ITERATIVE REGULARIZATION OF MINIMUM-RESIDUAL METHODS FOR LARGE-SCALE SYMMETRIC DISCRETE ILL-POSED PROBLEMS

YI HUANG† AND ZHONGXIAO JIA‡

Abstract. For large-scale symmetric discrete ill-posed problems, MR-II, a minimal residual method, is a competitive alternative to LSQR and CGLS. In this paper, we establish bounds for the distance between an underlying \(k\)-dimensional Krylov subspace and the subspace spanned by the \(k\) dominant eigenvectors. They show that the \(k\)-step MR-II captures the \(k\) dominant spectral components better for severely and moderately ill-posed problems than for mildly ill-posed problems, so that MR-II has better regularizing effects for the first two kinds of problems than for the third kind. By the bounds, we derive an estimate for the accuracy of the rank \(k\) approximation generated by the symmetric Lanczos process. We analyze the regularization of MINRES and compare it with MR-II, showing why it is generally not enough to compute a best possible regularized solution and when it is better than MR-II. Our general conclusions are that MINRES and MR-II have only the partial regularization for general symmetric ill-posed problems and mildly ill-posed problems, respectively. Numerical experiments confirm our assertions. Furthermore, they illustrate that MR-II has the full regularization for severely and moderately ill-posed problems and MINRES has only the partial regularization independent of the degree of ill-posedness. The experiments also indicate that MR-II is as equally effective as and more efficient than LSQR for symmetric ill-posed problems.

Key words. Symmetric ill-posed problem, regularization, severely, moderately, mildly, MR-II, MINRES, CGLS, LSQR, hybrid

AMS subject classifications. 15A18, 65F22, 65J20

1. Introduction. We consider the iterative solution of the large-scale symmetric discrete ill-posed problem

\[(1.1) \quad Ax = b, \quad A \in \mathbb{R}^{n \times n}, \quad b \in \mathbb{R}^n,\]

where \(A\) is symmetric and extremely ill conditioned with its singular values decaying gradually to zero without a noticeable gap. This kind of problem arises in many science and engineering areas, such as signal processing and image deblurring [14]. In particular, the right-hand side \(b\) is affected by noise, caused by measurement errors, modeling errors or discretization errors, i.e.,

\[b = \hat{b} + e,\]

where \(e \in \mathbb{R}^n\) represents a white noise vector and \(\hat{b} \in \mathbb{R}^n\) denotes the noise-free right-hand side, and it is supposed that \(\|e\| < \|b\|\). Because of the presence of noise \(e\) in \(b\) and the high ill-conditioning of \(A\), the naive solution \(x_{\text{naive}} = A^{-1}b\) of (1.1) is meaningless and far from the true solution \(x_{\text{true}} = A^{-1}\hat{b}\). Therefore, one needs to use regularization methods to determine a regularized solution so that it is close to \(x_{\text{true}} = A^{-1}\hat{b}\) as much as possible [12, 14].

(1.1) of small and moderate size for a general \(A\) can be efficiently solved by the singular value decomposition (SVD) of \(A\). For \(A\) symmetric, its SVD is related to its
spectral decomposition as follows:

\[
A = V \Lambda V^T = V \Omega \Sigma V^T = U \Sigma V^T,
\]

where \( U = (u_1, u_2, \ldots, u_n) = V \Omega \) and \( V = (v_1, v_2, \ldots, v_n) \) are orthogonal, whose columns are the left and right singular vectors of \( A \), respectively, the diagonal matrix \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \) with the singular values labeled as \( \sigma_1 > \sigma_2 > \cdots > \sigma_n > 0 \), \( \Omega = \text{diag}(\pm 1) \) is a signature matrix such that \( \sigma_i = |\lambda_i| \) with the \( \lambda_i \) the eigenvalues of \( A \), and \( A = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \). Obviously, \( u_i = \pm v_i \) with the \( \pm \) sign depending on \( \Omega \). With (1.2), we can express the naive solution of (1.1) as

\[
x_{\text{naive}} = \sum_{i=1}^{n} \frac{v_i^T b}{\lambda_i} v_i = \sum_{i=1}^{n} \frac{v_i^T \hat{b}}{\lambda_i} v_i + \sum_{i=1}^{n} \frac{u_i^T e}{\lambda_i} v_i = x_{\text{true}} + \sum_{i=1}^{n} \frac{u_i^T e}{\lambda_i} v_i.
\]

Throughout the paper, we assume that \( \hat{b} \) satisfies the discrete Picard condition: On average, the coefficients \( |u_i^T b| = |v_i^T \hat{b}| \) decay faster than the singular values \( \sigma_i \), i.e., \( |\lambda_i| \). This is a necessary hypothesis for the stability of regularized solutions of (1.1); see [12]. To be definitive, for the sake of simplicity we assume that these coefficients satisfy a widely used model [12, 14]:

\[
|v_i^T \hat{b}| = |\lambda_i|^{1+\beta}, \quad \beta > 0, \quad i = 1, 2, \ldots, n.
\]

The assumption of the white noise means that all the \( |v_i^T e| \) are nearly equal. Let \( k_0 \) denote the transition point for which \( |v_{k_0+1}^T \hat{b}| \approx |v_{k_0+1}^T e| \), which classifies the eigenvalues \( \lambda_i \) into the large ones for \( i \leq k_0 \) and the small ones for \( i > k_0 \). Here and hereafter, whether an eigenvalue is large or small is characterized by its absolute value. Therefore, similar to the truncated SVD (TSVD) method, for \( A \) symmetric we can use a truncated spectral decomposition method to capture the \( k_0 \) dominant spectral components and compute a best possible regularized solution [12, 14]:

\[
x_{k_0} = \sum_{i=1}^{k_0} \frac{v_i^T b}{\lambda_i} v_i = A_{k_0}^1 b,
\]

where \( A_{k_0} = U_{k_0} \Sigma_{k_0} \Sigma_{k_0} V_{k_0}^T \) with \( U_{k_0} \) and \( V_{k_0} \) the first \( k_0 \) columns of \( U \) and \( V \), respectively, \( \Sigma_{k_0} = \text{diag}(\sigma_1, \ldots, \sigma_{k_0}) \) and \( \dagger \) the Moore-Penrose generalized inverse.

The other most famous direct regularization is Tikhonov regularization, which takes its simplest form

\[
\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \mu^2 \|x\|^2,
\]

where the norm used is the 2-norm of a matrix or vector and \( \mu \geq 0 \) is referred as the regularization parameter. Both the Tikhonov regularization and the truncated spectral decomposition method can be regarded as parameter-filtered methods, whose solutions are of the form

\[
x_{\text{filt}} = \sum_{i=1}^{n} f_i \frac{v_i^T b}{\lambda_i} v_i = \sum_{i=1}^{n} f_i \frac{u_i^T b}{\sigma_i} v_i,
\]

where the filter factors \( f_i = \frac{\lambda_i^2}{\lambda_i^2 + \mu^2} \) for the Tikhonov regularization, and \( f_i = 1, \quad i = 1, 2, \ldots, k_0 \) and \( f_i = 0, \quad i = k_0 + 1, \ldots, n \) for the the truncated spectral decomposition.
method. The regularized solution of form (1.6) is said to be a filtered SVD solution [12, 14]. An appropriate choice of $\mu$ must be such that $f_j \approx 1$ for $k_0$ large eigenvalues and $f_j \approx 0$ for $n-k_0$ small eigenvalues by avoiding the noise deteriorating the solution [12, 14]. Many techniques have been developed for finding an optimal $\mu$, such as discrepancy principle, the L-curve criterion and generalized cross validation; see, e.g., [1, 12, 19, 26] for relevant comparisons and discussions of the classical and new ones.

