Some Results on Regularization of LSQR and CGLS for Large-Scale Discrete Ill-Posed Problems*

Yi Huang\textsuperscript{a}, Zhongxiao Jia\textsuperscript{b}

\textsuperscript{a} Department of Mathematical Sciences, Tsinghua University, 100084 Beijing, China
\textsuperscript{b} Department of Mathematical Sciences, Tsinghua University, 100084 Beijing, China,
E-mail: huangyi10@mails.tsinghua.edu.cn, jiazx@tsinghua.edu.cn

Abstract

For large-scale discrete ill-posed problems, LSQR, a Lanczos bidiagonalization process based Krylov method, is most often used. It is well known that LSQR has natural regularizing properties, where the number of iterations plays the role of the regularization parameter. In this paper, for severely and moderately ill-posed problems, we establish quantitative bounds for the distance between the \(k\)-dimensional Krylov subspace and the subspace spanned by \(k\) dominant right singular vectors. They show that the \(k\)-dimensional Krylov subspace may capture the \(k\) dominant right singular vectors for severely and moderately ill-posed problems, but it seems not the case for mildly ill-posed problems. These results should be the first step towards estimating the accuracy of the rank-\(k\) approximation generated by Lanczos bidiagonalization. We also derive some other results, which help further understand the regularization effects of LSQR. We draw to a conclusion that a hybrid LSQR should generally be used for mildly ill-posed problems. We report numerical experiments to confirm our theory. We present more definitive and general observed phenomena, which will derive more research.

Keywords: Ill-posed problem, regularization, severely, moderately, mildly, Lanczos bidiagonalization, LSQR, CGLS, hybrid method.

Mathematics Subject Classification (2010): 15A18, 65F22, 65J20

1 Introduction

We consider the iterative solution of large-scale discrete ill-posed problems

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

(1.1)

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 with smooth kernels; see [23] for some applications. In particular, the right-hand side \(b\) is affected by noise, caused by measurement or discretization errors, i.e.,

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

\*This work was supported in part by the National Basic Research Program of China 2011CB302400 and the National Science Foundation of China (No. 11371219)
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 \( x_{naive} = A^\dagger b \) of (1.1) is meaningless and far from the true solution \( x_{true} = A^\dagger \hat{b} \), where the superscript \( \dagger \) denotes the Moore-Penrose generalized inverse of a matrix. Therefore, it is necessary to use regularization techniques to determine a meaningful solution to approximate \( x_{true} = A^\dagger \hat{b} \).

A small or moderately sized (1.1) is generally solved by the aid of singular value decomposition (SVD) of \( A \):

\[
A = U \left( \begin{array}{c} \Sigma \\ 0 \end{array} \right) V^T, \tag{1.2}
\]

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 entries of the diagonal matrix \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \in \mathbb{R}^{n \times n} \) are the singular values of \( A \), labelling in decreasing order \( \sigma_1 > \sigma_2 > \cdots > \sigma_n > 0 \). With the SVD of \( A \) we can express the naive solution of (1.1) as

\[
x_{naive} = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^{n} \frac{u_i^T \hat{b}}{\sigma_i} v_i + \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i = x_{true} + \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i. \tag{1.3}
\]

Throughout the paper, we assume that \( \hat{b} \) satisfies the discrete Picard condition: On the average, the the coefficients \(| u_i^T \hat{b} | \) decay faster than the singular values. To be definitive, for simplicity we assume that these coefficients satisfy the model:

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

The assumption of the Gaussian white noise means that \( | u_i^T e | \) is nearly a constant for all \( i \). From (1.3), we see that for large singular values, the \( u_i^T \hat{b} / \sigma_i \) are dominant, compared to the corresponding \( u_i^T e / \sigma_i \). On the other hand, from some \( i \) upwards, whenever \(| u_i^T \hat{b} | \leq | u_i^T e | \), the small singular values magnify the noise coefficients \( | u_i^T e | / \sigma_i \), and the second term makes large contribution to the solution and thus must be suppressed. Let \( k_0 \) denote the transition point where \(| u_{k_0+1}^T \hat{b} | \approx | u_{k_0+1}^T e | \), the above considerations can be described as follows:

\[
x_{TSVD}^k = \begin{cases} \sum_{i=1}^{k} \frac{u_i^T \hat{b}}{\sigma_i} v_i \approx \sum_{i=1}^{k} \frac{u_i^T \hat{b}}{\sigma_i} v_i, & k \leq k_0; \\
\sum_{i=1}^{k} \frac{u_i^T \hat{b}}{\sigma_i} v_i \approx \sum_{i=1}^{k_0} \frac{u_i^T \hat{b}}{\sigma_i} v_i + \sum_{i=k_0+1}^{k} \frac{u_i^T e}{\sigma_i} v_i, & k > k_0. \end{cases}
\]

