SOME RESULTS ON THE REGULARIZATION OF LSQR AND CGLS
FOR LARGE-SCALE DISCRETE ILL-POSED PROBLEMS∗

YI HUANG† AND ZHONGXIAO JIA‡

Abstract. LSQR, a Lanczos bidiagonalization based Krylov method, and its mathematically equivalent CGLS applied to normal equations system, are commonly used for large-scale discrete ill-posed problems. It is well known that LSQR and CGLS have regularizing effects, where the number of iterations plays the role of the regularization parameter. However, it has been unknown whether the regularizing effects are good enough to find best possible regularized solutions. In this paper, we establish bounds for the distance between the k-dimensional Krylov subspace and the k-dimensional dominant right singular space. They show that the Krylov subspace captures the dominant right singular space better for severely and moderately ill-posed problems than for mildly ill-posed problems. Our general conclusions are that LSQR has better regularizing effects for the first two kinds of problems than for the third kind, and a hybrid LSQR with additional regularization, in general, is needed for mildly ill-posed problems. Exploiting the established bounds, we derive an estimate for the accuracy of the rank k approximation generated by Lanczos bidiagonalization. Numerical experiments illustrate that the regularizing effects of LSQR are good enough to compute best possible regularized solutions for severely and moderately ill-posed problems, but a hybrid LSQR must be used for mildly ill-posed problems.

Key words. Ill-posed problem, regularization, severely, moderately, mildly, Lanczos bidiagonalization, LSQR, CGLS, hybrid

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

1. Introduction. We consider the large-scale linear discrete ill-posed problem

\[ \min_{x \in \mathbb{R}^n} \|Ax - b\|, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \]

where the norm \( \|\cdot\| \) is the 2-norm of a vector or matrix, and the matrix \( A \) is 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 restoration, typically when discretizing the first-kind Fredholm integral equations [17, 19]. In particular, the right-hand side \( b \) is affected by noise, caused by measurement or discretization errors, i.e.,

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

where \( e \in \mathbb{R}^m \) represents the Gaussian white noise vector and \( \hat{b} \in \mathbb{R}^m \) denotes the noise-free right-hand side. Because of the presence of noise \( e \) in \( b \) and the ill-conditioning of \( A \), the naive solution \( x_{\text{naive}} = A^\dagger b \) of (1.1) is meaningless and far from the true solution \( x_{\text{true}} = A^\dagger \hat{b} \), where the superscript \( \dagger \) denotes the Moore-Penrose generalized inverse of a matrix. Therefore, it is necessary to use regularization techniques to determine a best possible approximation to \( x_{\text{true}} = A^\dagger \hat{b} \).

Let

\[ A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^T, \]

∗This work was supported in part by the National Basic Research Program of China 2011CB302400 and the National Science Foundation of China (No. 11371219)
†Department of Mathematical Sciences, Tsinghua University, Beijing 100084, People’s Republic of China. (huangyi10@mails.tsinghua.edu.cn)
‡Corresponding author. Department of Mathematical Sciences, Tsinghua University, Beijing 100084, People’s Republic of China. (jiazx@tsinghua.edu.cn)
where $U = (u_1, u_2, \ldots, u_m) \in \mathbb{R}^{m \times m}$ and $V = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^{n \times n}$ are orthogonal matrices, and the diagonal matrix $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \in \mathbb{R}^{n \times n}$ with the $\sigma_i$s being the singular values of $A$ labeled in decreasing order $\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$. Then we obtain

\begin{equation}
\begin{aligned}
    x_{\text{naive}} &= \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i + \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i = x_{\text{true}} + \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i.
\end{aligned}
\end{equation}

Throughout the paper, we assume that $\hat{b}$ satisfies the discrete Picard condition: On average, the coefficients $|u_i^T b|$ decay faster than the singular values. To be definitive, for simplicity we assume that these coefficients satisfy a widely used model [17, 19]:

\begin{equation}
|u_i^T b| = \sigma_i^{1+\beta}, \quad \beta > 0, \quad i = 1, 2, \ldots, n.
\end{equation}

The assumption of the Gaussian white noise means that all the $|u_i^T e|$ are nearly equal. Let $k_0$ be such that $|u_{k_0+1}^T b| \approx |u_{k_0+1}^T e|$. Then $k_0$ is called the transition point, and the truncated SVD (TSVD) method computes

\begin{equation}
x_k^\text{TSVD} = \begin{cases}
    \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i \approx \sum_{i=1}^{k_0} \frac{u_i^T b}{\sigma_i} v_i, & k \leq k_0; \\
    \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i \approx \sum_{i=1}^{k_0} \frac{u_i^T b}{\sigma_i} v_i + \sum_{i=k_0+1}^{k} \frac{u_i^T e}{\sigma_i} v_i, & k > k_0,
\end{cases}
\end{equation}

which can be written as $x_k^\text{TSVD} = A_k^T \hat{b}$, the solution of the modified problem that replaces $A$ by its best rank $k$ approximation $A_k = U_k \Sigma_k V_k^T$ in (1.1), where $U_k = (u_1, \ldots, u_k)$, $V_k = (v_1, \ldots, v_k)$ and $\Sigma_k = \text{diag}(\sigma_1, \ldots, \sigma_k)$. Therefore, the goal of regularization is to capture the $k_0$ dominant SVD components and suppress the other $n-k_0$ small ones.

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

\begin{equation}
\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \lambda^2 \|x\|^2,
\end{equation}

where $\lambda$ is a nonnegative real number, often referred as the regularization parameter. Both the TSVD method and the Tikhonov regularization method can be regarded as parameter-filtered methods, whose solutions can be written in the form

\begin{equation}
x_{\text{filt}} = \sum_{i=1}^{n} f_i \frac{u_i^T b}{\sigma_i} v_i,
\end{equation}

where the filter factors $f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$ for the Tikhonov regularization, and $f_i = 1, i = 1, 2, \ldots, k$ and $f_i = 0, i = k + 1, \ldots, n$ for the TSVD method. An appropriate choice of $\lambda$ must be such that $f_i \approx 1$ for $k_0$ large singular values and $f_i \approx 0$ for $n-k_0$ small singular values so as to avoid the noise deteriorating the solution. Many techniques have been developed, such as discrepancy principle, the L-curve and generalized cross validation; see, e.g., [1, 2, 17, 24, 31] for comparisons of the classical and new ones.

