SOME RESULTS ON THE REGULARIZATION OF LSQR FOR LARGE-SCALE DISCRETE ILL-POSED PROBLEMS[*]

YI HUANG[†] AND ZHONGXIAO JIA[‡]

Abstract. LSQR, a Lanczos bidiagonalization based Krylov subspace iterative 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 long been unknown whether the regularizing effects are good enough to find best possible regularized solutions. Here a best possible regularized solution means that it is at least as accurate as the best regularized solution obtained by the truncated singular value decomposition (TSVD) method. 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 is generally 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, stronger than our theory predicts, but they are not for mildly ill-posed problems and additional regularization is needed.

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

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

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

$$(1.1) \qquad \min_{x \in \mathbb{R}^n} \|Ax - b\|, \quad A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m, \ m \geq n,$$

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 Fredholm integral equations of the first-kind [20, 22]. 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, and it is supposed that $\|e\| < \|\hat{b}\|$. 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 to determine a best possible approximation to $x_{true} = A^\dagger \hat{b}$ [14, 18, 20, 22].

[†]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)

The solution of (1.1) can be analyzed by the SVD of A:

$$(1.2) \qquad A = U \begin{pmatrix} \Sigma \\ \mathbf{0} \end{pmatrix} V^T,$$

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 $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \in \mathbb{R}^{n \times n}$ are the singular values of A, which are assumed to be simple throughout the paper and labelled in decreasing order $\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$. With (1.2), we obtain

$$(1.3) \qquad x_{naive} = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^{n} \frac{u_i^T \hat{b}}{\sigma_i} v_i + \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i = x_{true} + \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i.$$

Throughout the paper, we assume that $\hat{b}$ satisfies the discrete Picard condition: On average, the coefficients $| u_i^T \hat{b} |$ decay faster than the singular values. To be definitive, for simplicity we assume that these coefficients satisfy a widely used model in the literature, e.g., [20, p. 81, 111 and 153] and [22, p. 68]:

$$(1.4) \qquad | u_i^T \hat{b} | = \sigma_i^{1+\beta}, \quad \beta > 0, \ i = 1, 2, \ldots, n.$$

Let k_0 be the transition point such that $|u_{k_0}^T \hat{b}| > |u_{k_0+1}^T e|$ and $|u_{k_0+1}^T \hat{b}| \leq |u_{k_0+1}^T e|$ [22, p. 98]. Then the TSVD method computes

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

which can be written as $x_k^{TSVD} = A_k^\dagger b$, 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 $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$. Remarkably, x_k^{TSVD} is the minimum-norm least squares solution of the perturbed problem that replaces A in (1.1) by its best rank k approximation A_k, and the best possible TSVD solution of (1.1) by the TSVD method is $x_{k_0}^{TSVD}$ [22, p. 98]. A number of approaches have been proposed for determining k_0, such as discrepancy principle, discrete L-curve and generalized cross validation; see, e.g., [1, 2, 20, 27, 35] for comparisons of the classical and new ones. In our numerical experiments, we use the L-curve criterion in the TSVD method and hybrid LSQR. The TSVD method has been widely studied; see, e.g., [4, 20, 22, 29].

For a small and moderate (1.1), the TSVD method has been used as a general-purpose reliable and efficient numerical method for solving (1.1). As a result, we will take the TSVD solution $x_{k_0}^{TSVD}$ as a standard reference when assessing the regularizing effects of iterative solvers and accuracy of iterates under consideration in this paper.

As well known, it is generally not feasible to compute SVD when (1.1) is large. In this case, one typically projects (1.1) onto a sequence of low dimensional Krylov subspaces and gets a sequence of iterative solutions [18, 20, 22, 37]. The Conjugate Gradient (CG) method has been used when A is symmetric definite [18]. As a CG-type method applied to the semidefinite linear system $Ax = b$ or the normal equations system $A^T A x = A^T b$, the CGLS algorithm has been studied; see [6, 20, 22] and the references therein. The LSQR algorithm [33], which is mathematically equivalent to

CGLS, has attracted great attention, and is known to have regularizing effects and exhibits semi-convergence (see [20, p. 135], [22, p. 110], and also [5, 19, 23, 32]): The iterates tend to be better and better approximations to the exact solution x_{true} and their norms increase slowly and the residual norms decrease. In later stages, however, the noise e starts to deteriorate the iterates, so that they will start to diverge from x_{true} and instead converge to the naive solution x_{naive}, while their norms increase considerably and the residual norms stabilize. Such phenomenon is due to the fact that a projected problem inherits the ill-conditioning of (1.1). That is, as the iterations proceed, the noise progressively enters the solution subspace, so that a small singular value of the projected problem appears and the regularized solution is deteriorated.