Thus, we see that the solution norm increases slowly for \( k \leq k_0 \), while it increases dramatically for \( k > k_0 \). With a similar analysis, the residual norm decreases steadily with \( k \) up to \( k_0 \), and it decreases much more slowly for \( k > k_0 \). As a result, the transition point \( k_0 \) is exactly the balance point for which a regularized solution \( x_{TSVD}^k \) contains the information of \( x_{true} \) as much as possible and, on the other hand, dampens the effects of noise \( e \) as much as possible. For details, we refer the reader to [21, 23]. Precisely, it is known [21, 23] that \( k_0 \) is just the one that achieves the overall corner of the L-curve of \( \log \| b - Ax_{TSVD}^k \|, \log \| x_{TSVD}^k \| \), \( k = 1, 2, \ldots, n \). The number \( k_0 \) distinguishes the singular values \( \sigma_i \) into two parts: the large ones for \( i \leq k_0 \) and the small ones for \( i > k_0 \). Therefore, the goal of regularization is to capture the SVD information according to the \( k_0 \) large singular values. It is significant to note that \( x_{TSVD}^k \) is the solution of the modified problem that replaces \( A \) by its best rank \( k \) approximation \( A_k = U_k \Sigma_k V_k^T \) to \( A \) in (1.1), where \( U_k = [u_1, \ldots, u_k], V_k = [v_1, \ldots, v_k] \) and \( \Sigma_k = \text{diag}(\sigma_1, \ldots, \sigma_k) \).
Besides the above truncated SVD (TSVD) regularization method, the other most famous direct regularization method is the Tikhonov regularization, which takes its simplest form

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

where $\lambda$ is a nonnegative real number, often referred as the regularization parameter.

Both the TSVD regularization method and the Tikhonov regularization method can be regarded as parameter-filtered methods, whose solutions can be written in the form

$$x_{filt} = \sum_{i=1}^{n} f_i \frac{u_i^T b}{\sigma_i} v_i,$$

where the filter factors $f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$ for the Tikhonov regularization, and $f_i = 1$, $i = 1, 2, \ldots, k_0$ and $f_i = 0$, $i = k_0 + 1, \ldots, n$ for the TSVD method. A choice of $\lambda$ must be such that $f_i \approx 1$ for large singular values and $f_i \approx 0$ for small singular values by avoiding the noise deteriorating the solution [21, 23].

However, it is impractical to compute SVD when (1.1) is large. In this case, one typically projects it onto a sequence of much smaller problems via iterative methods based on Krylov subspaces. The Conjugate Gradient (CG) method has been widely studied and used as an iterative regularization method when $A$ is symmetric definite [19]. Closely related to CG but applied to normal equations, the CGLS algorithm has been studied for solving (1.1); see [5, 23] and the references therein. The LSQR algorithm [31], which is mathematically equivalent to CGLS, has attracted great attention [4, 20, 30]; for a systematic account, see [21, 23]. LSQR and CGLS exhibit the semi-convergence: the iterative solutions improve from the beginning to some iteration, then the noise starts to deteriorate the solutions drastically and their norms become very large while the residual norms stabilize. A purely iterative method can be directly applied to $Ax = b$ or $\min_x \|Ax - b\|$ and might obtain an acceptable approximate solution by an early termination of the iterations. The semi-convergence phenomenon is due to the fact that a projected problem inherits the ill-conditioning of (1.1) when certain iterations are performed. That is, after some iteration, the appearance of a small singular value, called Ritz value, of the projected problem will amplify the noise considerably [21, 23].

Whenever the mentioned methods do not capture all the dominant SVD components before a small Ritz value appears, their natural regularization is not enough. One then has to combine them with some additional regularization techniques in order to obtain an acceptable regularized solution. The basic idea of a hybrid method is to first project the large-scale problem onto a Krylov subspace of small dimension, and then regularize the projected problem explicitly. The idea was first suggested by O’Leary and Simmons in their bidiagonalization-regularization procedure [30], later described by Bjööck [4], and then further been studied in [12, 27, 28, 29]. A hybrid method aims to remove the effect of small Ritz values, and expand the Krylov subspace until it captures all the dominant SVD components needed. As the projected problem is always of small size, many direct regularization methods, such as the TSVD or Tikhonov regularization, can be used.

Other minimum-residual methods have also gained attention for solving (1.1). For problems with $A$ symmetric, MINRES and its preferred variant MR-II have been shown to have natural regularizing properties [19]. When $A$ is nonsymmetric and multiplication with $A^T$ is difficult or impractical to compute, GMRES and its preferred variant RRGMRES seem natural candidate methods [10, 28]. The hybrid approaches based on the Arnoldi process have been introduced and studied in [9, 11, 15, 27, 29]. We remark that, although GMRES and its variants seem suitable for some ill-posed problems, the methods based on the Arnoldi process do not have general regularizing
properties, as has been addressed in [23, 25]. Since the Arnoldi process does not involve $A^T$, the methods can be expected to be good for $A$ nearly symmetric, but they may fail for a general non-symmetric $A$. For any regularization method, an appropriate choice of the regularization parameter is crucial. Many techniques have been developed, such as discrepancy principle, the L-curve and generalized cross validation; see, e.g., [1, 2, 21, 26, 33] for relevant comparisons and discussions of the classical and new ones. Recently, Gazzola et al. [15, 16, 17, 29] have studied the methods based on the Lanczos bidiagonalization, the Arnoldi process and the nonsymmetric Lanczos process for the severely ill-posed problems (1.1). They have described a general framework of the hybrid methods and present Krylov-Tikhonov methods with different parameter choice strategies employed.