However, it is impractical to compute SVD when (1.1) is large. In this case, one typically projects it onto a sequence of lower dimensional Krylov subspaces and gets iterative solutions. The Conjugate Gradient (CG) method has been used when $A$ is
symmetric definite [15]. Closely related to CG but applied to the normal equations system $A^T Ax = A^T b$, the CGLS algorithm has been studied; see [5, 19] and the references therein. The LSQR algorithm [29], which is mathematically equivalent to CGLS, has attracted great attention, and is known to have regularizing effects and exhibits the semi-convergence [4, 16, 17, 19, 20, 28]: The iterative solutions improve from the beginning to some iteration, then the noise starts to deteriorate the solutions dramatically and their norms become large while 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 [17, 19].

Unfortunately, for a given (1.1), the semi-convergence does not tell us whether or not an iterative solver has found a best possible regularized solution. For Krylov subspace based iterative solvers, their regularizing effects critically rely on how well the underlying $k$-dimensional Krylov subspace captures the $k$ dominant SVD components of $A$. The richer information the Krylov subspace contains on the $k$ dominant SVD components, the less possible 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 the 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, its hybrid variant, e.g., a hybrid LSQR, is needed that combines the solver with additional regularization [4, 11, 25, 26, 27, 28], which aims to remove the effects of small Ritz values and expand the Krylov subspace until it captures all the dominant SVD components needed and the method obtains a best possible regularized solution. The regularization of LSQR and CGLS has been receiving intensive attention over years. Except some informal analysis and qualitative arguments, however, there has been no definitive result or assertion on their full or partial regularization hitherto [17, 19].

Other minimum-residual methods have also gained attention for solving (1.1). For problems with $A$ symmetric, MINRES and its preferred variant MR-II are alternatives and have been shown to have regularizing effects [15]. When $A$ is nonsymmetric and multiplication with $A^T$ is difficult or impractical to compute, GMRES and its preferred variant RRGMRES are candidates [9, 26]. The hybrid approaches based on the Arnoldi process have been introduced in [8, 10, 25]. We point out that, although GMRES and its variants seem suitable for some ill-posed problems, RRGMRES does not have general regularizing effects, as has been addressed in [19, 22]. Recently, Gazzola et al. [12, 13, 14, 27] have studied more methods based on the Lanczos bidiagonalization, the Arnoldi process and the nonsymmetric Lanczos process, and they have described a general framework of the hybrid methods and presented Krylov-Tikhonov methods with different parameter choice strategies employed.

In order to study the regularization of an iterative solver, we need to introduce a precise definition of the degree of ill-posedness, which comes from [21] and has been widely adopted [17, 19]: If there exists an $\alpha > 0$ such that the singular values $\sigma_j = O(j^{-\alpha})$, then the problem is characterized 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 severely ill-posed can be extended to the problem with $\sigma_j = O(\rho^{-j})$ with $\rho > 1$ considerably. It is clear that the singular values $\sigma_j$ of a severely ill-posed problem decay exponentially
at the same rate $\rho^{-1}$, while those of a moderately or mildly ill-posed problem decay more and more slowly at the rate $(j+\frac{2}{\alpha})^\alpha \approx 1$ with increasing $j$.

In this paper, we focus on LSQR and establish some quantitative results on the regularization of LSQR. By them, we will draw some definitive conclusions. We establish bounds for the $F$-norm distance between the underlying $k$-dimensional Krylov subspace and the $k$-dimensional dominant right singular space. There has been no rigorous and quantitative result on the distance before. The results indicate that the $k$-dimensional Krylov subspace captures the $k$-dimensional dominant right singular space better for severely and moderately ill-posed problems than for mildly ill-posed problems. As a result, LSQR has better regularization for the first two kinds of problems than for the third kind. Furthermore, by the bounds and the analysis on them, we draw a definitive assertion that LSQR generally has only the partial regularization for mildly ill-posed problems, so that a hybrid LSQR with additional regularization is needed to compute a best possible regularized solution. We use the bounds to derive an estimate for the accuracy of the rank $k$ approximation, generated by Lanczos bidiagonalization, to $A$, which is closely related to the regularization of LSQR. Our results help further understand the regularization of LSQR. We also derive some other results. Numerical experiments confirm our theory that LSQR has only the partial regularization for mildly ill-posed problems and a hybrid LSQR is needed to compute best possible regularized solutions. Strikingly, the experiments demonstrate that LSQR has the full regularization for severely and moderately ill-posed problems. Our theory gives a partial support for the observed general phenomena. In the paper, all the computation is assumed in exact arithmetic. Since CGLS is mathematically equivalent to LSQR, all the assertions on LSQR apply to CGLS.

This paper is organized as follows. In Section 2, we describe the Lanczos bidiagonalization process and the LSQR algorithm. In Section 3, we present our results and analyze them. In Section 4, we report numerical experiments and observe some definitive and general phenomena. Finally, we conclude the paper in Section 5.

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 $\|\cdot\|_F$ the Frobenius norm of a matrix, and by $I$ the identity matrix with order clear from the context.

2. The LSQR algorithm. LSQR for solving (1.1) is based on the $k$-step Lanczos bidiagonalization process, which computes two orthonormal bases $\{q_1, q_2, \ldots, q_k\}$ and $\{p_1, p_2, \ldots, p_k\}$ for the Krylov subspaces $K_k(A^T A, A^T b)$ and $K_k(A A^T, b)$, respectively. We describe it as Algorithm 1.

\begin{algorithm}
\caption{k-step Lanczos bidiagonalization process}
\begin{algorithmic}[1]
\State Take $p_1 = b/\|b\| \in \mathbb{R}^m$, and define $\beta_1 q_0 = 0$ and $\alpha_{n+1} p_{n+1} = 0$
\For{$j = 1, 2, \ldots, k$}
\State $r_j = A^T p_j - \beta_j q_{j-1}$
\State $\alpha_j = \|r_j\|; q_j = r_j/\alpha_j$
\State $z_j = A q_j - \alpha_j p_j$
\State $\beta_{j+1} = \|z_j\|; p_{j+1} = z_j/\beta_{j+1}$
\EndFor
\State Define $Q_k = (q_1, q_2, \ldots, q_k)$ and $P_{k+1} = (p_1, p_2, \ldots, p_{k+1})$. Then Algorithm 1 can
\end{algorithm}
\end{algorithm}
be written in the matrix form

\begin{align}
AQ_k &= P_{k+1}B_k, \\
A^TP_{k+1} &= Q_kB_k^T + \alpha_{k+1}q_{k+1}e_{k+1}^T,
\end{align}

where \( e_{k+1} \) denotes the \((k+1)\)-th canonical basis vector of \( \mathbb{R}^{k+1} \) and

\[
B_k = \begin{pmatrix}
\alpha_1 \\
\beta_2 \\
\alpha_2 \\
\beta_3 & \ddots \\
\ddots & \ddots & \ddots \\
\beta_{k+1}
\end{pmatrix} \in \mathbb{R}^{(k+1) \times k}.
\]