However, it is impractical to compute the spectral decomposition of $A$ when (1.1) is large. In this case, one typically solves it iteratively via some Krylov subspace methods. For (1.1) with a general matrix $A$, the mathematically equivalent CGLS [4] and LSQR [24] have been shown to have regularizing effects [3, 10, 17, 22]; see also [12, 14] for a systematic account. They exhibit the semi-convergence: The iterates improve and the residual norms decrease from the beginning to some iteration, then the noise starts to deteriorate the solutions dramatically and their norms become large but the residual norms stabilize. The semi-convergence is due to the fact that the projected problem at some iteration starts to inherit the ill-conditioning of (1.1). That is, after some iteration, the appearance of a small singular value of the projected problem will amplify the noise considerably [12, 14].

For a given (1.1), a big problem is whether or not a purely iterative solver has already obtained a best possible regularized solution at the occurrence of semi-convergence. For Krylov subspace based iterative solvers, their regularizing effects critically rely on how well the underlying $k$-dimensional Krylov subspace captures the $k$ dominant right singular vectors of $A$ [12, 14]. The richer information the Krylov subspace contains on the $k$ dominant right singular vectors, the less possibly a small Ritz value appears and thus the better regularization the solver has. To precisely describe the regularizing effects, we introduce the term of full or partial regularization. If a purely iterative solver itself computes a best possible regularized solution at the occurrence of semi-convergence, it is said to have the full regularization; in this case, no additional regularization is needed. Otherwise, it is said to have the partial regularization; in this case, in order to compute a best possible regularized solution, its hybrid variant is needed that combines the solver with additional regularization. In order to analyze the regularizing effects of a Krylov subspace solver, we need the following definition of the degree of ill-posedness, which follows Hofmann’s book [16] and has been commonly used in the literature, e.g., [12, 14]: If there exists a positive real number $\alpha$ such that the singular values satisfy $\sigma_j = O(j^{-\alpha})$, then the problem is termed as mildly or moderately ill-posed if $\alpha \leq 1$ or $\alpha > 1$; if $\sigma_j = O(e^{-\alpha j})$ with $\alpha > 0$ considerably, $j = 1, 2, \ldots, n$, then the problem is termed severely ill-posed. More generally, the definition of severe ill-posedness can be extended to the problem with the singular values $\sigma_j = O(\rho^{-j})$ with $\rho > 1$ considerably.

Very recently, for LSQR and CGLS, Huang and Jia [17] have established quantitative bounds for the $F$-norm distance between the underlying $k$-dimensional Krylov subspace and the $k$-dimensional dominant right singular subspace. The results indicate that the $k$-dimensional Krylov subspace better captures the $k$ dominant right singular vectors for severely and moderately ill-posed problems than for mildly ill-posed problems. Consequently, LSQR and CGLS have better regularizing effects for the first two kinds of problems than for the third kind of problem. Precisely, the authors have shown that LSQR and CGLS generally have only the partial regularization for mildly ill-posed problems and additional regularization is needed to compute best possible regularized solutions.

For $A$ symmetric, the $k$-dimensional dominant eigenspace is identical to the $k$-
dimensional dominant left and right singular subspace. It is naturally desirable not to use CGLS and LSQR since they treat (1.1) as a general one and require to compute two multiplications with $A$ at each iteration. In this case, MINRES [25] and its variant MR-II [9, 11] are proper alternatives. MR-II was originally designed for solving singular and inconsistent systems, and it uses the starting vector $Ab$ and restricts the resulting Krylov subspace to the range of $A$. Thus, the iterates are orthogonal to the null space of $A$, which leads to the effects that MR-II computes the minimum-norm least squares solution [6, 9]. For (1.1), we are not interested in the minimum-norm least squares solution but a good regularized solution that approximates $x_{true}$. MINRES and MR-II have been shown to have regularizing effects and exhibit the semi-convergence [15, 18, 20], and MR-II generally provides better regularized solutions than MINRES, intuitively because the noise $e$ in the initial Krylov vector $Ab$ is filtered by multiplication with $A$ [10, 18]. Different algorithmic implementations associated with MR-II have been studied [9, 21]. Compared with CGLS and LSQR, each iteration of MINRES and MR-II needs only one multiplication with $A$. Thus, once they get best possible regularized solutions with roughly the same iterations as LSQR and CGLS, MINRES and MR-II are preferable since they halve the computational cost.

In this paper, we analyze the regularizing effects of MR-II, in a manner different from those used in [18, 20], and draw some definitive conclusions. We establish quantitative bounds for the $F$-norm distance between the underlying $k$-dimensional Krylov subspace and the $k$-dimensional dominant eigenspace. The bounds indicate that the Krylov subspace better captures the $k$ dominant eigenvectors for severely and moderately ill-posed problems than for mildly ill-posed problems. As a consequence, MR-II has better regularization for the first two kinds of problems than for the third kind. Furthermore, by the bounds and the analysis, we draw a general conclusion that MR-II generally has only the partial regularization for mildly ill-posed problems. We then use the results to estimate the accuracy of the rank $k$ approximation, generated by the symmetric Lanczos process, to $A$, a core problem on the regularizing effects of MR-II. For MINRES, we prove that it has a filtered SVD solution, by which we show why it, in general, has only the partial regularization, independent of the degree of ill-posedness of (1.1). As a result, a hybrid MINRES should be used. Furthermore, we take a closer look at the regularization of MINRES and MR-II in more detail, which, from a new perspective, shows that MINRES has only the partial regularization, and we prove that the $k$-th iterate by MINRES is always more accurate than the $(k-1)$-th iterate by MR-II until the occurrence of semi-convergence of MINRES.

This paper is organized as follows. In Section 2, we describe MINRES and prove that its iterates are filtered SVD solutions, followed by an analysis on its regularizing effects. In Section 3, we overview MR-II and present our theoretical results. In Section 4, we take a closer look at the regularization of MINRES and MR-II in more detail, which, from a new perspective, shows that MINRES has only the partial regularization, and we prove that the $k$-th iterate by MINRES is always more accurate than the $(k-1)$-th iterate by MR-II until the occurrence of semi-convergence of MINRES.

Throughout the paper, we denote by $K_k(C, w) = \text{span}\{w, Cw, \ldots, C^{k-1}w\}$ the $k$-dimensional Krylov subspace generated by the matrix $C$ and the vector $w$, by $\| \cdot \|$ and $\| \cdot \|_F$ the 2-norm of a matrix or vector and the Frobenius norm of a matrix, respectively, and by $I$ the identity matrix with order clear from the context.
2. Regularizing effects of MINRES. MINRES [25] is based on the symmetric Lanczos process that constructs an orthonormal basis of the Krylov subspace $K_k(A, b)$. Let $\bar{q}_1 = b/\|b\|$. The Lanczos process can be written in the matrix form

$$A\bar{Q}_k = \bar{Q}_{k+1}\bar{T}_k,$$

where $\bar{Q}_{k+1} = [\bar{q}_1, \bar{q}_2, \ldots, \bar{q}_{k+1}]$ has orthonormal columns, and $\bar{T}_k \in \mathbb{R}^{(k+1)\times k}$ is a tridiagonal matrix with its subdiagonal entries positive and leading $k \times k$ submatrix symmetric.

At iteration $k$, MINRES solves $\|b - A\bar{x}^{(k)}\| = \min_{x \in K_k(A, b)} \|b - Ax\|$ for the iterate $\bar{x}^{(k)} = \bar{Q}_k\bar{y}^{(k)}$ with

$$\bar{y}^{(k)} = \arg \min_{y \in \mathbb{R}^k} \|b\| e_1 - \bar{T}_k y,$$

where $e_1$ is the first canonical vector of dimension $k + 1$.

Similar to the CGLS iterates [12, p. 146], we can establish the following result.

**Theorem 2.1.** For MINRES to solve (1.1) with the starting vector $\bar{q}_1 = b/\|b\|$, the $k$-th iterate $\bar{x}^{(k)}$ has the form

$$\bar{x}^{(k)} = \sum_{i=1}^n f_i^{(k)} v_i^T b \lambda_i v_i,$$

where $f_i^{(k)} = 1 - \prod_{j=1}^k \frac{g^{(k)}_j - \lambda_i}{g^{(k)}_j}$, $i = 1, 2, \ldots, n$, and $g^{(k)}_j$, $j = 1, 2, \ldots, k$ are the harmonic Ritz values of $A$ with respect to $K_k(A, b)$.