The definition of degree of the ill-posedness follows Hofmann’s book [24]: if there exists a positive real number $\alpha$ such that the singular values satisfy $\sigma_j = \mathcal{O}(j^{-\alpha})$, then $\alpha$ is called the degree of ill-posedness, and the problem is characterized as mildly or moderately ill-posed if $\alpha \leq 1$ or $\alpha > 1$. On the other hand, if $\sigma_j = \mathcal{O}(e^{-\alpha j})$ with $\alpha > 0$, then the problem is termed severely ill-posed. More generally, the definition of severely ill-posed can be extended to the problem with the singular values $\sigma_j = \mathcal{O}(\rho^{-j})$ with $\rho > 1$ decaying to zero exponentially. Severely ill-posed problems are common in many applications, such as computerized tomography when X-ray measurements are available only for some directions, and signal processing when recovering a signal through an ideal low pass filter with a known bandwidth [14, 18]. Also, moderately ill-posed problems arise in many areas of science and engineering, notably in combustion problems when the radial distribution of a quantity in an axis-symmetric flame is inferred from line-of-sight measurements, and atmospheric chemistry when deducing the source distribution from vertically averaged concentration measurements [6, 13].

In this paper, for severely and moderately ill-posed problems, we establish quantitative bounds for $\|\sin \Theta\|_F$ between the $k$-dimensional Krylov subspace and the subspace spanned by $k$ dominant right singular vectors. There has been no rigorous and quantitative result in this regard before. The results indicate that the $k$-dimensional Krylov subspace may capture the $k$ dominant right singular vectors for severely and moderately ill-posed problems, while it seems not the case for mildly ill-posed problems. So the hybrid LSQR should be used for the last kind of problem. The results should be the first step towards to estimating the accuracy of the rank-$k$ approximation generated by Lanczos bidiagonalization to $A$, a core problem that understands the regularization of LSQR. We also derive some other results, which may be helpful to understand the regularization of LSQR. Numerically, we have observed more definitive and general phenomena on LSQR for severely and moderately ill-posed problems, which will derive more research. Throughout the paper, all the computation is assumed in exact arithmetic. Since CGLS is mathematically equivalent to LSQR, all the results proved for LSQR are valid for CGLS.

This paper is organized as follows. In Section 2, we describe the Lanczos bidiagonalization process and the LSQR algorithm, and review its regularizing properties with both intuitive and theoretical considerations. In Section 3 we present our results and analyze them. In Section 4 we give numerical examples and observe some definitive and general phenomena. Finally, we conclude the paper in Section 5.

Some notations are introduced now. 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 Regularizing properties of LSQR and CGLS

LSQR has been widely applied for solving (1.1) in the context of Tikhonov regularization for large (1.1). It 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 $\mathcal{K}_k(A^T A, A^T b)$ and $\mathcal{K}_k(A A^T, b)$, respectively. We describe it as Algorithm 1.

Algorithm 1 $k$-step Lanczos bidiagonalization process

1. Take $p_1 = b/\|b\| \in \mathbb{R}^m$, and define $\beta_1 q_0 = 0$ and $\alpha_{n+1} p_{n+1} = 0$.

2. For $j = 1, 2, \ldots, k$, do
   
   \[ r_j = A^T p_j - \beta_j q_{j-1} \]
   \[ \alpha_j = \|r_j\|; q_j = r_j/\alpha_j \]
   \[ z_j = A q_j - \alpha_j p_j \]
   \[ \beta_{j+1} = \|z_j\|; p_{j+1} = z_j/\beta_{j+1} \]

Endfor.

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 be written in the matrix form

\[ A Q_k = P_{k+1} B_k, \]  \( (2.1) \)
\[ A^T P_{k+1} = Q_k B_k^T + \alpha_{k+1} q_{k+1} e_{k+1}^T, \]  \( (2.2) \)

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

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

At the $k$-th iteration, LSQR computes the approximation $x^{(k)} = Q_k y^{(k)}$, where

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

Furthermore, we have

\[ \min_{x \in \mathcal{K}_k} \|b - Ax\| = \min_{y \in \mathbb{R}^k} \|b - A Q_k y\| \\
= \min_{y \in \mathbb{R}^k} \|b\| P_{k+1} e_1 - P_{k+1} B_k y\| \\
= \min_{y \in \mathbb{R}^k} \|b\| e_1 - B_k y\| \\
= \left\|\|b\| e_1 - B_k y^{(k)}\right\|, \]

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

The above relation shows that the residual norm of the original problem is just that of the projected problem and can thus be computed cheaply. As mentioned previously, LSQR has natural regularizing properties, and the semi-convergence will occur at some step $k$: the residual norms will decrease monotonically and then stabilize at iteration $k$ upwards, while the solution norms increase quickly after that iteration. That is to say, the regularized solutions become better approximations to the true solution until some iteration $k$, and the noise will dominate the regularized solutions after that iteration. Thus, the iteration number $k$ plays the role of the regularization parameter. However, the semi-convergence does not necessarily mean that LSQR delivers an acceptable regularized solution. This is because the projected problem may become ill conditioned but the subspace does not yet capture all the needed dominant SVD components of $A$ associated with large singular values. In this case, in order to get a best possible regularized solution, it is necessary to use a hybrid LSQR method that combines LSQR with some additional regularization, e.g., the TSVD or Tikhonov regularization, used within the projected problems.

LSQR is mathematically equivalent to CGLS, the CG method applied to the normal equation $A^T A x = A^T b$.

Hansen [21, p. 146] describes the convergence of the individual SVD components of the CG iterations. We summarize his analysis as the following theorem, by which we briefly review the regularizing properties of LSQR and CGLS.