At the \( k \)-th iteration, LSQR computes the regularized solution \( x^{(k)} = Q_ky^{(k)} \) with

\[
y^{(k)} = \arg \min_{y \in \mathbb{R}^k} ||b||_1 - B_ky||.
\]

Note that \( P_{k+1}^Tb = ||b||_1e_1 \). We get

\begin{equation}
x^{(k)} = Q_kb^{(k)} = ||b||_1Q_kB_k^Te_1 = Q_kB_k^Tp_{k+1}^Tb.
\end{equation}

As stated in the introduction, LSQR exhibits the semi-convergence at some iteration: The regularized solutions \( x^{(k)} \) become better approximations to \( x_{\text{true}} \) until some iteration \( k \), and the noise will dominate the \( x^{(k)} \) after that iteration. The iteration number \( k \) plays the role of the regularization parameter. However, the semi-convergence does not necessarily mean that LSQR finds a best possible regularized solution as \( B_k \) may become ill conditioned before \( k \leq k_0 \) but \( x^{(k)} \) does not yet contain all the needed \( k_0 \) dominant SVD components of \( A \). In this case, in order to get a best possible regularized solution, one has to use a hybrid LSQR method in the way described in the introduction.

3. The regularization of LSQR. As stated in the introduction, the regularizing effects of LSQR critically depends on what information \( \mathcal{K}_k(A^TA, A^Tb) \) contains and provides. Note that the eigenvalues and eigenvectors of \( A^TA \) are \( \sigma_j^2 \), \( j = 1, 2, \ldots, n \) and the right singular vectors of \( A \), and \( B_k^TB_k \) is the projected matrix of \( A^Tb \) over the subspace \( \mathcal{K}_k(A^TA, A^Tb) \) [5]. We have a general claim adapted from [30] and exploited widely in [17, 19]: The more information it contains on the \( k \) dominant right singular vectors, the more possibly and more accurately the \( k \) Ritz values approximate large singular values of \( A \); on the other hand, the less information it contains on the other \( n-k \) right singular vectors, the less accurate a small Ritz value, if it appears. For our problem, since the small singular values of \( A \) are clustered and close to zero, a small Ritz value will show up until \( k \) increases to some point, and it starts to appear more late when \( \mathcal{K}_k(A^TA, A^Tb) \) contains less information on the other \( n-k \) right singular vectors. In this sense, we say that LSQR has better regularization.

Using the definition of canonical angles \( \Theta(\mathcal{X}, \mathcal{Y}) \) between the two subspaces \( \mathcal{X} \) and \( \mathcal{Y} \) of the same dimension [32, p. 250], we have the following theorem, which, to some extent, shows how well the subspace \( \mathcal{K}_k(A^TA, A^Tb) \), on which LSQR and CGLS work, captures the \( k \)-dimensional dominant right singular space.

**Theorem 3.1.** Let the SVD of \( A \) be defined as (1.2). Assume that the singular values of \( A \) are of the form \( \sigma_j = \mathcal{O}(e^{-\alpha j}) \) with \( \alpha > 0 \) considerably. Let
\( \mathcal{V}_k = \text{span}\{V_k\} \) be the subspace spanned by the columns of \( V_k = (v_1, v_2, \ldots, v_k) \), and \( \mathcal{V}_k = \mathcal{K}_k(A^T A, A^T b) \). Then

\[
\| \sin \Theta(\mathcal{V}_k, \mathcal{V}_k^\perp) \|_F \leq \frac{\sigma_{k+1}}{\sigma_k} \frac{|u_{k+1}^T b|}{|u_k^T b|} \sqrt{k(n-k)}O(1), \quad k = 1, 2, \ldots, n - 1. \tag{3.1}
\]

**Proof.** Let \( \bar{U} = (u_1, u_2, \ldots, u_n) \) consist of the first \( n \) columns of \( U \) defined in (1.2). We see \( \mathcal{K}_k(\Sigma^2, \Sigma U^T b) \) is spanned by the columns of the \( n \times k \) matrix \( DT_k \) with

\[
D = \text{diag}(\sigma_i \bar{U}_i^T b), \quad T_k = \begin{pmatrix}
1 & \sigma_1^2 & \ldots & \sigma_{k-1}^2 \\
1 & \sigma_2^2 & \ldots & \sigma_{k-1}^2 \\
\vdots & \vdots & \ddots & \vdots \\
1 & \sigma_n^2 & \ldots & \sigma_{k-1}^2
\end{pmatrix}.
\]

Partition the matrices \( D \) and \( T_k \) as follows:

\[
D = \begin{pmatrix}
D_1 & 0 \\
0 & D_2
\end{pmatrix}, \quad T_k = \begin{pmatrix}
T_{k1} \\
T_{k2}
\end{pmatrix},
\]

where \( D_1, T_{k1} \in \mathbb{R}^{k \times k} \). Since \( T_{k1} \) is a Vandermonde matrix with \( \sigma_j \) distinct for \( 1 \leq j \leq k \), it is nonsingular. Thus, by the SVD of \( A \), we have

\[
\mathcal{K}_k(A^T A, A^T b) = \text{span}\{VDT_k\} = \text{span}\left\{ V \begin{pmatrix} D_1 T_{k1} \\ D_2 T_{k2} \end{pmatrix} \right\} = \text{span}\left\{ V \begin{pmatrix} I \\ \Delta_k \end{pmatrix} \right\}
\]

with \( \Delta_k = D_2 T_{k2} T_{k1}^{-1} D_1^{-1} \). 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 \( \mathcal{V}_k^\perp \).