**Proof.** From [23], the residual $\bar{r}^{(k)} = b - A\bar{x}^{(k)}$ of the MINRES iterate $\bar{x}^{(k)}$ can be written as

$$\bar{r}^{(k)} = \chi_k(A)b,$$

where the residual polynomial $\chi_k(t)$ has the form

$$\chi_k(t) = \prod_{j=1}^k \frac{g^{(k)}_j - t}{g^{(k)}_j},$$

with $g^{(k)}_j$ the harmonic Ritz values of $A$ with respect to $K_k(A, b)$. From (2.3), we get

$$\bar{x}^{(k)} = (I - \chi_k(A))A^{-1}b.$$

Substituting $A = VAV^T$ into the above gives

$$\bar{x}^{(k)} = \sum_{i=1}^n f_i^{(k)} v_i^T b \lambda_i v_i,$$

where

$$f_i^{(k)} = 1 - \prod_{j=1}^k \frac{g^{(k)}_j - \lambda_i}{g^{(k)}_j}, \quad i = 1, 2, \ldots, n.$$  

(2.2) shows that the MINRES iterate $\bar{x}^{(k)}$ is a filtered SVD solution. For a general symmetric $A$, the harmonic Ritz values have an attractive feature: they usually
favor extreme, i.e., algebraically large and small, eigenvalues of $A$, provided that a Krylov subspace contains substantial information on all the eigenvectors $v_i$ [23]. In our current context, if at least a small harmonic Ritz value starts to appear for some $k \leq k_0$, i.e., $|\theta^{(k)}_k| \leq |\lambda_{k_0+1}|$, the corresponding filter factors $f_i^{(k)}$, $i = k + 1, \ldots, n$ are not small, meaning that $\hat{x}^{(k)}$ is already deteriorated and becomes a poorer regularized solution. If no small harmonic Ritz value appears before $k \leq k_0$, the $\hat{x}^{(k)}$ are expected to become better approximations to $x_{\text{true}}$ as $k$ increases. Unfortunately, since $K_k(A, b)$ includes the noisy $b = \hat{b} + e$, which contains substantial components $v_i$ corresponding to all the eigenvalues $\lambda_i$, a small harmonic Ritz value generally appears for $k \leq k_0$. This demonstrates that, in general, MINRES only has the partial regularization and is not enough to get a best possible regularized solution. We will come back to the regularization of MINRES in Section 4.

3. Regularizing effects of MR-II. MR-II [9] is a variant of MINRES applied to $K_k(A, Ab)$. The method is based on the symmetric Lanczos process

$$(3.1) \quad AQ_k = Q_{k+1}T_k,$$

where $Q_{k+1} = (q_1, q_2, \ldots, q_{k+1})$ has orthonormal columns with $q_1 = Ab/\|Ab\|$, and $T_k \in \mathbb{R}^{(k+1) \times k}$ is a tridiagonal matrix with the diagonals $\alpha$ and subdiagonals $\beta_i$, $i = 1, 2, \ldots, k$ and the leading $k \times k$ submatrix symmetric. The columns of $Q_k$ form an orthonormal basis of $K_k(A, Ab)$. At iteration $k$, MR-II computes the iterate $x^{(k)} = Q_ky^{(k)}$ to minimize the 2-norm of the residual over $K_k(A, Ab)$, which is obtained by solving the projected problem

$$(3.2) \quad y^{(k)} = \arg\min_{y \in \mathbb{R}^n} \|b - Q_{k+1}T_ky\|.$$

Next we establish a number of results that help better understand the regularization of MR-II. We first focus on what information $K_k(A, Ab)$ contains and provides. In terms of the definition of canonical angles $\Theta(\mathcal{X}, \mathcal{Y})$ between the two subspaces $\mathcal{X}$ and $\mathcal{Y}$ of the same dimension [27, p. 250], we present the following result.

**Theorem 3.1.** Let $A = V \Sigma V^T = V \Lambda V^T$ be defined as in (1.2), and assume that the singular values of $A$ are of the form $\sigma_j = |\lambda_j| = \mathcal{O}(e^{-\alpha j})$ with $\alpha > 0$ considerably. Let $V_k = \text{span}\{v_k\}$ be the subspace spanned by the columns of $V_k = (v_1, v_2, \ldots, v_k)$, and $V_k^* = K_k(A, Ab)$. Then

$$(3.3) \quad \|\sin \Theta(V_k, V_k^*)\|_F \leq \frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{|v_k^T b|}{|v_k^T b|} \sqrt{k(k - 1)\mathcal{O}(1)}, \quad k = 1, 2, \ldots, n - 1.$$ 

**Proof.** Note that $K_k(\Lambda, \Lambda V^T b)$ is spanned by the columns of the $n \times k$ matrix $DB_k$ with

$$D = \text{diag} (\lambda_i V^T b), \quad B_k = \begin{pmatrix} 1 & \lambda_1 & \ldots & \lambda_{k-1}^k \\ 1 & \lambda_2 & \ldots & \lambda_{k-1}^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \ldots & \lambda_{k-1}^{k-1} \end{pmatrix}.$$ 

Partition the matrices $D$ and $B_k$ as follows:

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}, \quad B_k = \begin{pmatrix} B_{k1} \\ B_{k2} \end{pmatrix},$$
where $D_1, B_{k1} \in \mathbb{R}^{k \times k}$. Since $B_{k1}$ is a Vandermonde matrix with $\lambda_j$ distinct for $1 \leq j \leq k$, it is nonsingular. Since $K_k(A, Ab) = VK_k(AV^Tb)$, we have

$$K_k(A, Ab) = \text{span}\{VDT_k\} = \text{span}\left\{V \begin{pmatrix} D_1B_{k1} \\ D_2B_{k2} \end{pmatrix} \right\} = \text{span}\left\{V \begin{pmatrix} I \\ \Delta_k \end{pmatrix} \right\}$$

with $\Delta_k = D_2B_{k2}B_{k1}^{-1}D_1^{-1}$. Define $Z_k = V \begin{pmatrix} I \\ \Delta_k \end{pmatrix}$. Then $Z_k^T Z_k = I + \Delta_k^T \Delta_k$, and the columns of $Z_k(Z_k^T Z_k)^{-\frac{1}{2}}$ form an orthonormal basis of $V_k^\perp$.

Write $V = (V_k, V_k^\perp)$. By definition, we obtain

$$\|\sin \Theta(V_k, V_k^\perp)\|_F = \left\| (V_k^\perp)^T Z_k Z_k^T Z_k^{-\frac{1}{2}} \right\|_F$$

$$= \left\| (V_k^\perp)^T V \begin{pmatrix} I \\ \Delta_k \end{pmatrix} (I + \Delta_k^T \Delta_k)^{-\frac{1}{2}} \right\|_F$$

$$= \left\| \Delta_k (I + \Delta_k^T \Delta_k)^{-\frac{1}{2}} \right\|_F$$

$$\leq \|\Delta_k\|_F \left\| (I + \Delta_k^T \Delta_k)^{-\frac{1}{2}} \right\|$$

$$\leq \|\Delta_k\|_F = \left\| D_2B_{k2}B_{k1}^{-1}D_1^{-1} \right\|_F.$$

We now estimate $\left\| B_{k2}B_{k1}^{-1} \right\|_F$. It is easily justified that the $i$-th column of $B_{k1}^{-1}$ consists of the coefficients of the Lagrange polynomial

$$L_i^{(k)}(\lambda) = \prod_{j=1, j \neq i}^k \frac{\lambda_j - \lambda}{\lambda_j - \lambda_i}$$

that interpolates the elements of the $i$-th canonical basis vector $e_i^{(k)} \in \mathbb{R}^k$ at the abscissas $\lambda_1, \ldots, \lambda_k$. Consequently, the $i$-th column of $B_{k2}B_{k1}^{-1}$ is

$$B_{k2}B_{k1}^{-1} e_i^{(k)} = \left( L_i^{(k)}(\lambda_{k+1}), \ldots, L_i^{(k)}(\lambda_n) \right)^T,$$

from which we obtain

$$B_{k2}B_{k1}^{-1} = \begin{pmatrix} L_1^{(k)}(\lambda_{k+1}) & L_2^{(k)}(\lambda_{k+1}) & \ldots & L_k^{(k)}(\lambda_{k+1}) \\ L_1^{(k)}(\lambda_{k+2}) & L_2^{(k)}(\lambda_{k+2}) & \ldots & L_k^{(k)}(\lambda_{k+2}) \\ \vdots & \vdots & \ddots & \vdots \\ L_1^{(k)}(\lambda_n) & L_2^{(k)}(\lambda_n) & \ldots & L_k^{(k)}(\lambda_n) \end{pmatrix}.$$