**Theorem 2.1.** For LSQR to solve (1) with the starting vector $p_1 = b/\|b\|$ and CGLS with $x^{(0)} = 0$, the $k$-step iterate $x^{(k)}$ satisfies

$$x^{(k)} = \sum_{i=1}^{n} f_i^{(k)} u_i^T \frac{b}{\sigma_i} v_i,$$

where $f_i^{(k)} = 1 - \prod_{j=1}^{k} \left(\frac{\theta_j^{(k)}}{\theta_j^{(k)} - \sigma_i^2}\right)^2$, $i = 1, 2, \ldots, n$, and $\theta_j^{(k)}$, $j = 1, 2, \ldots, k$ are the singular values of $B_k$.

The above expression shows that the regularized solution has the form of filtered SVD solution. Because of the fast convergence of the Ritz values to the largest singular values of $A$, the filter factors $f_i^{(k)}$ roughly equal 1 for the dominant singular values, and they may be small for the small ones. However, once at least a small Ritz value starts to appear, i.e., $\theta_k^{(k)} \leq \sigma_{k+1}$, the corresponding filter factors $f_i^{(k)}$, $i = k + 1, \ldots, n$ are not small, causing that the regularized solution blows up and becomes meaningless. This theorem illustrates the regularizing effects of LSQR and CGLS, and explains why the semi-convergence occurs and why it is enough or not to get an acceptable approximate solution. In the next section we will give more theoretical considerations so as to further and better understand the regularization of LSQR and CGLS.

### 3 Regularization of LSQR and CGLS

As we have seen from (1.4) and (2.3), the regularizing properties of LSQR depend on how the singular values, the Ritz values, of the projected matrix $B_k$ approximate those of $A$. Obviously, their convergence behavior critically depends on what information $K_k(A^T A, A^T b)$ contains and provides.

Using the definition of canonical angles $\Theta(\mathcal{X}, \mathcal{Y})$ between the two subspaces $\mathcal{X}$ and $\mathcal{Y}$ [34, p. 250], we have the following important theorem, which quantitatively shows that, for severely ill-posed problems, the $k$-step Lanczos bidiagonalization process may well capture the $k$ dominant right singular vectors.
Theorem 3.1. Let $A = U \left( \sum \right) V^T = \bar{U} \Sigma V^T$ be the SVD of $A$ as defined in (1.2) with $\bar{U} = (u_1, u_2, \ldots, u_n)$. Assume that the singular values of $A$ are of the form $\sigma_j = O(e^{-\alpha j})$, and the noise-free right-hand side $\bar{b}$ satisfies the discrete Picard condition. Let $\mathcal{V}_k = \text{span}\{V_k\}$ be the subspace spanned by the first $k$ dominant right singular vectors, and $\mathcal{V}_k^* = \mathcal{K}_k(A^T A, A^T b)$. Then

$$\| \sin \Theta(\mathcal{V}_k, V_k^*) \|_F \leq \frac{\sigma_k^{k+1} |u_k^T b|}{\sigma_k |u_k^T b|} \sqrt{k(n - k)O(1), \ k = 1, 2, \ldots, n - 1.} \quad (3.1)$$

Proof. By the SVD of $A$, we see the Krylov subspace $\mathcal{K}_k(\Sigma^2, \Sigma \bar{U}^T b)$ is spanned by the columns of $n \times k$ matrix $DT_k$ with

$$D = \text{diag}(\sigma_i \bar{U}^T b), \ T_k = \begin{pmatrix} 1 & \sigma_2^2 & \cdots & \sigma_{2k-2}^2 \\ 1 & \sigma_2 & \cdots & \sigma_{2k-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \sigma_n^2 & \cdots & \sigma_{2k-2}^2 \end{pmatrix}.$$ 

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

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

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, 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 $\mathcal{V}_k^*$.

Write $V = [V_k, V_k^*]$. By definition we obtain

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

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

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

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

$$= \frac{\| \Delta_k \|_F}{\sqrt{1 + \sigma_{\min}^2(\Delta_k)}}$$

$$\leq \| \Delta_k \|_F = \| D_2 T_{k2} T_{k1}^{-1} D_1^{-1} \|_F, \quad (3.2)$$

where $\sigma_{\min}(C)$ denotes the smallest singular value of a matrix $C$.

Now, we use a transformation technique to 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

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

$$T_{k2}T_{k1}^{-1} e_i^{(k)} = (L_i^{(k)}(\sigma_{k+1}^2), \ldots, L_i^{(k)}(\sigma_n^2))^T.$$ 

Therefore, we obtain

$$T_{k2}T_{k1}^{-1} = \begin{pmatrix}
L_1^{(k)}(\sigma_{k+1}^2) & L_2^{(k)}(\sigma_{k+1}^2) & \cdots & L_k^{(k)}(\sigma_{k+1}^2) \\
L_1^{(k)}(\sigma_{k+2}^2) & L_2^{(k)}(\sigma_{k+2}^2) & \cdots & L_k^{(k)}(\sigma_{k+2}^2) \\
\vdots & \vdots & \ddots & \vdots \\
L_1^{(k)}(\sigma_n^2) & L_2^{(k)}(\sigma_n^2) & \cdots & L_k^{(k)}(\sigma_n^2)
\end{pmatrix}.
$$ (3.3)

Since $|L_i^{(k)}(\lambda)|$ is monotonic for $\lambda < \sigma_k^2$, it is bounded by $|L_i^{(k)}(0)|$. Furthermore, for $i = 1, 2, \ldots, k$, we have

$$|L_i^{(k)}(0)| \leq |L_i^{(k)}(0)| = \left| \prod_{j=1}^{k-1} \frac{\sigma_j^2}{\sigma_j^2 - \sigma_k^2} \right| = \mathcal{O}\left( \prod_{j=1}^{k-1} \frac{1}{1 - e^{-2(j-1)\alpha}} \right) = \mathcal{O}(1),$$

from which and (3.3) it follows that

$$\|T_{k2}T_{k1}^{-1}\|_F \leq \sqrt{k} \|T_{k2}T_{k1}^{-1}e_k^{(k)}\| \leq \sqrt{k(n-k)}|L_k^{(k)}(0)| = \sqrt{k(n-k)}\mathcal{O}(1).$$