As far as an iterative solver for solving (1.1) is concerned, a central problem is whether or not a pure iterative solver has already obtained a best possible regularized solution at semi-convergence, namely whether or not the regularized solution at semi-convergence is at least as accurate as $x_{k_0}^{TSVD}$. As it appears, for Krylov subspace based iterative solvers, their regularizing effects critically rely on how well the underlying k-dimensional Krylov subspace captures the k-dimensional dominant right singular subspace of A. The richer information the Krylov subspace contains on the k-dimensional dominant right singular subspace, the less possible a small Ritz value of the resulting projected problem appears and thus the better regularizing effects the solver has. To precisely describe the regularizing effects of an iterative solver, we introduce the term of *full* or *partial* regularization: If the iterative solver itself computes a best possible regularized solution at semi-convergence, it is said to have the full regularization; in this case, no additional regularization is needed. Here, as defined in the abstract, a best possible regularized solution means that it is at least as accurate as the best regularized solution obtained by the truncated singular value decomposition (TSVD) method. Otherwise, it is said to have the partial regularization; in this case, in order to compute a best possible regularized solution, its hybrid variant, e.g., a hybrid LSQR, is needed that combines the solver with additional regularization [5, 13, 28, 30, 31, 32], 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 study of the regularizing effects of LSQR and CGLS has been receiving intensive attention for years; see [20, 22] and the references therein. However, there has yet been no definitive result or assertion on their full or partial regularization.

To proceed, we need the following definition of the degree of ill-posedness, which follows Hofmann's book [24] and has been commonly used in the literature, e.g., [20, 22]: If there exists a positive real number α such that the singular values satisfy $\sigma_j = \mathcal{O}(j^{-\alpha})$, then the problem is termed as mildly or moderately ill-posed if $\alpha \leq 1$ or $\alpha > 1$; if $\sigma_j = \mathcal{O}(e^{-\alpha j})$ with $\alpha > 0$ considerably, $j = 1, 2, \ldots, n$, then the problem is termed severely ill-posed. It is clear that the singular values σ_j of a severely ill-posed problem decay exponentially at the same rate ρ^{-1}, while those of a moderately or mildly ill-posed problem decay more and more slowly at the decreasing rate $\left(\frac{j}{j+1}\right)^{\alpha}$ approaching one with increasing j, which, for the same j, is smaller for the moderately ill-posed problem than it for the mildly ill-posed problem.

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 [18]. When A is nonsymmetric and multiplication with A^T is difficult or impractical to compute, GMRES and its preferred variant RRGMRES are candidates [10, 30]. The hybrid approach based on

the Arnoldi process was first introduced in [11], and has been studied in [9, 11, 12, 28]. Recently, Gazzola *et al.* [15, 16, 17, 31] have studied more methods based on the Lanczos bidiagonalization, the Arnoldi process and the nonsymmetric Lanczos process for the severely ill-posed problem (1.1). They have described a general framework of the hybrid methods and present Krylov-Tikhonov methods with different parameter choice strategies employed.

In this paper, we focus on LSQR. We derive bounds for the 2-norm distance between the underlying k-dimensional Krylov subspace and the k-dimensional 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 regularizing effects for the first two kinds of problems than for the third kind. By the bounds and the analysis on them, we draw a definitive conclusion that LSQR generally has only the partial regularization for mildly ill-posed problems, so that a hybrid LSQR with additional explicit regularization is needed to compute a best possible regularized solution. We also 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 to further understand the regularization of LSQR, though they appear less sharp. In addition, we derive a bound on the diagonal entries of the bidiagonal matrices generated by the Lanczos bidigonalization process, showing how fast they decay. 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. Throughout 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 LSQR algorithm, and then present our theoretical results on LSQR with a detailed analysis. In Section 3, we report numerical experiments to justify the partial regularization of LSQR for mildly ill-posed problems. We also report some definitive and general phenomena observed. Finally, we conclude the paper in Section 4.