For a fixed $\lambda$ satisfying $|\lambda| \leq |\lambda_{k+1}|$, let $i_0 = \arg \max_{i=1,2,\ldots,k} |L_i^{(k)}(\lambda)|$. Then we have

$$|L_{i_0}^{(k)}(\lambda)| = \prod_{j=1, j \neq i_0}^k \left| \frac{\lambda_j - \lambda}{\lambda_j - \lambda_{i_0}} \right| \leq \prod_{j=1, j \neq i_0}^k \left| \frac{|\lambda_j - \lambda|}{|\lambda_j - \lambda_{i_0}|} \right| \leq \prod_{j=1, j \neq i_0}^k \left| \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_j| - |\lambda_{i_0}|} \right|. $$
Therefore, for \( i = 1, 2, \ldots, k \) and \( \alpha > 0 \) considerably we get:

\[
|L^{(k)}_i(\lambda)| = \prod_{j=1, j \neq i_0}^{i_0 - 1} \left| \frac{\lambda_j + |\lambda_{k+1}|}{\lambda_j - |\lambda_{i_0}|} \right| = \prod_{j=1}^{i_0 - 1} \frac{\lambda_j + |\lambda_{k+1}|}{\lambda_j - |\lambda_{i_0}|} \cdot \prod_{j=i_0 + 1}^{k} \frac{\lambda_j + |\lambda_{k+1}|}{\lambda_j - |\lambda_{i_0}|} \cdot \prod_{j=i_0 + 1}^{k} \frac{\lambda_j}{\lambda_j - |\lambda_j|} \\
= \prod_{j=1}^{i_0 - 1} \frac{1 + O\left((e^{-(k-j)+1})^{\alpha}\right)}{1 - O\left((e^{-(i_0-j)})^{\alpha}\right)} \cdot \prod_{j=i_0 + 1}^{k} \frac{1}{1 - O\left((e^{-(j-i_0)})^{\alpha}\right)} \cdot \prod_{j=i_0 + 1}^{k} \frac{1}{O\left((e^{-(j-i_0)})^{\alpha}\right)} \\
= \prod_{j=1}^{i_0 - 1} \frac{1 + O\left((e^{-(k-j)+1})^{\alpha}\right)}{1 - O\left((e^{-(i_0-j)})^{\alpha}\right)} \cdot \prod_{j=i_0 + 1}^{k} \frac{1}{1 - O\left((e^{-(j-i_0)})^{\alpha}\right)} \cdot \prod_{j=i_0 + 1}^{k} \frac{1}{O\left((e^{-(j-i_0)})^{\alpha}\right)} \\
= \frac{1 + \sum_{j=1}^{k} O\left((e^{-(k-j)+1})^{\alpha}\right)}{1 + O\left((e^{-(k-i_0)+1})^{\alpha}\right)} \cdot \frac{1 + \sum_{j=1}^{i_0 - 1} O\left((e^{-(j-i_0)})^{\alpha}\right)}{1 + \sum_{j=1}^{k-i_0} O\left((e^{-(j-i_0)})^{\alpha}\right)} \cdot \frac{1}{1 + \sum_{j=1}^{k-i_0} O\left((e^{-(j-i_0)})^{\alpha}\right)} \\
= O(1),
\]

when \( i_0 = k \) or near to it by noticing that the last second quantity becomes smaller for the other \( i_0 \) away from \( k \), from which and (3.4) it follows that

\[
\|B_{k_2}B^{-1}_{k_1}\|_F = \sqrt{k(n-k)}O(1).
\]

As a result, for \( i = 1, 2, \ldots, n - 1 \) we have

\[
\|\sin \Theta(V_k, V_k^*)\|_F \leq \|D_2B_{k_2}B^{-1}_{k_1}D_1^{-1}\|_F \\
\leq D_2 \|B_{k_2}B^{-1}_{k_1}\|_F \|D_1^{-1}\| \\
\leq \frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{|v_{k+1}^T b|}{|v_k^T b|} \|B_{k_2}B^{-1}_{k_1}\|_F \\
= \frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{|v_{k+1}^T b|}{|v_k^T b|} \sqrt{k(n-k)}O(1). \tag{\*}
\]

**Remark 3.1** The theorem holds for a general severely ill-posed problem since for \( |\lambda| \leq |\lambda_{k+1}| \) we can similarly justify \( \max_{i=1, 2, \ldots, k} |L^{(k)}_i(\lambda)| = O(1) \) for exponentially decaying singular values \( \sigma_j = O(\rho^{-j}) \) with \( \rho > 1 \) considerably. Furthermore, the theorem can be extended to moderately ill-posed problems with the singular values \( \sigma_j = O(j^{-\alpha}) \), \( \alpha > 1 \) considerably and \( k \) not big. This is because for \( i = 1, 2, \ldots, k \)
and $|\lambda| \leq |\lambda_{k+1}|$ we can justify, in a similar way to the proof of Theorem 3.1, that

$$|L_i^{(k)}(\lambda)| \leq \prod_{j=1, j \neq i}^{k} \left| \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_j| - |\lambda_{k+1}|} \right|$$

$$= \prod_{j=1}^{i-1} \frac{1 + O\left(\left(\frac{1}{i-k}\right)^\alpha\right)}{1 - O\left(\left(\frac{1}{i-k}\right)^\alpha\right)} \cdot \prod_{j=i+1}^{k} O\left(\left(\frac{1}{i-k}\right)^\alpha\right) + 1 \cdot O\left(\left(\frac{1}{i-k}\right)^\alpha\right) - 1 = O(1).$$

Unfortunately, for mildly ill-posed problems we have $\max_{i=1,2,\ldots,k} |L_i^{(k)}(\lambda)| > 1$ considerably for $|\lambda| \leq |\lambda_{k+1}|$ and $\alpha < 1$.

The theorem and the above analysis mean that $V_k^s$ captures $V_k$ considerably better for severely and moderately ill-posed problems than for mildly ill-posed problems. In other words, our results imply that $V_k^s$ contains substantial information on the other $n-k$ eigenvectors for mildly ill-posed problems, causing that a small Ritz value appear very possibly for some $k \leq k_0$, especially when $k_0$ is not small. As a result, MR-II has better regularization for severely and moderately ill-posed problems than for mildly ill-posed problems. Since MR-II has at most the full regularization for severely and moderately ill-posed problems, MR-II generally has only the partial regularization for mildly ill-posed problem.

In comparison of the results, i.e., Theorem 3.1, in [17] on LSQR and CGLS, we find that $K_k(A,Ab)$ can be as comparably effective as $K_k(A^T,Ab)$, on which LSQR and CGLS work, for capturing the $k$-dimensional dominant eigenspace.