Therefore, according to the discrete Picard condition, for $i = 1, 2, \ldots, n-1$ we have

$$\|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^*)\|_F \leq \|D_2T_{k2}T_{k1}^{-1}D_1^{-1}\|_F$$

$$\leq |D_2| \|T_{k2}T_{k1}^{-1}\|_F \|D_1^{-1}\|$$

$$\leq \frac{\sigma_{k+1}}{\sigma_k} \frac{|u_{k+1}^T b|}{|u_k^T b|} \|T_{k2}T_{k1}^{-1}\|_F$$

$$\leq \frac{\sigma_{k+1}}{\sigma_k} \frac{|u_{k+1}^T b|}{|u_k^T b|} \sqrt{k(n-k)}\mathcal{O}(1).$$

\[\square\]

**Remark 3.1.** Observe that the hypothesis $\sigma_j = \mathcal{O}(e^{-\alpha j})$ is just used to estimate $|L_k^{(k)}(0)|$. In the same way, it is easily justified that $|L_k^{(k)}(0)| = \mathcal{O}(1)$ for a general severely ill-posed problem with the exponentially decaying singular values $\sigma_j = \mathcal{O}(\rho^{-j})$ with $\rho > 1$ considerably. However, the result is not limited to severely ill-posed problems only, and it can be extended to moderately ill-posed problems with the singular values $\sigma_j = \mathcal{O}(j^{-\alpha})$, $\alpha > 1$, we obtain

$$|L_k^{(k)}(0)| = \left| \prod_{j=1}^{k-1} \frac{\sigma_j^2}{\sigma_j^2 - \sigma_k^2} \right| = \mathcal{O}\left( \prod_{j=1}^{k-1} \frac{1}{1 - (\frac{k}{j})^{2\alpha}} \right).$$

For $k$ not big or $\alpha > 1$ considerably, it is seen that $\mathcal{O}\left( \prod_{j=1}^{k-1} \frac{1}{1 - (\frac{k}{j})^{2\alpha}} \right) = \mathcal{O}(1)$. As a consequence, the result in Theorem 3.1 holds for moderately ill-posed problem with $\alpha > 1$ considerably. Nonetheless, for mildly ill-posed problems, there is no analogous assertion since $|L_k^{(k)}(0)| \gg 1$ for $\alpha < 1$.  

8
The above theorem establishes a close relationship between the Krylov subspace \( \mathcal{V}_k \) and the dominant right singular subspace \( \mathcal{V}_k^* \). It means that the \( k \)-dimensional Krylov subspace may well capture the \( k \) dominant right singular vectors for severely and moderately ill-posed problems, while it may not be the case for mildly ill-posed problems. Since \( \mathcal{V}_k^* \) may contain substantial SVD components corresponding to small singular values for mildly ill-posed problems, it is very possible that a small Ritz value appears before all the needed dominant SVD components are captured, causing that LSQR has only the partial regularization and a hybrid LSQR should be used.

**Remark 3.2.** (3.1) indicates that, the faster the singular values decay, the more accurately the Krylov subspace captures the the dominant right singular subspace. Furthermore, consider the coefficient

\[
c_k \triangleq \frac{|u_{k+1}^T b|}{|u_k^T b|} = \frac{|u_{k+1}^T \hat{b} + u_{k+1}^T e|}{|u_k^T \hat{b} + u_k^T e|} = \frac{|\sigma_{k+1}^{1+\beta} + u_{k+1}^T e|}{|\sigma_k^{1+\beta} + u_k^T e|}.
\]

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 more accurate \( \mathcal{V}_k^* \) to \( \mathcal{V}_k \) is. For \( k > k_0 \), since all the \( |u_{k+1}^T b| \approx |u_k^T e| \) almost remain the same, \( c_k \approx 1 \) and \( \| \sin \Theta(\mathcal{V}_k, \mathcal{V}_k^*) \|_F \) decay a bit more slowly after the \( k_0 \)-th iteration.

**Remark 3.3.** (3.1) should not be sharp, partly because we have replaced all the \( |L_i^{(k)}(\sigma_j^2)| \) by a possibly considerably bigger bound \( |L_i^{(k)}(0)| \) for \( i = 1, 2, \ldots, k, j = k + 1, \ldots, n \). More importantly, it seems more reasonable that the factor \( \sqrt{n-k} \) should be replaced by a much smaller factor \( \mathcal{O}(1) \).

Let us investigate more and get insight into LSQR. For LSQR to solve (1.1) with the starting vector \( p_1 = b/\|b\|, \) define

\[
\gamma_k = \| A - P_{k+1} B_k Q_k^T \|
\]

which measures the quality of the rank-\( k \) approximation, generated by Lanczos bidiagonalization, to \( A \). We present the following estimate for \( \gamma_k \).