Throughout the paper, we denote by $\mathcal{K}_k(C, w) = span\{w, Cw, \ldots, C^{k-1}w\}$ 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 regularization of LSQR. LSQR for solving (1.1) is based on the Lanczos bidiagonalization process, which starts with $p_1 = b/\|b\|$ and, at step (iteration) k, computes two orthonormal bases $\{q_1, q_2, \ldots, q_k\}$ and $\{p_1, p_2, \ldots, p_k\}$ of the Krylov subspaces $\mathcal{K}_k(A^TA, A^Tb)$ and $\mathcal{K}_k(AA^T, b)$, respectively.

Define the matrices $Q_k = (q_1, q_2, \ldots, q_k)$ and $P_{k+1} = (p_1, p_2, \ldots, p_{k+1})$. Then the k-step Lanczos bidiagonalization can be written in the matrix form

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

where e_{k+1} denotes the $(k + 1)$-th canonical basis vector of $\mathbb{R}^{k+1}$ and the quantities $\alpha_i, i = 1, 2, \ldots, k + 1$ and $\beta_i, \ i = 2, \ldots, k + 1$ denote the diagonal and subdiagonal

elements of the $(k + 1) \times k$ lower bidiagonal matrix B_k, respectively. At iteration k, LSQR computes the solution $x^{(k)} = Q_k y^{(k)}$ with

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

Note that $P_{k+1}^T b = \|b\| e_1$. We get

(2.3) $$x^{(k)} = Q_k y^{(k)} = \|b\| Q_k B_k^\dagger e_1 = Q_k B_k^\dagger P_{k+1}^T b.$$

As stated in the introduction, LSQR exhibits semi-convergence at some iteration: The iterates $x^{(k)}$ become better approximations to x_{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, 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, as described in the introduction. The significance of (2.3) is that the LSQR iterates can be interpreted as the minimum-norm least squares solutions of the perturbed problems that replace A in (1.1) by its rank k approximations $Q_k B_k^\dagger P_{k+1}^T$, whose nonzero singular values are just those of B_k. If the singular values of B_k approximate the k large singular values of A in natural order for $k = 1, 2, \ldots, k_0$, then LSQR must have the full regularization, and the regularized solution $x^{(k_0)}$ is best possible and is as comparably accurate as the best possible regularized solution $x_{k_0}^{TSVD}$ by the TSVD method.

Hansen's analysis [20, p. 146] shows that the LSQR iterates have the filtered SVD expansions:

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

where $f_i^{(k)} = 1 - \prod_{j=1}^k \frac{\theta_j^{(k)} - \sigma_i}{\theta_j^{(k)}}$, $i = 1, 2, \ldots, n$, and $\theta_j^{(k)}$, $j = 1, 2, \ldots, k$ are the singular values of B_k. In our context, if we have $\theta_k^{(k)} \leq \sigma_{k_0+1}$ for some $k \leq k_0$, the factors $f_i^{(k)}$, $i = k+1, \ldots, n$ are not small, meaning that $x^{(k)}$ is already deteriorated and becomes a poorer regularized solution, namely, LSQR surely does not have full regularization. As a matter of fact, in terms of the best possible solution $x_{k_0}^{TSVD}$, it is easily justified that the full regularization of LSQR is equivalent to requiring that the singular values of B_k approximate the k largest singular values of A in natural order for $k = 1, 2, \ldots, k_0$, so it is impossible to have $\theta_k^{(k)} \leq \sigma_{k_0+1}$ for $k \leq k_0$.