**Remark 3.2** (3.3) and the above analysis indicate that $V_k^s$ captures $V_k$ better for severely ill-posed problems than for moderately ill-posed problems. There are two reasons for this. The first is that the factors $|\lambda_{k+1}|/|\lambda_k|$ are basically fixed constants for severely ill-posed problems as $k$ increases, and they are smaller than the counterparts for moderately ill-posed problems unless the degree $\alpha$ of ill-posedness is far bigger than one and $k$ small. The second is that the $O(1)$ are smaller for severely ill-posed problems than for moderately ill-posed problems.

Recall the discrete Picard condition (1.4), and consider the coefficients

$$c_k = \frac{|v_{k+1}^T b|}{|v_k^T b|} = \frac{\gamma_{k+1} b + v_{k+1}^T e}{\tilde{v}_k b + v_k^T e} \approx \frac{|\lambda_{k+1}|^{1+\beta} + |v_{k+1}^T e|}{|\lambda_k|^{1+\beta} + |v_k^T e|}.$$ 

We see that, the larger $\beta$ is, the smaller $c_k \approx \frac{|\lambda_{k+1}|^{1+\beta}}{|\lambda_k|^{1+\beta}} < 1$ for $k \leq k_0$, and thus the better $V_k^s$ may not capture $V_k$ so well after iteration $k_0$.

**Remark 3.3** In the proof, for $|\lambda| \leq |\lambda_{k+1}|$ we have replaced $|\lambda_j - \lambda|$ by possibly bigger $|\lambda_j| + |\lambda_{k+1}|$. However, when $A$ is positive (or negative) definite, i.e., $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$, $|\lambda_j - \lambda|$ can be simply replaced by smaller $\lambda_j$. This indicates that $K_k(A,Ab)$ may capture the dominant eigenspace better for $A$ positive (negative) definite than for $A$ having both positive and negative eigenvalues.

**Remark 3.4** (3.3) should not be sharp. As we have seen from the proof, the factor $\sqrt{n-k}$ in it seems unavoidable. However, we conjecture that the factor $\sqrt{n-k}$ should be replaced by a much smaller factor $O(1)$. Unfortunately, we are currently unable to remove $\sqrt{n-k}$.

Let us get more insight into the regularization of MR-II. Recall (3.1), and define

$$\gamma_k = \|A - Q_{k+1}T_kQ_k^T\|,$$
which measures the quality of the rank $k$ approximation $Q_{k+1}T_kQ_k^T$ to $A$. Based on (3.3), we can derive the following estimate for $\gamma_k$.

**Theorem 3.2.** Assume that (1.1) is severely or moderately ill posed. Then

\[(3.5) \quad |\lambda_{k+1}| \leq \gamma_k \leq (1 + \eta_k)|\lambda_{k+1}|,\]

where $\eta_k$ is positive. Note that $Q_{k+1}T_kQ_k^T$ is of rank $k$. The lower bound in (3.5) is trivial since the error norm of the best $k$ approximation to $A$ with respect to the 2-norm is $\sigma_{k+1} = |\lambda_{k+1}|$. We next prove the upper bound. From (3.1), we obtain

\[(3.6) \quad \|A - AQ_kQ_k^T\| = \|A - AQ_kQ_k^T\| = \|A(I - Q_kQ_k^T)\|.

From Theorem 3.1, it is known that $V_k = K_k(A, Ab) = \text{span}\{Q_k\}$. Let $V_k = (v_1, v_2, \ldots, v_k)$ and $\Lambda_k = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$. Then we obtain

\[
\|A - AQ_kQ_k^T\| = \|(A - V_k\Lambda_kV_k^T + V_k\Lambda_kV_k^T)(I - Q_kQ_k^T)\| \\
\leq \|(A - V_k\Lambda_kV_k^T)(I - Q_kQ_k^T)\| + \|V_k\Lambda_kV_k^T(I - Q_kQ_k^T)\| \\
\leq |\lambda_{k+1}| + \|\Lambda_k\| \|V_k^T(I - Q_kQ_k^T)\| \\
= |\lambda_{k+1}| + \|\lambda_1\| \sin \Theta(V_k, V_k^*) \\
\leq (1 + \eta_k)|\lambda_{k+1}|.\]

Our numerical experiments will indicate that $\gamma_k \approx \sigma_{k+1} = |\lambda_{k+1}|$ for severely and moderately ill-posed problems. This implies that $Q_{k+1}T_kQ_k^T$ is a very best rank $k$ approximation to $A$ with the approximate accuracy $\sigma_{k+1}$. Recall from (1.5) that the TSVD or truncated spectral decomposition method generates the best regularized solution $x_{k0} = A_{k0}^*b$. As a result, if $\gamma_{k0} \approx \sigma_{k0+1}$, the MR-II iterate $x^{(k0)} = Q_{k0}T_{k0}^*Q_{k0+1}^Tb$ is reasonably close to the TSVD solution $x_{k0}$ provided that $\sigma_{k0+1}$ is a reasonably small. This means that MR-II has the full regularization and does not need any additional regularization. These phenomena appear to be of generality and thus should have strong theoretical supports. Note that $\eta_k$ in (3.5) is generally considered larger than one. Compared with the observations, it appears that estimate (3.5) is pessimistic. Given the general observations that $\gamma_k \approx \sigma_{k+1}$ for severely and moderately ill-posed problems, we believe that our derivation must miss something essential that is unfortunately unknown to us now, except that we have replaced $\|\sin \Theta(V_k, V_k^*)\|$ by a bigger $\|\sin \Theta(V_k, V_k^*)\|F$, which may amplify $\eta_k$ roughly by a multiple $\sqrt{k}$ since $\|\sin \Theta(V_k, V_k^*)\|F \geq \frac{1}{\sqrt{k}} \|\sin \Theta(V_k, V_k^*)\|F$.

Recall that $\beta_i, i = 1, 2, \ldots, k$ denote the subdiagonals of $T_k$ defined by (3.1). The size of $\beta_k$ decides when to stop the Lanczos process, depending on the context. We establish an intimate and interesting relationship between $\beta_k$ and $\gamma_k$, showing how fast $\beta_k$ decays.

**Theorem 3.3.** For $k = 1, 2, \ldots, n - 1$ we have

\[(3.7) \quad \beta_k \leq \gamma_k.\]

**Proof.** Suppose the Lanczos process runs to completion. Then we have

\[A = Q_n\hat{T}_nQ_n^T,\]
where $Q_n \in \mathbb{R}^{n \times n}$ is orthogonal, and

$$
\hat{T}_n = \begin{pmatrix}
\alpha_1 & \beta_1 \\
\beta_1 & \alpha_2 & \beta_2 \\
& \ddots & \ddots \\
& & \ddots & \beta_{n-1} \\
& & & \beta_{n-1} & \alpha_n \\
\end{pmatrix}
$$

is symmetric tridiagonal. Thus, from (3.1) we have

$$
\gamma_k = \|A - Q_{k+1}T_kQ_k^T\| = \|Q_n^TAQ_n^{T}(A - Q_{n+1}^TQ_{n+1}T_kQ_k^T)Q_n\|
$$

$$
= \|\hat{T}_n - \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} T_k & I \\ I & 0 \end{pmatrix} \| = \|G_k\|,
$$

where

$$
G_k = \begin{pmatrix}
\beta_k & \alpha_{k+1} & \beta_{k+1} & \alpha_{k+2} & \beta_{k+2} & \ddots & \ddots \\
\alpha_{k+1} & \beta_{k+1} & \alpha_{k+2} & \beta_{k+2} & \ddots & \ddots \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots & \ddots & \beta_{n-1} \\
& & & \ddots & \beta_{n-1} & \alpha_n \\
\end{pmatrix}
\in \mathbb{R}^{(n-k+1) \times (n-k)},
$$

from which we get

$$
\beta_k = \|e^T_kG_k\| \leq \|G_k\| = \gamma_k.
$$

The theorem indicates that $\beta_k$ decays at least as fast as $\gamma_k$.