**Theorem 3.2.** Assume that (1.1) is severely ill-posed with the singular value \( \sigma_j = \mathcal{O}(e^{-\alpha j}), \alpha > 0 \). Moreover, assume that the noise-free right-hand side \( \hat{b} \) satisfies the discrete Picard condition. Then

\[
\gamma_k \leq (1 + \eta_k) \mathcal{O}(1) \sigma_{k+1}, \tag{3.4}
\]

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

**Proof.** From (2.1), we obtain

\[
\| A - P_{k+1} B_k Q_k^T \| = \| A - A Q_k Q_k^T \|.
\]

According to Theorem 3.1, it is known that \( \mathcal{V}_k^* = \mathcal{K}_k(A^T A, A^T b) = \text{span}\{Q_k\} \) with \( Q_k \) having orthonormal columns. Therefore, we obtain

\[
\| A - A Q_k Q_k^T \| = \| (A - U_k \Sigma_k V_k^T + U_k \Sigma_k V_k^T)(I - Q_k Q_k^T) \|
\]

\[
\leq \| (A - U_k \Sigma_k V_k^T)(I - Q_k Q_k^T) \| + \| U_k \Sigma_k V_k^T(1 - Q_k Q_k^T) \|
\]

\[
\leq \sigma_{k+1} + \| \Sigma_k \| \| V_k^T (I - Q_k Q_k^T) \|
\]

\[
= \sigma_{k+1} + \| V_k^T (I - Q_k Q_k^T) \|
\]

\[
\leq (1 + \eta_k) \mathcal{O}(1) \sigma_{k+1}.
\]
We remark that our estimate for \((3.4)\) should be rough, partly because we have used \(\| \sin \Theta(V_k, V_k^s) \|_F\) to replace \(\| \sin \Theta(V_k, V_k^s) \|\) in the proof. Numerically, it has been extensively observed in the literature that the \(\gamma_k\) decay as fast as \(\sigma_{k+1}\) for severely ill-posed problems; see, e.g., [3, 16, 17]. These experiments illustrate that our \((3.4)\) is too overestimate and pessimistic. These observed phenomena are of generality and thus should have strong theoretical supports for severely ill-posed problems. As it has appeared that \(V_k^s\) may capture \(V_k\) effectively, we believe that there are definitely some other serious factors hidden and unknown to us at the time of our current work. How to dig those hidden important factors and find out a comprehensive mechanism that fits into estimating \(\gamma_k\) more accurately is both difficult and crucial to 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 to analyzing \(\gamma_k\), we have derived a quantitative bound for \(\| \sin \Theta(V_k, V_k^s) \|_F\), which should play a key role in deriving accurate estimates for \(\gamma_k\). On the other hand, a more accurate bound for \(\| \sin \Theta(V_k, V_k^s) \|\) other than \(\| \sin \Theta(V_k, V_k^s) \|_F\) is very appealing, but we definitively need a more subtle analysis approach than what we have done for \(\| \sin \Theta(V_k, V_k^s) \|_F\).

We next present more results on \(\alpha_{k+1}\), which may help further understand LSQR. We give first a refinement of a result in Gazzola et al. [17].

**Theorem 3.3.** Let \(B_k = W_k\Theta_kS_k^T\) be the full SVD of \(B_k\), where \(W_k \in \mathbb{R}^{(k+1) \times (k+1)}\) and \(\Theta_k \in \mathbb{R}^{(k+1) \times k}\), and let \(\tilde{U}_k = P_{k+1}W_k, \tilde{V}_k = Q_kS_k\). Then
\[
A\tilde{V}_k - \tilde{U}_k\Theta_k = 0, \tag{3.5}
\]
\[
\left\| A^T\tilde{U}_k - \tilde{V}_k\Theta_k^T \right\| = \alpha_{k+1}. \tag{3.6}
\]

*Proof.* From (2.1) and \(B_k = W_k\Theta_kS_k^T\), we obtain
\[
A\tilde{V}_k = AQ_kS_k = P_{k+1}B_kS_k = \tilde{U}_k\Theta_k.
\]
Then from (2.2), we have
\[
A^T\tilde{U}_k = A^TP_{k+1}W_k = Q_kB_k^TW_k + \alpha_{k+1}q_{k+1}e_{k+1}^TW_k = Q_kS_k\Theta_k^T + \alpha_{k+1}q_{k+1}e_{k+1}^TW_k = \tilde{V}_k\Theta_k^T + \alpha_{k+1}q_{k+1}e_{k+1}^TW_k = \tilde{V}_k\Theta_k^T + \alpha_{k+1}q_{k+1}w_{k+1}^T,
\]
where \(w_{k+1}\) denotes the \(k\)-th row of the orthonormal matrix \(W_k\). Note that \(\|q_{k+1}\| = \|w_{k+1}\| = 1\). Then we get
\[
\left\| A^T\tilde{U}_k - \tilde{V}_k\Theta_k \right\| = \alpha_{k+1}\| q_{k+1}w_{k+1}^T \| = \alpha_{k+1}. \tag{3.7}
\]

We remark that it is an inequality other than an equality in (3.6) of [17].

As is well known, LSQR is one way to achieve the Ritz-Galerkin process on the Krylov subspace \(K_k(A^TA, A^Tb)\). The singular values of \(B_k\) are approximations to some singular values of \(A\). For a general singular value distribution, only some of the \(k\) Ritz values rather than all of them become good approximations to both largest and smallest singular values; if the small singular values are clustered, small Ritz values are expected not to appear very early, and they converge very slowly [32]. For severely and even moderately ill-posed problems, since the small singular values of \(A\) are
both clustered and close to zero, this means that all the Ritz values may not be small for not big 
k. So all of them may only be approximations to large singular values of \( A \). A small Ritz value
will show up until \( k \) increases up to some point. Therefore, the theorem shows that once the entry
\( \alpha_k \) becomes small for not big \( k \), the \( k \) singular values of \( B_k \) may be good approximations to the
largest singular values of \( A \).