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

\[
\| \sin \Theta(\mathcal{V}_k, \mathcal{V}_k^\perp) \|_F = \| (V_k^\perp)^T Z_k(Z_k^T Z_k)^{-\frac{1}{2}} \|_F
\]

\[
= \left\| (V_k^\perp)^T V \left( I + \Delta_k^T \Delta_k \right)^{-\frac{1}{2}} \right\|_F
\]

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

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

\[
\leq \| \Delta_k \|_F = \| D_2 T_{k2} T_{k1}^{-1} D_1^{-1} \|_F. \tag{3.2}
\]

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

\[
L_i^{(k)}(\lambda) = \prod_{j=1, j \neq i}^{k} \frac{\sigma_j^2 - \lambda}{\sigma_i^2 - \sigma_j^2}
\]

that interpolates the elements of the \( i \)-th canonical basis vector \( e_i^{(k)} \in \mathbb{R}^k \) at the abscissas \( \sigma_1^2, \ldots, \sigma_k^2 \). Consequently, the \( i \)-th column of \( T_{k2} T_{k1}^{-1} \) is

\[
T_{k2} T_{k1}^{-1} e_i^{(k)} = \left( L_i^{(k)}(\sigma_{k+1}^2), \ldots, L_i^{(k)}(\sigma_n^2) \right)^T,
\]
from which we obtain

\[
(3.3) \quad T_{k_2T_{k_1}^{-1}} = \begin{pmatrix}
L_1^{(k)}(\sigma_{k+1}^2) & L_2^{(k)}(\sigma_{k+2}^2) & \cdots & L_k^{(k)}(\sigma_{k+1}^2) \\
L_1^{(k)}(\sigma_{k+2}^2) & L_2^{(k)}(\sigma_{k+2}^2) & \cdots & L_k^{(k)}(\sigma_{k+2}^2) \\
\vdots & \vdots & \ddots & \vdots \\
L_1^{(k)}(\sigma_n^2) & L_2^{(k)}(\sigma_n^2) & \cdots & L_k^{(k)}(\sigma_n^2)
\end{pmatrix}.
\]

Since \(|L_i^{(k)}(\lambda)|\) is monotonic for \(\lambda < \sigma_i^2\), it is bounded by \(|L_i^{(k)}(0)|\). Furthermore, let \(|L_i^{(k)}(0)| = \max_{i=1,2,\ldots,k} |L_i^{(k)}(0)|\). Then for \(i = 1,2,\ldots,k\) and \(\alpha > 0\) considerably we have

\[
|L_i^{(k)}(0)| \leq |L_i^{(k)}(0)| = \prod_{j=1}^{k} \left| \frac{\sigma_j^2}{\sigma_j^2 - \sigma_{i_0}} \right| = \prod_{j=1}^{i_0-1} \frac{1}{1 - \mathcal{O}(e^{-2(i_0-j)\alpha})} \prod_{j=i_0+1}^{k} \frac{1}{1 - \mathcal{O}(e^{-2(j-i_0)\alpha}) - 1}
\]

\[
= \frac{1}{\prod_{j=1}^{i_0-1} \mathcal{O}(e^{-2(i_0-j)\alpha})} \frac{1}{\prod_{j=i_0+1}^{k} \mathcal{O}(e^{-2(j-i_0)\alpha})}
\]

\[
= \frac{1 + \sum_{j=1}^{i_0-1} \mathcal{O}(e^{-2j\alpha})}{\prod_{j=i_0+1}^{k} \mathcal{O}(e^{-2(j-i_0)\alpha})}
\]

\[
= \mathcal{O}(1)
\]

when \(i_0 = k\) or near to it by noticing that the last second quantity becomes smaller for the other \(i_0\). From this and (3.3) it follows that

\[
\|T_{k_2T_{k_1}^{-1}}\|_F \leq \sqrt{k} \|T_{k_2T_{k_1}^{-1}}e_k^{(k)}\| \leq \sqrt{k(n-k)}|L_i^{(k)}(0)| = \sqrt{k(n-k)}\mathcal{O}(1).
\]

Therefore, for \(i = 1,2,\ldots,n-1\) and \(\alpha > 0\) considerably we have

\[
\| \sin \Theta(\mathcal{V}_k, \mathcal{V}_k^\perp) \|_F \leq \| D_2T_{k_2T_{k_1}^{-1}}D_1^{-1} \|_F \leq \| D_2 \| \| T_{k_2T_{k_1}^{-1}} \|_F \| D_1^{-1} \| \leq \frac{\sigma_{k+1}}{\sigma_k} \frac{|u_{k+1}^T b|}{|u_k^T b|} \| T_{k_2T_{k_1}^{-1}} \|_F \leq \frac{\sigma_{k+1}}{\sigma_k} \frac{|u_{k+1}^T b|}{|u_k^T b|} \sqrt{k(n-k)}\mathcal{O}(1).
\]

**Remark 3.1** In the same way, we can justify that \(\max_{i=1,2,\ldots,k} |L_i^{(k)}(0)| = \mathcal{O}(1)\) for the exponentially decaying singular values \(\sigma_j = \mathcal{O}(\rho^{-j})\) with \(\rho > 1\) considerably, so the theorem still holds for a general severely ill-posed problem. Furthermore, the theorem can be extended to moderately ill-posed problems with the singular values \(\sigma_j = \mathcal{O}(j^{-\alpha})\), \(\alpha > 1\) considerably and \(k\) not big since, in a similar manner to the
proof of Theorem 3.1, we can prove
\[ |L^{(k)}_{i_0}(0)| = \prod_{j=1}^{k} \left| \frac{\sigma_j^2}{\sigma_{i_0}^2 - \sigma_j^2} \right| = \prod_{j=1}^{i_0-1} \frac{1}{\left(1 - \mathcal{O}(\frac{1}{i_0})^2\right)} \cdot \prod_{j=i_0+1}^{k} \frac{1}{\mathcal{O}(\frac{1}{i_0})^2 - 1} = \mathcal{O}(1). \]

However, for mildly ill-posed problems, we have \( |L^{(k)}_{i_0}(0)| > 1 \) considerably for \( \alpha < 1 \).

The theorem and the above analysis mean that \( V^*_k \) captures \( V_k \) considerably better for severely and moderately ill-posed problems than mildly ill-posed problems. As a result, \( V^*_k \) captures more information on the other \( n - k \) right singular vectors for mildly ill-posed problems, making the appearance of a small Ritz value more possible.

\[ \text{Thus, we have } \text{LSQR has better regularization for the first two kinds of problems than for the third kind. Note that LSQR, at most, has the full regularization, i.e., there is no Ritz value less than } \sigma_{i_0+1} \text{ for } k \leq k_0 \text{, for severely and moderately ill-posed problems. Our analysis indicates that LSQR generally has only the partial regularization for mildly ill-posed problem and a hybrid LSQR should be used.} \]

**Remark 3.2** (3.1) and the above analysis also indicate that \( V^*_k \) 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 \( \sigma_{k+1}/\sigma_k \) are basically fixed constants \( \rho^{-1} \) for severely ill-posed problems, which are smaller than those ratios for moderately ill-posed problems unless \( \alpha \) is rather big and \( k \) small. The second is that the quantities \( L^{(k)}_{i_0}(0) \) for severely ill-posed problems are generally smaller than those for moderately ill-posed problems.