The regularizing effects of LSQR critically depend on what $\mathcal{K}_k(A^T A, A^T b)$ mainly contains and provides. Note that the eigenpairs of $A^T A$ are the squares of singular values and right singular vectors of A, and the tridiagonal matrix $B_k^T B_k$ is the projected matrix of $A^T A$ onto the subspace $\mathcal{K}_k(A^T A, A^T b)$, which is obtained by applying the symmetric Lanczos tridiagonalization process to $A^T A$ starting with $q_1 = A^T b / \|A^T b\|$ [6]. We have a general claim deduced from [6, 34] and exploited widely in [20, 22]: The more information the subspace $\mathcal{K}_k(A^T A, A^T b)$ contains on the k dominant right singular vectors, the more possible and accurate the k Ritz values approximate the k largest 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 is if it appears.

For our problem, since the small singular values of A are clustered and close to zero, it is expected that a small Ritz value will show up as k grows large, and it starts to appear more late when $\mathcal{K}_k(A^T A, A^T b)$ contains less information on the other $n - k$ right singular vectors. In this sense, we say that LSQR has better regularizing effects since $x^{(k)}$ contains more dominant SVD components.

Using the definition of canonical angles $\Theta(\mathcal{X}, \mathcal{Y})$ between the two subspaces $\mathcal{X}$ and $\mathcal{Y}$ of the same dimension [36, p. 250], we have the following theorem, which shows how well the subspace $\mathcal{K}_k(A^T A, A^T b)$, on which LSQR and CGLS work, captures the k-dimensional dominant right singular space.

THEOREM 2.1. *Let the SVD of A be* (1.2), *and assume that its singular values are distinct and satisfy* $\sigma_j = \mathcal{O}(e^{-\alpha j})$ *with* $\alpha > 0$. *Let* $\mathcal{V}_k = span\{V_k\}$ *be the subspace spanned by the columns of* $V_k = (v_1, v_2, \ldots, v_k)$, *and* $\mathcal{V}_k^s = \mathcal{K}_k(A^T A, A^T b)$. *Then*

$$(2.4) \qquad \| \sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s) \| = \frac{\|\Delta_k\|}{\sqrt{1 + \|\Delta_k\|^2}}$$

with the $(n - k) \times k$ *matrix* Δ_k *to be defined by* (2.6) *and*

$$(2.5) \quad \|\Delta_k\|_F \leq \frac{\sigma_{k+1}}{\sigma_k} \frac{\max_{j=k+1}^{n} |u_j^T b|}{\min_{j=1}^{k} |u_j^T b|} \sqrt{k(n-k)}(1 + \mathcal{O}(e^{-2\alpha})), \quad k = 1, 2, \ldots, n - 1.$$

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 $\mathcal{K}_k(\Sigma^2, \Sigma \bar{U}^T b)$ is spanned by the columns of the $n \times k$ matrix $D T_k$ with

$$D = \mathrm{diag}\left(\sigma_i \bar{U}^T b\right), \quad T_k = \begin{pmatrix} 1 & \sigma_1^2 & \cdots & \sigma_1^{2k-2} \\ 1 & \sigma_2^2 & \cdots & \sigma_2^{2k-2} \\ \vdots & \vdots & & \vdots \\ 1 & \sigma_n^2 & \cdots & \sigma_n^{2k-2} \end{pmatrix}.$$

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 σ_j distinct for $1 \leq j \leq k$, it is nonsingular. Thus, by the SVD of A, we have

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

with

$$(2.6) \qquad \Delta_k = D_2 T_{k2} T_{k1}^{-1} D_1^{-1}.$$

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 $\mathcal{V}_k^s$.

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

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

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

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

which proves (2.4) and indicates that $\|\sin\Theta(\mathcal{V}_k,\mathcal{V}_k^s)\|$ is monotonically increasing with respect to $\|\Delta_k\|$.

We next estimate $\|\Delta_k\|$. We have

$$\|\Delta_k\| \le \|\Delta\|_F = \left\|D_2 T_{k2} T_{k1}^{-1} D_1^{-1}\right\|_F \le \|D_2\| \left\|T_{k2} T_{k1}^{-1}\right\|_F \|D_1^{-1}\|$$

$$(2.7) \qquad = \frac{\sigma_{k+1}}{\sigma_k} \frac{\max_{j=k+1}^n |v_j^T b|}{\min_{j=1}^k |v_j^T b|} \left\|T_{k2} T_{k1}^{-1}\right\|_F .$$