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

**Theorem 3.4.** We have

\[ \alpha_{k+1} \leq \gamma_k. \]  

**Proof.** From (3.7), we know

\[ \alpha_{k+1} = \| A^T \tilde{U}_k - \tilde{V}_k \Theta_k^T \| . \]

Note that \( \tilde{U}_k^T \tilde{U}_k = I \). Therefore, we obtain

\[
\alpha_{k+1} = \left\| A^T \tilde{U}_k - \tilde{V}_k \Theta_k^T \right\| \\
= \left\| A^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k \tilde{U}_k^T \right\| \\
= \left\| A^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k \tilde{U}_k^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k \tilde{U}_k^T \tilde{U}_k \tilde{U}_k^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k \tilde{U}_k^T \right\| \\
\leq \left\| A^T - \tilde{V}_k \Theta_k^T \right\| \left\| \tilde{U}_k \tilde{U}_k^T \right\| \\
= \left\| A - \tilde{U}_k \Theta_k \tilde{U}_k \right\| = \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 a manner similar to \( \sigma_{k+1} \) for severely ill-posed problems, as observed in [3, 16, 17].

### 4 Numerical experiments

In this section, we report numerical experiments to illustrate the the regularizing properties of
LSQR. We choose several ill-posed examples from Hansen’s regularization toolbox [22]. All the
problems arise from the discretization of the first kind Fredholm integral equation

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

In each case, we generate a 256 × 256 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 \( \epsilon_{\text{mach}} = 2.22 \times 10^{-16} \) under the Microsoft
Windows 7 64-bit system.
4.1 Case for severely ill-posed problems

**Example 4.1.** This problem 'Shaw' arises from one-dimensional image restoration, and can be 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

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

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

**Example 4.2.** This problem 'Wing' can also be 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

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

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

We use the Matlab codes *shaw* and *wing* to generate triples $A, x_{true}, \hat{b}$. These two problems are severely ill-posed with the degree of ill-posedness $\alpha = 2$ for 'Shaw' and $\alpha = 4.5$ for 'Wing'.

In Figure 1, we display the curves of the sequence $\gamma_k$ with $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$, respectively. They illustrate that the quantities $\gamma_k$ decrease as fast as $\sigma_k+1$. Furthermore, we can see that the decaying curves with different noise levels almost stay unchanged. Recalling Remark 3.2, we conclude that $c_k$ and thus $\gamma_k$ may be affected by the noise level only very marginally. For severely ill-posed problems, we observe that $\gamma_k$ critically depends on the decaying rate of singular values, and there is no significant difference with different noise levels.

In Figure 2, we plot the relative error $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with different noise levels for these two problems. Obviously, LSQR exhibits the semi-convergence phenomenon. Moreover, for smaller noise level, we get better regularized approximate solutions at cost of more iterations, as expected.

4.2 Case for moderately ill-posed problems

**Example 4.3.** This problem 'Heat' arises from the inverse heat equation, and can be 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 $K(s, t) = k(s - t)$ with

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

**Example 4.4.** This problem is Phillips’ test problem. It can 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

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

$$b(s) = \exp\left( -\frac{1}{\pi}s \right) - \exp\left( -\frac{4}{\pi}s \right) / 2s.$$
Figure 1: (a)-(b): Plots of decaying behavior of the sequences $\gamma_k$ and $\sigma_{k+1}$ 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 $\sigma_{k+1}$ for the problem Wing with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right).
Figure 2: The relative error $\frac{\|x^{(k)} - x_{true}\|}{\|x_{true}\|}$ with respect to $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for the problems Shaw (left) and Wing (right).

We use the Matlab codes heat and phillips to generate $A, x_{true}, \hat{b}$. These two problems are regarded as moderately ill-posed.

From Figure 3, we see that the $\gamma_k$ decrease as fast as $\sigma_{k+1}$. However, unlike the case for severely ill-posed problems, $\gamma_k$ may not be as close as $\sigma_{k+1}$. Furthermore, the figure shows that the decaying curves of $\gamma_k$ are slightly affected by different noise levels. We can observe that $\gamma_k$ deviates from $\sigma_{k+1}$ more with larger noise level. Indeed, as the singular values for moderately ill-posed problems do not decay as fast as those for ill-posed problems, $\gamma_k$ is also affected by $c_k$. Since $c_k$ approximates one quickly with large noise level, which makes $\gamma_k$ not so close to $\sigma_{k+1}$ after $c_k \approx 1$. This also justifies our assertions in Remark 3.2. By comparing the behavior of $\gamma_k$ for severely and moderately ill-posed problem, we come to the conclusion that the $k$-step Lanczos bidiagonalization can generate more accurate rank-$k$ approximations for severely ill-posed problems than for moderately ill-posed problems. Based on these observations, it is expected that LSQR has better regularization for severely ill-posed problems. This is partly supported by our Theorem 3.1, which shows that Lanczos bidiagonalization captures $V_k$ more accurately for severely ill-posed problems than for moderately ill-posed problems.

In Figure 4, we depict the relative error 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.