Recall the discrete Picard condition (1.4), and consider the coefficients
\[ c_k = \frac{|u_{k+1}^T b|}{|u_k^T b|} = \frac{|u_{k+1}^T \hat{b} + u_{k+1}^T e|}{|u_k^T \hat{b} + u_k^T e|} \approx \frac{\sigma_{k+1}^{1+\beta} + |u_{k+1}^T e|}{\sigma_k^{1+\beta} + |u_k^T e|} \]

We observe that the larger \( \beta \) is, the smaller \( c_k \approx \frac{\sigma_{k+1}^{1+\beta}}{\sigma_k^{1+\beta}} < 1 \) for \( k \leq k_0 \) and thus the better \( V^*_k \) captures \( V_k \). For \( k > k_0 \), note that all the \( |u_k^T b| \approx |u_k^T e| \) almost remain the same. Thus, we have \( c_k \approx 1 \), and \( V^*_k \) may not capture \( V_k \) so well after iteration \( k_0 \).

**Remark 3.3** (3.1) should not be sharp. As we have seen from the proof, it seems that the presence of the factor \( \frac{\sigma_{k+1}}{\sigma_k} \frac{|u_{k+1}^T b|}{|u_k^T b|} \) is unavoidable. However, we conjecture that the factor \( \sqrt{n-k} \) is superfluous and should be replaced by a much smaller factor \( \mathcal{O}(1) \). Unfortunately, we are currently unable to remove it.

Let us investigate more and get insight into the regularization of LSQR. Define
\[ \gamma_k = \|A - P_{k+1}B_kQ_k^T\|, \]
which measures the quality of the rank \( k \) approximation \( P_{k+1}B_kQ_k^T \), which is generated by Lanczos bidiagonalization, to \( A \). Based on (3.1), we can derive the following estimate for \( \gamma_k \).

**Theorem 3.2.** Assume that (1.1) is severely or moderately ill posed. Then
\[ \sigma_{k+1} \leq \gamma_k \leq (1 + \eta_k)\sigma_{k+1}, \]

where \( \eta_k = \frac{\sigma_1\sqrt{k(n-k)}|u_{k+1}^T b|}{\sigma_k|u_k^T b|} \mathcal{O}(1). \)

**Proof.** Let \( A_k = U_k\Sigma_kV_k^T \) be the best rank \( k \) approximation to \( A \) with respect to the 2-norm, where \( U_k = (u_1, \ldots, u_k) \), \( V_k = (v_1, \ldots, v_k) \) and \( \Sigma_k = \text{diag}(\sigma_1, \ldots, \sigma_k) \).
Since the rank of $P_{k+1}B_kQ_k^T$ is $k$, the lower bound in (3.6) is trivial by noting that $\gamma_k \geq \|A - A_k\| = \sigma_{k+1}$. We now prove the upper bound. From (2.1) we obtain
\[
\|A - P_{k+1}B_kQ_k^T\| = \|A - AQ_kQ_k^T\|.
\]
It is known from Algorithm 1 that $V_k^T = Q_k(A^T A, A^T b) = \text{span}\{Q_k\}$ with $Q_k$ having orthonormal columns. Since $\|\sin \Theta(V_k, V_k^*)\| \leq \|\sin \Theta(V_k, V_k^*)\|_F$, from Theorem 3.1 we get
\[
\|A - AQ_kQ_k^T\| = \|(A - U_k\Sigma_k V_k^T + U_k\Sigma_k V_k^T)(I - Q_k Q_k^T)\| \\
\leq \|(A - U_k\Sigma_k V_k^T)(I - Q_k Q_k^T)\| + \|U_k\Sigma_k V_k^T (I - Q_k Q_k^T)\| \\
\leq \sigma_{k+1} + \|\Sigma_k\| \|V_k^T (I - Q_k Q_k^T)\| \\
= \sigma_{k+1} + \sigma_1 \|\sin \Theta(V_k, V_k^*)\| \\
\leq (1 + \eta_k)\sigma_{k+1}. \quad \square
\]

Numerically, it has been extensively observed in, e.g., [3, 13, 14] that the $\gamma_k$ decay as fast as $\sigma_{k+1}$ and, more precisely, $\gamma_k \approx \sigma_{k+1}$ for severely ill-posed problems, meaning that the $P_{k+1}B_kQ_k^T$ are very best rank $k$ approximations to $A$. As our experiments will indicate in detail, these observed phenomena are of generality for both severely and moderately ill-posed problems and thus should have theoretical supports. Compared to the observations, our estimate (3.6) appears pessimistic since $\eta_k$ in it is considerably bigger than one. Our less accurate estimate is partly due to the fact that we have used $\|\sin \Theta(V_k, V_k^*)\|_F$ to replace $\|\sin \Theta(V_k, V_k^*)\|$ in the proof, which may amplify $\eta_k$ roughly by a multiple $\sqrt{k}$ since $\|\sin \Theta(V_k, V_k^*)\| \leq \frac{1}{\sqrt{k}} \|\sin \Theta(V_k, V_k^*)\|_F$. More importantly, we believe that there are some other factors hidden and unknown to us that affect the estimate accuracy of $\gamma_k$ at the time of our current work. How to dig those hidden factors to accurately estimate $\gamma_k$ is crucial to completely understand the regularizing effects of LSQR. We now have no way other than leave this as future research. Fortunately, as the first key step towards analyzing $\gamma_k$, we have derived a bound for the key quantity $\|\sin \Theta(V_k, V_k^*)\|_F$, which has been used to estimate $\gamma_k$. More appealing is to derive accurate bounds for $\|\sin \Theta(V_k, V_k^*)\|$ other than $\|\sin \Theta(V_k, V_k^*)\|_F$, which definitely needs a more subtle analysis.

We next present some results on $\alpha_{k+1}$ in (2.2). If $\alpha_{k+1} = 0$, the Lanczos bidiagonalization process terminates, and we have found $k$ exact singular triples of $A$ [23]. In our context, since $A$ has only simple singular values and $b$ has components in all the left singular vectors, early termination is impossible in exact arithmetic, but small $\alpha_{k+1}$ is possible. In practice, Lanczos bidiagonalization stops when $\alpha_{k+1}$ is small enough, depending on the context. We investigate how fast $\alpha_{k+1}$ decays. We first give a refinement of a result in [14].