4. A closer look at the regularization of MINRES and MR-II. We have proved that MR-II have the better regularization than MINRES. The inferior regularization of MINRES was known a long time ago [9] and is simply due to the fact that $K_k(A, b)$ for MINRES includes the noisy $b$ and $K_k(A, Ab)$ for MR-II contains less information on $v_i$ corresponding to small eigenvalues since the noise $e$ in the starting vector $Ab$ is filtered by multiplication with $A$. In Section 2, we have given a formal analysis on the regularizing effects of MINRES. In what follows we shed more light on the regularization of MINRES and MR-II, and reveal some new features of them.

One might be confused that, since $K_{k-1}(A, Ab) \subset K_k(A, b)$, MINRES at iteration $k$ should provide an approximate solution that is at least as accurately as MR-II does at iteration $k - 1$. This is true for solving standard linear systems, but not the case for solving an ill-posed problem, for which we are concerned with regularized approximations to the true solution $x_{true}$ other than the naive solution $x_{naive}$. Our previous analysis has shown that a small harmonic Ritz value $|\theta_k^{(k)}| \leq \sigma_{k_0+1} = |\lambda_{k_0+1}|$ generally appears for MINRES before some iteration $k \leq k_0$. However, because $K_{k-1}(A, Ab) \subset K_k(A, b)$, the $k$-step MINRES must get a more accurate regularized solution than the $(k-1)$-step MR-II whenever $|\theta_k^{(k)}| > |\lambda_{k_0+1}|$, i.e., before the occurrence of semi-convergence of MINRES.

In view of the above, our general conclusion is that, to filter the effect of small harmonic Ritz values, a hybrid MINRES is generally needed that applies additional regularization to the projected problems after the iteration where the semi-convergence
occurs. Since $K_{k-1}(A, Ab) \subset K_k(A, b)$, in such a way, the $k$-step hybrid MINRES will get a regularized solution at least as accurately as the $(k-1)$-step MR-II does.

We can also explain the partial regularization of MINRES in terms of the rank $k$ approximation $Q_{k+1}T_kQ_k^T$ to $A$ as follows: Since the $k$-dimensional dominant eigenspace is identical to the $k$-dimensional left and right singular vectors, $K_k(A, b)$ contains substantial SVD components corresponding to all the singular values. As a result, it is very possible that the projected matrix $\bar{T}_k$ has a singular value smaller than $\sigma_{k_0+1}$ for some $k \leq k_0$. This means that $Q_{k+1}T_kQ_k^T$ is a poor rank $k$ approximation to $A$, causing, from (2.1), that $\|\bar{x}^{(k)}\| = \|Q_k\bar{y}^{(k)}\| = \|b\|\|\bar{T}_k^1e_1\|$ is generally large, i.e., $\bar{x}^{(k)}$ is already deteriorated. Conversely, if no singular value of $\bar{T}_k$ smaller than $\sigma_{k_0+1}$, the MINRES iterate $\bar{x}^{(k)}$ should be at least as accurate as the MR-II iterate $x^{(k-1)}$ because of $K_{k-1}(A, Ab) \subset K_k(A, b)$.

5. Numerical experiments. In this section, we report numerical experiments to illustrate the regularizing effects of MR-II and compare it with MINRES and the hybrid MINRES as well as LSQR. We demonstrate a few features: (i) the $\gamma_k$ decrease as fast as the $\sigma_k+1$, and MR-II has the full regularization and is as effective as and more efficient than LSQR for severely and moderately ill-posed problems. (ii) MR-II has only the partial regularization for mildly ill-posed problems, and additional regularization is needed. (iii) MINRES has only the partial regularization, independent of the degree of ill-posedness. (iv) the $k$-th iterate by MINRES is always more accurate than the $(k-1)$-th iterate by MR-II before the semi-convergence of MINRES. (v) the hybrid MINRES and MR-II are equally effective for mildly ill-posed problems.

We choose several symmetric ill-posed examples from Hansen’s regularization toolbox [13]. All the problems arise from the discretization of the first kind Fredholm integral equation

\begin{equation}
\int_a^b K(s, t)x(t)dt = b(s), \quad c < s < d.
\end{equation}

For each problem we use the corresponding code in [13] to generate a $1024 \times 1024$ matrix $A$, the true solution $x_{\text{true}}$ and noise-free right-hand $\hat{b}$. In order to simulate the noisy data, we generate the noise vector $e$ whose entries are normally distributed with mean zero and variance one. Defining the noise level $\varepsilon = \frac{\|e\|}{\|b\|}$, we use $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$, respectively, in the test examples. To simulate exact arithmetic, the full reorthogonalization is used during the Lanczos process. All the computations are carried out using Matlab 7.8 with the machine precision $\epsilon_{\text{mach}} = 2.22 \times 10^{-16}$ under the Microsoft Windows 7 64-bit system.

5.1. MR-II for severely and moderately ill-posed problems. We first illustrate the full regularization of MR-II for the following two severely ill-posed problems and show that it generates very best rank $k$ approximations $Q_{k+1}T_kQ_k^T$ to $A$.

Example 1. This problem ‘Shaw’ arises from one-dimensional image restoration, and is obtained by discretizing (5.1) with $[-\frac{\pi}{2}, \frac{\pi}{2}]$ as both integration intervals, where

\begin{equation}
K(s, t) = (\cos(s) + \cos(t))^2 \left(\frac{\sin(u)}{u}\right)^2, \quad u = \pi(\sin(s) + \sin(t)),
\end{equation}

\begin{equation}
x(t) = 2\exp(-6(t - 0.8)^2) + \exp(-2(t + 0.5)^2).
\end{equation}
Example 2. This problem ‘Foxgood’ is obtained by discretizing (5.1) with \([0, 1]\) as both integration intervals, in which

\[
K(s, t) = (s^2 + t^2)^{\frac{3}{2}}, \quad b(s) = \frac{1}{3} \left( (1 + s^2)^{\frac{3}{2}} - s^3 \right), \quad x(t) = t.
\]

In Figure 1, we display the curves of sequences \(\gamma_k\) with the noise levels \(\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}\), respectively. We see that \(\gamma_k \approx \sigma_{k+1} = |\lambda_{k+1}|\), almost independent of noise level \(\varepsilon\). Particularly, due to the round-offs in finite precision arithmetic, they level off at \(\varepsilon_{\text{mach}}\) when \(k = 20\) for ‘Shaw’ and when \(k = 37\) for ‘Foxgood’. The results indicate that the \(Q_{k+1}T_kQ_k^T\) are very best rank \(k\) approximations to \(A\) with the approximate accuracy \(\sigma_{k+1}\) so that \(T_k\) does not become ill-conditioned before \(k \leq k_0\). As a result, the regularized solutions \(x^{(k)}\) become increasingly better approximations to \(x_{\text{true}}\) until iteration \(k_0\), and they are deteriorated after that iteration. At iteration \(k_0\), \(x^{(k_0)}\) captures the \(k_0\) dominant SVD components of \(A\) and is a best possible regularized solution, i.e., MR-II has the full regularization for the test severely ill-posed problems. We will give more details on the full regularization of MR-II in Section 5.2.

In Figure 2, we plot the relative errors \(\|x^{(k)} - x_{\text{true}}\|/\|x_{\text{true}}\|\) with different noise levels for these two problems. Obviously, MR-II exhibits the semi-convergence. Moreover, for smaller noise levels, we get more accurate regularized solutions at cost of more iterations. This is expected since a bigger \(k_0\) is needed for a smaller \(\varepsilon\).

We next demonstrate the full regularization of MR-II for two moderately ill-posed problems.