So we need to estimate $\left\|T_{k2} T_{k1}^{-1}\right\|_F$. It is easily justified that the i-th column of T_{k1}^{-1} consists of the coefficients of the Lagrange polynomial

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

that interpolates the elements of the i-th canonical basis vector $e_i^{(k)} \in \mathbb{R}^k$ at the abscissas $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$. 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

$$(2.8) \qquad T_{k2} T_{k1}^{-1} = \begin{pmatrix} L_1^{(k)}(\sigma_{k+1}^2) & L_2^{(k)}(\sigma_{k+1}^2) & \cdots & L_k^{(k)}(\sigma_{k+1}^2) \\ L_1^{(k)}(\sigma_{k+2}^2) & L_2^{(k)}(\sigma_{k+2}^2) & \cdots & L_k^{(k)}(\sigma_{k+2}^2) \\ \vdots & \vdots & & \vdots \\ L_1^{(k)}(\sigma_n^2) & L_2^{(k)}(\sigma_n^2) & \cdots & L_k^{(k)}(\sigma_n^2) \end{pmatrix}.$$

Since $|L_i^{(k)}(\lambda)|$ is monotonic for $\lambda < \sigma_k^2$, it is bounded by $|L_i^{(k)}(0)|$. Furthermore, let $|L_{i_0}^{(k)}(0)| = \max_{i=1,2,\ldots,k} |L_i^{(k)}(0)|$. Then for $i = 1, 2, \ldots, k$ and $\alpha > 0$ we have

$$|L_i^{(k)}(0)| \le |L_{i_0}^{(k)}(0)| = \prod_{j=1, j\ne i_0}^k \left|\frac{\sigma_j^2}{\sigma_j^2 - \sigma_{i_0}^2}\right| = \prod_{j=1}^{i_0-1} \frac{\sigma_j^2}{\sigma_j^2 - \sigma_{i_0}^2} \cdot \prod_{j=i_0+1}^k \frac{\sigma_j^2}{\sigma_{i_0}^2 - \sigma_j^2}$$

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

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

$$(2.9) \qquad = \frac{\left(1 + \displaystyle\sum_{j=1}^{i_0} \mathcal{O}(e^{-2j\alpha})\right)\left(1 + \displaystyle\sum_{j=1}^{k-i_0+1} \mathcal{O}(e^{-2j\alpha})\right)}{\displaystyle\prod_{j=i_0+1}^k \mathcal{O}(e^{2(j-i_0)\alpha})}$$

by absorbing those higher order terms into $\mathcal{O}(\cdot)$. Note that in the above numerator we have

$$1 + \sum_{j=1}^{i_0} \mathcal{O}(e^{-2j\alpha}) = 1 + \mathcal{O}\left(\sum_{j=1}^{i_0} e^{-2j\alpha}\right) = 1 + \mathcal{O}\left(\frac{e^{-2\alpha}}{1 - e^{-2\alpha}}(1 - e^{-2i_0\alpha})\right),$$

and

$$1 + \sum_{j=1}^{k-i_0+1} \mathcal{O}(e^{-2j\alpha}) = 1 + \mathcal{O}\left(\sum_{j=1}^{k-i_0+1} e^{-2j\alpha}\right) = 1 + \mathcal{O}\left(\frac{e^{-2\alpha}}{1 - e^{-2\alpha}}(1 - e^{-2(k-i_0+1)\alpha})\right).$$

It is then easily seen that their product is

$$1 + \mathcal{O}\left(\frac{2e^{-2\alpha}}{1 - e^{-2\alpha}}\right) + \mathcal{O}\left(\left(\frac{e^{-2\alpha}}{1 - e^{-2\alpha}}\right)^2\right) = 1 + \mathcal{O}\left(\frac{2e^{-2\alpha}}{1 - e^{-2\alpha}}\right) = 1 + \mathcal{O}(e^{-2\alpha}).$$