**Theorem 3.3.** Let $B_k = W_k\Theta_k S_k^T$ be the SVD of $B_k$, where $W_k \in \mathbb{R}^{(k+1) \times (k+1)}$ and $S_k \in \mathbb{R}^{k \times k}$ are orthogonal, and $\Theta_k \in \mathbb{R}^{(k+1) \times k}$, and define $\tilde{V}_k = P_{k+1}W_k$ and $\tilde{V}_k = Q_k S_k$. Then
\[
\tilde{A}\tilde{V}_k - \tilde{U}_k\tilde{\Theta}_k = 0,
\]
\[
\|A^T \tilde{U}_k - \tilde{V}_k \tilde{\Theta}_k^T\| = \alpha_{k+1}.
\]

**Proof.** From (2.1) and $B_k = W_k\Theta_k S_k^T$, we obtain
\[
\tilde{A}\tilde{V}_k = AQ_k S_k = P_{k+1}B_k S_k = \tilde{U}_k \tilde{\Theta}_k.
\]
So (3.7) holds. From (2.2), we get
\[
A^T \tilde{U}_k = A^T P_{k+1} W_k = Q_k B_k \epsilon_{k} + \alpha_{k+1} q_{k+1} e_{k+1}^T W_k = Q_k S_h \Theta_k^T + \alpha_{k+1} q_{k+1} e_{k+1}^T W_k = \tilde{V}_k \Theta_k^T + \alpha_{k+1} q_{k+1} e_{k+1}^T W_k.
\]

Note that \(\|q_{k+1}\| = \|e_{k+1}^T W_k\| = 1\). Then we have
\[
\|A^T \tilde{U}_k - \tilde{V}_k \Theta_k^T\| = \alpha_{k+1} \|q_{k+1} e_{k+1}^T W_k\| = \alpha_{k+1}.
\]

We remark that it is an inequality other than the equality in a result of [14] similar to (3.8).

Recall the first paragraph of this section. This theorem shows that once the entry \(\alpha_{k+1}\) becomes small for not big \(k\), the singular values of \(B_k\) may approximate the large singular values of \(A\), and no small one appears for severely ill-posed problems and possibly moderately ill-posed problems.

As our final result, we establish an intimate and interesting relationship between \(\alpha_{k+1}\) and \(\gamma_k\), showing how fast \(\alpha_{k+1}\) decays.

**Theorem 3.4.** It holds that
\[
\alpha_{k+1} \leq \gamma_k.
\]

**Proof.** With the notations as in Theorem 3.3, we have \(P_{k+1} B_k \Theta_k^T = \tilde{U}_k \Theta_k \tilde{V}_k^T\). So, by (3.5), we have
\[
\gamma_k = \|A - \tilde{U}_k \Theta_k \tilde{V}_k^T\|.
\]

Note that \(\tilde{U}_k^T \tilde{U}_k = I\). Therefore, from (3.9) we obtain
\[
\alpha_{k+1} = \|A^T \tilde{U}_k - \tilde{V}_k \Theta_k^T\| = \|A^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k^T\| = \|A^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k^T \tilde{U}_k \tilde{U}_k^T\| \\
\leq \|A - \tilde{U}_k \Theta_k \tilde{V}_k^T\| \|\tilde{U}_k \tilde{U}_k^T\| = \gamma_k.
\]

The theorem indicates that \(\alpha_{k+1}\) decays at least as fast as \(\gamma_k\), which, in turn, means that \(\alpha_{k+1}\) may decrease in the same rate as \(\sigma_{k+1}\), as observed in [3, 13, 14] for severely ill-posed problems.

4. **Numerical experiments.** In this section, we report numerical experiments to illustrate the regularizing effects of LSQR. We will demonstrate that LSQR has the full regularization for severely and moderately ill-posed problems, stronger phenomena.
than our theory proves, but it only has the partial regularization for mildly ill-posed problems, in accordance with our theory, for which a hybrid LSQR is needed to compute best possible regularized solutions. We choose several ill-posed examples from Hansen’s regularization toolbox [18]. All the problems arise from the discretization of the first kind Fredholm integral equation

\[(4.1) \quad \int_a^b K(s, t)x(t)dt = b(s), \quad c < s < d.\]

For each problem we use the corresponding code of [18] to generate a 1024 × 1024 matrix A, true solution x_{true} and noise-free right-hand \( \hat{b} \). In order to simulate the noisy data, we generate the Gaussian noise vector e whose entries are normally distributed with mean zero and variance one. Defining the noise level \( \varepsilon = \frac{\|e\|}{\|\hat{b}\|} \), 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 bidiagonalization process. All the computations are carried out in Matlab 7.8 with the machine precision \( \varepsilon_{\text{mach}} = 2.22 \times 10^{-16} \) under the Microsoft Windows 7 64-bit system.

4.1. Case for severely ill-posed problems. We consider the following two severely ill-posed problems [18].

**Example 1.** This problem 'Shaw' arises from one-dimensional image restoration and is obtained by discretizing the first kind Fredholm integral equation (4.1) with \([-\frac{\pi}{2}, \frac{\pi}{2}]\) as both integration intervals. The kernel \( K(s, t) \) and the solution \( x(t) \) are given by

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

\[ x(t) = 2 \exp(-6(t - 0.8)^2) + \exp(-2(t + 0.5)^2). \]

**Example 2.** This problem 'Wing' has a discontinuous solution and is obtained by discretizing the first kind Fredholm integral equation (4.1) with \([0, 1]\) as both integration intervals. The kernel \( K(s, t) \) and the solution \( x(t) \) are given by

\[ K(s, t) = t \exp(-st^2), \quad b(s) = \frac{\exp(-\frac{1}{9}s) - \exp(-\frac{4}{9}s)}{2s}, \]

\[ x(t) = \begin{cases} 1, & \frac{1}{4} < t < \frac{2}{3}; \\ 0, & \text{elsewhere}. \end{cases} \]

These two problems are severely ill-posed, whose singular values \( \sigma_j = O(e^{-\alpha j}) \) with \( \alpha = 2 \) for 'Shaw' and \( \alpha = 4.5 \) for 'Wing', respectively.