Example 3. This problem ‘Gravity’ arises from one-dimensional gravity surveying model and is obtained by discretizing (5.1) with \([0, 1]\) as both integration intervals, where

\[
K(s, t) = \frac{1}{4} \left( \frac{1}{16} + (s - t)^2 \right)^{-\frac{3}{2}}, \quad x(t) = \sin(\pi t) + \frac{1}{2}\sin(2\pi t).
\]

Example 4. This problem is Phillips’ test problem and is obtained by discretizing (5.1) with \([-6, 6]\) as both integration intervals, where

\[
K(s, t) = \begin{cases} 
1 + \cos \left( \frac{\pi(s-t)}{3} \right), & |s - t| < 3; \\
0, & |s - t| \geq 3,
\end{cases}
\]

\[
x(t) = \begin{cases} 
1 + \cos \left( \frac{\pi t}{3} \right), & |t| < 3; \\
0, & |t| \geq 3,
\end{cases}
\]

\[
b(s) = (6 - |s|) \left( 1 + \frac{1}{2} \cos \left( \frac{\pi s}{3} \right) \right) + \frac{9}{2\pi} \sin \left( \frac{\pi |s|}{3} \right).
\]

Figure 3 depicts the decaying curves of \(\gamma_k\) for different noise levels. From it we see that the \(\gamma_k\) decrease as fast as the \(|\lambda_{k+1}|\). However, slightly different from severely ill-posed problems, \(\gamma_k\) may not be so close to \(|\lambda_{k+1}|\). This confirms our theory since the \(\eta_k\) defined by (3.5) are generally bigger for moderately ill-posed problems than for severely ill-posed problems. For ‘Gravity’, we have observed that \(\gamma_k\) and \(\sigma_{k+1}\) level off at \(\varepsilon_{\text{mach}}\) at \(k = 50\) due to round-offs in finite precision arithmetic. After \(k = 50\), \(\gamma_k\) and \(\sigma_{k+1}\) are purely round-offs and unreliable. As a whole, though the \(k\)-step Lanczos process can generate more accurate rank \(k\) approximations for severely ill-posed problems than for moderately ill-posed problems, the \(\gamma_k\) are still excellent approximations to the \(\sigma_{k+1}\), so that MR-II has the full regularization.
Fig. 1. (a)-(b): Plots of decaying behavior of the sequences $\gamma_k$ and $|\lambda_{k+1}|$ for the problem Shaw with $\varepsilon = 10^{-2}$ (left) and $\varepsilon = 10^{-3}$ (right); (c)-(d): Plots of decaying behavior of the sequences $\gamma_k$ and $|\lambda_{k+1}|$ for the problem Foxgood with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right).

Fig. 2. The relative error $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for the problems Shaw (left) and Foxgood (right).
In Figure 4, we depict the relative errors of $x^{(k)}$, and observe analogous phenomena to those for severely ill-posed problems. A distinction is that since $\sigma_j$ does not decay as fast as that for a severely ill-posed problem, MR-II needs more iterations to achieve the semi-convergence for moderately ill-posed problems with the same noise level.

5.2. Comparison of MR-II, LSQR and MINRES. For 'Shaw', 'Foxgood', 'Gravity' and 'Phillips', we compare the regularizing effects of LSQR and MR-II. It is remarkable to stress that, for these four problems, we have made numerical experiments on LSQR, confirming that it has the full regularization and obtains regularized solutions which are as best as the hybrid LSQR with the additional TSVD regularization applied to the projected problems using L-curve criterion. We refer to [17] for details on 'Shaw' and 'Phillips'.

As verified in Section 5.1, we have $\gamma_k \approx \sigma_{k+1}$ for the above four test problems. These observations reflect that the pure MR-II has the full regularization for severely and moderately ill-posed problems and no additional regularization is needed. Taking LSQR as a reference, we now give more direct justifications on the full regularization of MR-II.

In the sequel, we only report the results for the noise level $\varepsilon = 10^{-3}$. Results for the other two $\varepsilon$ are analogous and thus omitted unless stated otherwise.
In Figure 5, we plot the relative error \( \| x^{(k)} - x_{true} \| / \| x_{true} \| \) with respect to LSQR and MR-II. Obviously, MR-II behaves almost the same as LSQR, especially for severely ill-posed problems. This illustrates that MR-II has very similar regularizing effects as LSQR and thus has the full regularization. Theoretically, this is expected for severely and moderately ill-posed problems with \( A \) symmetric since the results in [17] and Theorems 3.1–3.2 indicate that the Krylov subspaces \( K_k(A^T A, A b) \) and \( K_k(A, A b) \) capture the \( k \) dominant \( v_i, i = 1, 2, \ldots, k \) with comparable quality for the same \( k \). Therefore, MR-II and LSQR should have comparable regularizing effects and get regularized solutions with comparable quality. However, MR-II is preferred for these problems, because multiplications with \( A \) are halved, compared to LSQR.

As already proved and numerically shown in [17], a hybrid LSQR should be used to compute best possible regularized solutions for mildly ill-posed problems. We now test MR-II, MINRES and their hybrid variants for the following mildly ill-posed problem ‘Deriv2’, showing that MR-II and MINRES has only the partial regularization, and one must use their hybrid variants to compute a best possible regularized solution.

**Example 5.** The problem ‘Deriv2’ is mildly ill-posed, which is obtained by discretizing (5.1) with \([0, 1]\) as both integration intervals, where the kernel \( K(s, t) \) is the Green’s function for the second derivative:

\[
K(s, t) = \begin{cases} 
  s(t - 1), & s < t; \\
  t(s - 1), & s \geq t,
\end{cases}
\]

and the solution \( x(t) \) and the right-hand side \( b(s) \) are given by

\[
x(t) = \begin{cases} 
  t, & t < \frac{1}{2}; \\
  1 - t, & t \geq \frac{1}{2},
\end{cases} \quad b(s) = \begin{cases} 
  (4s^3 - 3s)/24, & s < \frac{1}{2}; \\
  (-4s^3 + 12s^2 - 9s + 1)/24, & s \geq \frac{1}{2}.
\end{cases}
\]

Figure 6 (a) shows that the relative errors of approximate solutions obtained by the hybrid MINRES and MR-II with the additional TSVD regularization applied to the projected problems reach a considerably smaller minimum level than their pure versions for ‘Deriv2’. For this problem, before MINRES or MR-II captures all the dominant spectral components needed, a small singular value of \( T_k \) or \( \bar{T}_k \) appears and starts to deteriorate the approximate solutions. In contrast, their hybrid variants expand Krylov subspaces until enough dominant spectral components are
Fig. 5. The relative errors $\|x^{(k)} - x_{\text{true}}\|/\|x_{\text{true}}\|$ with respect to LSQR and MR-II for the problems Shaw, Foxgood, Gravity, Phillips (from top left to bottom right).

Fig. 6. (a): The relative errors $\|x^{(k)} - x_{\text{true}}\|/\|x_{\text{true}}\|$ with respect to MINRES, hybrid MINRES, MR-II, and hybrid MR-II; (b): The L-curves with respect to MINRES and MR-II for the problem Deriv2.
captured and additional regularization effectively dampens the spectral components corresponding to small eigenvalues. For example, we see from Figure 6 (a) that the semi-convergence of the MR-II occurs at iteration \( k = 3 \), which is also observed by the corner of the L-curve depicted by Figure 6 (b). However, as shown by Figure 6 (a), such regularization indicated by semi-convergence is not enough, and the hybrid MR-II uses a larger six dimensional Krylov subspace \( K_6(A, Ab) \) to improve the solutions and get a best possible regularized solution, whose residual norm is smaller than that obtained by the pure MR-II. After \( k = 6 \), the regularized solutions almost stabilize with the minimum error as \( k \) increases. We observe similar phenomena for MINRES and its hybrid variant, where we find that the relative error by the hybrid MINRES reaches the same minimum level as that by the hybrid MR-II. In the meantime, from Figure 6 (a) we observe that the pure MR-II is better than the pure MINRES since it computes a better regularized solution than MINRES at the occurrence of semi-convergence.

In what follows we now compare MINRES and MR-II and get more insight into their regularizing effects. We show that (i) MINRES generally has only the partial regularization even for severely and moderately ill-posed problems, (ii) the regularized solution by MR-II is always more accurate than that by MINRES when semi-convergence occurs to them, and (iii) the regularized solutions \( \tilde{x}^{(k)} \) by MINRES are always more accurate that the counterparts \( x^{(k-1)} \) by MR-II until the occurrence of semi-convergence of MINRES.

