*, Inverse Problems, 20 (2004), pp. 1747–1758.
3. D. Calvetti, B. Lewis, and L. Reichel, *GMRES-type methods for inconsistent systems*, Lin. Alg. Appl., 316 (2000), pp. 157–169.
4. D. Calvetti, B. Lewis, and L. Reichel, *GMRES, L-curves, and discrete ill-posed problems*, BIT, 42 (2002), pp. 44–65.

5. D. Calvetti, B. Lewis, and L. Reichel, *On the regularizing properties of the GMRES method*, Numer. Math., 91 (2002), pp. 605–625.
6. B. Fischer, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, Wiley Teubner, Stuttgart, 1996.
7. B. Fischer, M. Hanke, and M. Hochbruck, *A note on conjugate-gradient type methods for indefinite and/or inconsistent linear systems*, Numer. Algo., 11 (1996) pp. 181–187.
8. M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman Scientific & Technical, Essex, 1995.
9. M. Hanke, *On Lanczos based methods for the regularization of discrete ill-posed problems*, BIT, 41 (2001), pp. 1008–1018.
10. M. Hanke and J. G. Nagy, *Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques*, Inverse Problems, 12 (1996), pp. 157–173.
11. P. C. Hansen, *Regularization Tools. A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algo., 6 (1004), pp. 1–35.
12. P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images – Matrices, Spectra and Filtering*, SIAM, Philadephia, 2006 (to appear).
13. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
14. M. E. Kilmer, *On the regularizing properties of Krylov subspace methods*, unpublished; results presented at BIT 40th Anniversary meeting, Lund, Sweden, 2000.
15. M. E. Kilmer and G. W. Stewart, *Iterative regularization and MINRES*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 613–628.
16. C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Num. Anal., 12 (1975), pp. 617-629.
17. Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.