In Figure 1, we display the curves of the sequences \( \gamma_k \) with \( \varepsilon = 10^{-2}, 10^{-3}, 10^{-4} \), respectively. The figure clearly illustrates that the quantities \( \gamma_k \) decrease as fast as \( \sigma_{k+1} \) and both of them level off at the level of \( \varepsilon_{\text{mach}} \) for \( k \) no more than 20. We can see that the decaying curves with different noise levels are almost the same, which means that the size of \( \|e\| \) has little effect on \( \gamma_k \). Furthermore, we observe that \( \gamma_k \approx \sigma_{k+1} \) for severely ill-posed problems, indicating that the \( P_{k+1}B_kQ_k^T \) are very best rank \( k \) approximations to \( A \) with the approximate accuracy \( \sigma_{k+1} \) and that \( B_k \) does not
become ill-conditioned before \( k \leq k_0 \). As a result, the regularized solutions \( x^{(k)} \) become 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 \( A \) only and suppress the other \((n-k_0)\) SVD components, so that it is a best possible regularized solution. Consequently, the pure LSQR has the full regularization for severely ill-posed problems. We will give more direct justifications on this property in Section 4.3.

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, LSQR exhibits the semi-convergence phenomenon. Moreover, for smaller noise level, we get better regularized solutions at the cost of more iterations, as expected.

4.2. Case for moderately ill-posed problems. We now consider the following two moderately ill-posed problems [18].

Example 3. This problem 'Heat' arises from the inverse heat equation, and is obtained by discretizing Volterra integral equation of the first kind, a class of equations that is moderately ill-posed, with \([0, 1]\) as integration interval. The kernel
SOME RESULTS ON REGULARIZATION OF LSQR AND CGLS

13

Fig. 2. The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for the problems Shaw (left) and Wing (right).

$$K(s, t) = k(s - t)$$ with

$$k(t) = \frac{t^{-3/2}}{2\sqrt{\pi}} \exp\left(-\frac{1}{4t}\right).$$

Example 4. This problem is Phillips’ test problem and is be obtained by discretizing the first kind Fredholm integral equation (4.1) with $[-6, 6]$ as both integration intervals. The kernel $K(s, t)$, the solution $x(t)$ and the right-hand side $b(s)$ are given by

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

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

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

From Figure 3, we see that $\gamma_k$ decreases as fast as $\sigma_{k+1}$. However, slightly different from severely ill-posed problems, we can see that the $\gamma_k$ are not be so close to the $\sigma_{k+1}$. This indicates that the $k$-step Lanczos bidiagonalization generates more accurate rank $k$ approximations for severely ill-posed problems than for moderately ill-posed problems. The reason is that $K_k(A^T A, A^T b)$ captures the $k$ dominant right singular vectors better for severely ill-posed problems than for moderately ill-posed problems, as our theory shows. Nonetheless, we have seen that, for the test moderately ill-posed problems, all the $\gamma_k$ are still excellent approximations to the $\sigma_{k+1}$, so that LSQR still has the full regularization.

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 now LSQR needs more iterations for moderately ill-posed problems with the same noise level, as should be the case.
Fig. 3. (a)-(b): Plots of decaying behavior of the sequences $\gamma_k$ and $\sigma_{k+1}$ for the problem Heat with $\varepsilon = 10^{-2}$ (left) and $\varepsilon = 10^{-3}$ (right); (c)-(d): Plots of decaying behavior of the sequences $\gamma_k$ and $\sigma_{k+1}$ for the problem Phillips with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right).

Fig. 4. The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for the problems Heat (left) and Phillips (right).
4.3. Comparison of LSQR with and without additional TSVD regularization. For the previous four severely and moderately ill-posed problems, we now compare the regularizing effects of the pure LSQR and the hybrid LSQR with the additional TSVD regularization used within projected problems, in which we determine the TSVD solutions for projected problems using the L-curve criterion. We will show that LSQR has the full regularization and a best possible regularized solution has been found for each problem at the semi-convergence of LSQR, so no additional regularization is needed.

In the sequel, we only report the results for the noise level $\varepsilon = 10^{-3}$. Results for other $\varepsilon$ are analogous and thus omitted.

Figures 5 (a)-(b) and Figures 6 (a)-(b) indicate that the relative errors of approximate solutions obtained by the two methods reach the same minimum level, and the hybrid LSQR just stabilizes the regularized solutions with the minimum error. This means that the pure LSQR itself has already found a best possible regularized solution at some iteration $k$ and no additional regularization is needed. Our only task is to determine such $k$, which is the iteration where $\|x^{(k+1)}\|$ starts to increase dramatically while its residual norm remains almost unchanged. The L-curve criterion fits nicely into this task. In these examples, we also choose $x_{\text{reg}} = \text{arg min}_k \|x^{(k)} - x_{\text{true}}\|$ for the pure LSQR. Figure 5 (c) and Figures 6 (c)-(d) show that the regularized solutions are generally good approximations to the true solutions. However, we should point out that for the problem 'Wing' with a discontinuous solution, the large relative error indicates that the regularized solution is a poor approximation to the true solution, as depicted in Figure 5 (d). Such phenomenon is due to the unsuitable smoothing term $\lambda^2\|x\|$ in Tikhonov regularization. For such solutions, instead of Tikhonov regularization, Total Variation Regularization should be used, in which the smoothing term $\lambda^2\|x\|$ is replaced by $\lambda^2\|Lx\|_1$ with $L \neq I$ some $p \times n$ matrix and $\|\cdot\|$ the 1-norm.

In what follows, we compare the regularizing effects of the pure LSQR and hybrid LSQR for mildly ill-posed problems, showing that LSQR has only the partial regularization and a hybrid LSQR should be used for this kind of problem.

**Example 5.** The problem 'deriv2' is mildly ill-posed, which is obtained by discretizing the first kind Fredholm integral equation (4.1) with $[0,1]$ as both integration intervals. The kernel $K(s,t)$ is 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 7 (a) shows that the relative errors of approximate solutions by the hybrid LSQR reach a considerably smaller minimum level than those by the pure LSQR, a clear indication that LSQR has the partial regularization. On the other hand, the hybrid LSQR expands the Krylov subspace until it contains enough dominant SVD components and, meanwhile, additional regularization effectively dampen the SVD components corresponding to small singular values. For instance, the semi-convergence of the pure LSQR occurs at iteration $k = 3$, but it is not enough. As the hybrid LSQR shows, we need a larger six dimensional Krylov subspace $K_6(A^T A, A^T b)$ to construct a best possible regularized solution. We also choose $x_{\text{reg}} = \text{arg min}_k \|x^{(k)} - x_{\text{true}}\|$
Fig. 5. (a)-(b): The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to LSQR and LSQR with additional TSVD regularization for $\epsilon = 10^{-3}$; (c)-(d): The regularized solutions $x_{reg}$ for the pure LSQR for the problems Shaw (left) and Wing (right).

for the pure LSQR and the hybrid LSQR. Figure 7 (b) indicates that the regularized solution obtained by the hybrid LSQR is a considerably better approximation to $x_{true}$ than that by the pure LSQR, especially in the non-smooth middle part of $x_{true}$.