Figures 7 and 8 display numerous curves for severely and moderately ill-posed problems, respectively. Clearly, the regularized solutions by MR-II are always more accurate than those by MINRES at the occurrence of semi-convergence of MINRES. This is because that a small singular value of the projected matrix \( \tilde{T}_k \) appears before a regularized solution becomes best, causing that its error does not reach the same error level as that obtained by MR-II. For instance, we see from Figures 7 (a) and (c) that the best possible regularized solution by MR-II includes seven dominant spectral or SVD components, i.e., \( k_0 = 7 \), but \( \tilde{T}_k \) has already a small singular value lying between \( \sigma_7 \) and \( \sigma_8 \) at iteration \( k = 5 \), making the relative error starts to increase dramatically at that iteration. However, all the singular values of \( T_k \) in MR-II are very good approximations to the large singular values of \( A \) in natural order, and no small singular value of \( T_k \) deteriorates the regularized solution before it becomes best at iteration \( k = 7 \). Similar phenomena are observed for ‘Foxgood’, as indicated by Figures 7 (b) and (d), where \( \tilde{T}_k \) has a small singular value from \( k = 3 \) upwards, causing that the regularized solutions become increasingly more poorly. We have analogous findings for moderately ill-posed problems, as shown by Figure 8, where the \( k \) singular values of \( \tilde{T}_k \) do not approximate the \( k \) large singular values of \( A \) in natural order and the smallest singular value of \( \tilde{T}_k \) is always considerably smaller than \( \sigma_k \) after very few iterations and is between some other two singular values which are increasingly farther from \( \sigma_{k+1} \) with increasing \( k \). This justifies our assertion that MINRES only has the partial regularization and one must use a hybrid MINRES with additional regularization to remove the effects of small singular values of \( \tilde{T}_k \).

We have more observations from Figures 7–8. It is clear that, for each test problem, the \( \tilde{x}^{(k)} \) obtained by MINRES are more accurate than the corresponding \( x^{(k-1)} \) obtained by MR-II until the occurrence of semi-convergence of MINRES. Afterwards, the regularized solutions \( \tilde{x}^{(k)} \) are deteriorated more and more seriously.

Finally, we show that the more expensive and sophisticated hybrid MINRES with the additional TSVD regularization within projected problems works as effectively as
ITERATIVE REGULARIZATION OF MINIMUM-RESIDUAL METHODS

Fig. 7. (a)-(b): The relative errors $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ with respect to MINRES and MR-II; (c)-(d): Plots of the singular values (circles for MINRES, stars for MR-II) of the projected matrix and the ones (solid lines) of $A$, for the problems Shaw (left) and Foxgood (right).

6. Conclusions. For a general large-scale problem (1.1), LSQR and CGLS are most popularly used to compute its regularized solution. However, for $A$ symmetric, MR-II may be a better alternative, as it halves the computational cost of LSQR and CGLS if they use roughly the same iterations to achieve semi-convergence.

We have proved that MR-II captures all the needed dominant spectral information as effectively as LSQR for symmetric severely and moderately ill-posed problems. Numerical experiments have demonstrated that MR-II has the full regularization and can compute the best possible regularized solutions for these two kinds of problems. Our theory has given a partial theoretical support for these observations. On the other hand, our theory and experiments have shown that MR-II has only the partial regularization
and its hybrid variant is necessary for mildly ill-posed problems, while MINRES has only the partial regularization independent of the degree of ill-posedness. We have given illuminating experiments to confirm some new features of MR-II and MINRES that are found in this paper. As a comparison of MR-II and LSQR, our experiments have indicated that MR-II is as effective as LSQR but more efficient than the latter.

For future work, more appealing is an accurate estimate for \( \| \sin \Theta(V_k, V_k^*) \| \) other than \( \| \sin \Theta(V_k, V_k^*) \|_F \). This needs a more subtle analysis and plays a central role in accurately estimating the accuracy of the rank \( k \) approximation generated by the Lanczos process, which is the core problem of completely understanding the regularizing effects of MR-II. Our bounds in Theorems 3.1–3.2 are less sharp and need to be improved on substantially.

REFERENCES

[1] F. Bauer and M. A. Lukas, Comparing parameter choice for regularization of ill-posed problems. Math. Comput. Simulation, 81(2011), pp. 1795–1841.

[2] F. S. V. Bazán, M. C. C. Cunha and L. S. Borges, Extension of GKB-FP algorithm to large-scale general-form Tikhonov regularization. Numer. Linear Algebra Appl., 21 (2014), pp. 316–339.
Fig. 9. The relative errors \( \|x^{(k)} - x_{true}\|/\|x_{true}\| \) with respect to MR-II and MINRES with additional TSVD regularization for the problems Shaw, Foongood, Gravity, Phillips (from top left to bottom right).

[3] Å. Björck, A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. BIT Numer. Math., 28 (1988), pp. 659–670.

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

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

[6] B. Fischer, M. Hanke and M. Hochbruck, A note on conjugate-gradient type methods for indefinite and/or inconsistent linear systems. Numer. Algorithms., 11 (1996), pp. 181–187.

[7] S. Gazzola, Regularization techniques based on Krylov methods for ill-posed linear systems. Ph. D. thesis, Dept. of Mathematics, University of Padua, Italy, 2014.

[8] S. Gazzola, P. Novati and M. R. Russo, On Krylov projection methods and Tikhonov regularization. Available online from http://www.math.unipd.it/~gazzola/pubs.html.

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

[10] M. Hanke, On Lanczos based methods for the regularization of discrete ill-posed problems. BIT Numer. Math., 41 (2001), pp. 1008–1018.

[11] M. Hanke and J. G. Nagy, Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques, Inverse Probl., 12 (1996), pp. 157–173.

[12] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, Philadelphia, PA, 1998.

[13] P. C. Hansen, Regularization tools version 4.0 for Matlab 7.3. Numer. Algorithms., 46 (2007), pp. 189–194.

[14] P. C. Hansen, Discrete Inverse Problems: Insight and Algorithms. SIAM, Philadelphia, PA, 2010.

[15] P. C. Hansen and T. K. Jensen, Noise propagation in regularizing iterations for image de-
blurring. Electron. Trans. Numer. Anal., 31 (2008), pp. 204–220.

[16] B. Hofmann, Regularization for Applied Inverse and Ill-posed Problems. Teubner, Stuttgart, Germany, 1986.

[17] Y. Huang and Z. Jia, Some results on the regularization of LSQR and CGLS for large-scale discrete ill-posed problems. arXiv: math/1503.01864v1, 2015.

[18] T. K. Jensen and P. C. Hansen, Iterative regularization with minimum-residual methods. BIT Numer. Math., 47 (2007), pp. 103–120.

[19] M. E. Kilmer and D. P. O’Leary, Choosing regularization parameters in iterative methods for ill-posed problems. SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221.

[20] M. E. Kilmer and G. W. Stewart, Iterative regularization and minres. SIAM J. Matrix Anal. Appl., 21 (1999), pp. 613–628.

[21] A. Neuman, L. Reichel and H. Sadok, Algorithms for range restricted iterative methods for linear discrete ill-posed problems. Numer. Algorithms., 59 (2012), pp. 325–331.

[22] D. P. O’Leary and J. A. Simmons, A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems. SIAM. J. Sci. Statist. Comput., 2 (1981), pp. 474–489.

[23] C. C. Paige, B. N. Parlett and H. A. Van der Vorst, Approximate solutions and eigenvalue bounds from Krylov subspaces. Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.

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

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

[26] L. Reichel and G. Rodriguez, Old and new parameter choice rules for discrete ill-posed problems. Numer. Algorithms., 63 (2013), pp. 65–87.

[27] G. W. Stewart, Matrix Algorithms. Volume II: Eigensystems, SIAM, Philadelphia, PA, 2001.