4.3 Comparison of LSQR with and without additional TSVD regularization

Example 4.5. For the severely ill-posed problems ’Shaw’ and ’Wing’, and the moderately ill-posed problems ’Heat’ and ’Phillips’, we compare the regularizing properties of LSQR and LSQR with the additional TSVD regularization used within projected problems, showing that the pure
Figure 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).
Figure 4: The relative error $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for the problems Heat (left) and Phillips (right).

LSQR is enough to obtain a best possible regularized solution for each problem and no additional regularization is needed for these problems.

Figures 5 (a)-(b) and Figures 6 (a)-(b) indicate that the relative errors obtained by two methods reach the same minimum level, and the additional regularization just stabilizes the regularized solution with the minimum error. This means there is an iteration step $k$ such that the $k$-step LSQR itself finds a best possible approximate solution and no additional regularization is needed. Our task is to determine such $k$, which is the iteration where $\|x^{(k+1)}\|$ starts to increase drastically but its residual norm remains almost unchanged. If it is the case, we get a best possible acceptable approximate solution $x^{(k)}$. This is just the L-curve criterion. In these examples, we also choose $x_{reg} = \arg \min_k \|x^{(k)} - x_{true}\|$ for pure LSQR. Figure 5 (c) and Figures 6 (c)-(d) show that the regularized solutions are generally excellent approximations to the true solutions. However, we should point out that for the problem 'Wing' with a discontinuous solution as depicted in Figure 5 (d), the large relative error indicates that the regularized solution is a poor approximation to the true solution. Such phenomenon is due to the unsuitable smoothing term $\lambda^2 \|x\|$ in the Tikhonov regularization for a discontinuous solution, which should be replaced by a more general form $\lambda^2 \|Lx\|$ with $L \neq I$ some $p \times n$ matrix [23].

**Example 4.6.** For mildly ill-posed problems, we compare the regularizing properties of LSQR and LSQR with the additional TSVD regularization used within projected problems, showing that LSQR has only the partial regularization and a hybrid LSQR should be used for this kind of problem.

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 5: (a)-(b): The relative error $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to LSQR and LSQR with additional TSVD regularization; (c)-(d): The regularized solution $x_{reg}$ for the pure LSQR for the problems Shaw (left) and Wing (right).
Figure 6: (a)-(b): The relative error $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ obtained by the pure LSQR and LSQR with the additional TSVD regularization; (c)-(d): The regularized solution $x_{reg}$ for the pure LSQR for the problems Heat (left) and Pillips (right).
Figure 7: The relative error $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ and the regularized solution $x_{reg}$ with respect to the pure LSQR and LSQR with the additional TSVD regularization for the problem Deriv2.

We use the Matlab code deriv2 to generate triples $A, x_{true}, \hat{b}$. Figure 7 (a) shows that the relative error by the LSQR with the additional TSVD regularization reaches a considerably smaller minimum level than the pure LSQR. This means that before the pure LSQR itself captures all the dominant SVD components needed, a small Ritz value appears and starts to deteriorate the approximate solutions. On the other hand, the hybrid LSQR has the ability to expand the Krylov subspace until it contains enough dominant SVD components and meanwhile to effectively dampen the SVD components corresponding to small singular values. More precisely, the semi-convergence of the pure LSQR occurs at iteration $k = 4$, but it is not enough. As the hybrid LSQR shows, we need a larger six dimensional Krylov subspace $K_6(A^TA, A^Tb)$ to construct the best regularized solution. This indicates that a hybrid LSQR should be used for mildly ill-posed problems. We also choose $x_{reg} = \arg\min_k \|x^{(k)} - x_{true}\|$ for the pure LSQR and LSQR with the additional TSVD regularization. Figure 7 (b) indicates that the regularized solution obtained by the hybrid LSQR is a considerably better approximation to $x_{true}$ than the pure LSQR, especially in the non-smooth middle part of $x_{true}$.

5 Conclusions

For many large-scale problems (1.1), LSQR and CGLS are the most efficient approaches to computing regularized solutions. The methods have natural regularizing properties and exhibit the semi-convergence. If a small Ritz value appears before the methods capture all the needed dominant SVD components, the methods must be equipped with an additional regularization technique. Otherwise, the pure LSQR is expected to be enough to compute a best possible regularized solution without any additional regularization needed.

We have proved that for severely and moderately ill-posed problems the $k$-dimensional Krylov subspace may well capture the first $k$ dominant right singular vectors. However, it may not be the case for mildly ill-posed problems since it is very possible that the Krylov subspace contains substantial SVD components corresponding to small singular values, so that a small Ritz value starts to appear before all the needed dominant SVD components are captured. As a result, a hybrid LSQR should be used for mildly ill-posed problems. For severely and moderately ill-posed
problems, a direct consequence of the above is that the $k$ singular values of $B_k$ may all approximate the $k$ large singular values of $A$, and no small Ritz value generally appears. Our theory has given a partial support that the pure LSQR may be enough to compute best possible approximate solutions without using additional regularization for severely and moderately ill-posed problems. Our numerical experiments have demonstrated that the pure LSQR works very well for severely and moderately ill-posed problems. Such findings seem to be of generality. Unfortunately, our experiments have demonstrated that a hybrid LSQR should be used to compute best possible regularized solutions for mildly ill-posed problems.

Theorem 3.1 on $\sin \Theta$ should play the important role in analyzing the accuracy of the rank-$k$ approximation, generated by Lanczos bidiagonalization, to $A$. A detailed and accurate analysis of such rank-$k$ approximation is the core of 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 and a thorough analysis of the intrinsic regularizing effects of LSQR and CGLS, we still have a long way to go, and more research is obviously needed.

References