5. Conclusions. For large-scale discrete ill-posed problems, LSQR and CGLS are commonly used methods. The methods have regularizing effects and exhibit the semi-convergence. However, if a small Ritz value appears before LSQR captures all the needed dominant SVD components, it has only the partial regularization and must be equipped with additional regularization so that best possible regularized solutions can be found. Otherwise, LSQR has the full regularization and can compute best possible regularized solutions without using additional regularization.

We have proved that the underlying $k$-dimensional Krylov subspace captures the $k$ dimensional dominant right singular space better for severely and moderately ill-posed problems than for mildly ill-posed problems. This makes LSQR have better regularization for the first two kinds of problems than for the third kind. Furthermore, we have shown that LSQR generally has only the partial regularization for mildly ill-posed problems. Numerical experiments have demonstrated that LSQR has the full regularization for severely and moderately ill-posed problems, stronger than our theory predicts, and it has the partial regularization for mildly moderately ill-posed
Some results on regularization of LSQR and CGLS

Fig. 6. (a)-(b): The relative errors \( \| x^{(k)} - x_{\text{true}} \| / \| x_{\text{true}} \| \) obtained by the pure LSQR and LSQR with the additional TSVD regularization for \( \varepsilon = 10^{-3} \); (c)-(d): The regularized solutions \( x_{\text{reg}} \) for the pure LSQR for the problems Heat (left) and Phillips (right).

Fig. 7. The relative errors \( \| x^{(k)} - x_{\text{true}} \| / \| x_{\text{true}} \| \) and the regularized solution \( x_{\text{reg}} \) with respect to LSQR and LSQR with the additional TSVD regularization for the problem Deriv2 and \( \varepsilon = 10^{-3} \).
problems, compatible with our assertion. Together with the observations [3, 13, 14], it appears that the full regularization of LSQR on severely and moderately ill-posed problems are of generality.

As we have stated, an accurate estimate for \(\|\sin \Theta(V^k, V^s_k)\|\) is more appealing than for \(\|\sin \Theta(V^k, V^s_k)\|_F\), as it plays a crucial role in analyzing the accuracy of the rank \(k\) approximation, generated by Lanczos bidiagonalization, to \(A\). Accurate bounds for the accuracy of such rank \(k\) approximation are the core of completely understanding the regularizing effects of LSQR. Since CGLS is mathematically equivalent to LSQR, our results apply to CGLS as well. Our current work has helped to better understand the regularization of LSQR and CGLS. But for a complete understanding of the intrinsic regularizing effects of LSQR and CGLS, we still have a long way to go and do more research.

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 and L. S. Borges, GKB-FP: an algorithm for large-scale discrete ill-posed problems. BIT Numer. Math., 50 (2010), pp. 481–507.
[3] 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.
[4] Å. Björck, A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. BIT Numer. Math., 28 (1988), pp. 659–670.
[5] Å. Björck, Numerical Methods for Least Squares Problems. SIAM, Philadelphia, PA, 1996.
[6] D. Calvetti, G. H. Golub and L. Reichel, Estimation of the L-curve via Lanczos bidiagonalization. BIT Numer. Math., 39 (1999), pp. 603–619.
[7] D. Calvetti, P. C. Hansen and L. Reichel, L-curve curvature bounds via Lanczos bidiagonalization. Electron. Trans. Numer. Anal., 14 (2002), pp. 20–35.
[8] D. Calvetti, B. Lewis and L. Reichel, On the choice of subspace for iterative methods for linear ill-posed problems. Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.
[9] D. Calvetti, B. Lewis and L. Reichel, GMRES, L-curve, and discrete ill-posed problems. BIT Numer. Math., 42 (2002), pp. 44–65.
[10] D. Calvetti and L. Reichel, Tikhonov regularization of large linear problems. BIT Numer. Math., 43 (2003), pp. 263–283.
[11] J. Chuang, J. G. Nagy and D. P. O’Leary, A weighted GCV method for Lanczos hybrid regularization. Electron. Trans. Numer. Anal., 28 (2008), pp. 149–167.
[12] S. Gazzola and P. Novati, Multi-parameter Arnoldi-Tikhonov methods. Electron. Trans. Numer. Anal., 40 (2013), pp. 452–475.
[13] S. Gazzola, Regularization techniques based on Krylov methods for ill-posed linear systems. Ph. d. thesis, Dept. of Mathematics, University of Padua, Italy, 2014.
[14] 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.
[15] M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems. Longman, Essex, 1995.
[16] M. Hanke, On Lanczos based methods for the regularization of discrete ill-posed problems. BIT Numer. Math., 41 (2001), pp. 1008–1018.
[17] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, Philadelphia, PA, 1998.
[18] P. C. Hansen, Regularization tools version 4.0 for Matlab 7.3. Numer. Algorithms., 46 (2007), pp. 189–194.
[19] P. C. Hansen, Discrete Inverse Problems: Insight and Algorithms. SIAM, Philadelphia, PA, 2010.
[20] I. Hnětynková, M. Plešinger, Z. Strakoš, The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data. BIT Numer. Math., 49 (2009), pp. 669–696.
[21] B. Hofmann, Regularization for Applied Inverse and Ill-posed Problems. Teubner, Stuttgart, Germany, 1986.
[22] T. K. Jensen and P. C. Hansen, Iterative regularization with minimum-residual methods.
[23] Z. Jia and D. Niu, An implicitly restarted refined bidiagonalization Lanczos method for computing a partial singular value decomposition. SIAM J. Matrix Anal. Appl., 25 (2003), pp. 246–265.

[24] 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.

[25] B. Lewis and L. Reichel, Arnoldi-Tikhonov regularization methods. J. Comput. Appl. Math., 226 (2009), pp. 92–102.

[26] 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.

[27] P. Novati and M. R. Russo, A GCV based Arnoldi-Tikhonov regularization method. BIT Numer. Math., 54 (2014), pp. 501–521.

[28] 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.

[29] 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.

[30] B. N. Parlett, The Symmetric Eigenvalue Problem. SIAM, Philadephia, PA, 1998.

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

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