On the other hand, by definition, the denominator $\prod_{j=i_0+1}^{k}\left(\frac{\sigma_{i_0}}{\sigma_j}\right)^2 = \prod_{j=i_0+1}^{k} \mathcal{O}(e^{2(j-i_0)\alpha})$ in (2.9) is exactly one for $i_0 = k$, and it is strictly bigger than one for $i_0 < k$. Therefore, for any k, we have $|L_{i_0}^{(k)}(0)| = \max_{i=1,2,\ldots,k} |L_i^{(k)}(0)| = 1 + \mathcal{O}(e^{-2\alpha})$. From this and (2.8) it follows that

$$\left\|T_{k2}T_{k1}^{-1}\right\|_F \leq \sqrt{k}\left\|T_{k2}T_{k1}^{-1}e_k^{(k)}\right\| \leq \sqrt{k(n-k)}|L_{i_0}^{(k)}(0)| = \sqrt{k(n-k)}(1 + \mathcal{O}(e^{-2\alpha})).$$

Therefore, for $i = 1, 2, \ldots, n-1$ and $\alpha > 0$ considerably, from (2.7) we have

$$(2.10) \qquad \|\Delta_k\|_F \leq \frac{\sigma_{k+1}}{\sigma_k}\frac{\max_{j=k+1}^n |u_j^T b|}{\min_{j=1}^k |u_j^T b|}\sqrt{k(n-k)}(1 + \mathcal{O}(e^{-2\alpha})). \qquad \square$$

Remark 2.1 We point out that (2.5) should not be sharp. As we have seen from the proof, the factor $\frac{\sigma_{k+1}}{\sigma_k}\frac{\max_{j=k+1}^n |u_j^T b|}{\min_{j=1}^k |u_j^T b|}$ seems intrinsic and unavoidable, but the factor $\sqrt{k(n-k)}$ in (2.5) is an overestimate and can certainly be reduced. (2.10) is an overestimate since $|L_i^{(k)}(0)|$ for i not near to k is considerably smaller than $|L_{i_0}^{(k)}(0|$, but we replace all them by their maximum $1 + \mathcal{O}(e^{-2\alpha})$. In fact, our derivation clearly illustrates that the smaller i is, the smaller $|L_i^{(k)}(0)|$ than $|L_k^{(k)}(0)|$.

Recall the discrete Picard condition (1.4). Then

$$(2.11) \qquad c_k = \frac{\max_{j=k+1}^n |u_j^T b|}{\min_{j=1}^k |u_j^T b|} = \frac{\max_{j=k+1}^n(|u_j^T \hat{b} + u_j^T e|)}{\min_{j=1}^k(|u_j^T \hat{b} + u_j^T e|)} \approx \frac{\sigma_{k+1}^{1+\beta} + |u_{k+1}^T e|}{\sigma_k^{1+\beta} + |u_k^T e|}.$$

We observe that $c_k \approx \frac{\sigma_{k+1}^{1+\beta}}{\sigma_k^{1+\beta}} < 1$ almost remains constant for $k \leq k_0$. For $k > k_0$, note that all the $|u_k^T b| \approx |u_k^T e|$ almost remain the same. Thus, we have $c_k \approx 1$, meaning that $\mathcal{V}_k^s$ does not capture $\mathcal{V}_k$ as well as it does for $k \leq k_0$.

Remark 2.2 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 2.1, we can obtain by the first order Taylor expansion

$$|L_{i_0}^{(k)}(0)| \approx |L_k^{(k)}(0)| = \prod_{j=1}^{k-1} \frac{\sigma_j^2}{\sigma_j^2 - \sigma_k^2}$$

$$= \prod_{j=1}^{k-1} \frac{1}{1 - \mathcal{O}((\frac{j}{k})^{2\alpha})} \approx 1 + \sum_{j=1}^{k-1} \mathcal{O}\left(\left(\frac{j}{k}\right)^{2\alpha}\right) = \mathcal{O}(1),$$

which, unlike $(1+\mathcal{O}(e^{-2\alpha}))$ for severely ill-posed problems, depends on k and increases slowly with k for $\alpha > 1$ considerably. However, for mildly ill-posed problems, from above we have $|L_{i_0}^{(k)}(0)| > 1$ considerably for $\alpha < 1$.

Remark 2.3 A combination of (2.4) and (2.5) and the above analysis indicate that $\mathcal{V}_k^s$ captures $\mathcal{V}_k$ better for severely ill-posed problems than for moderately ill-posed problems. There are two reasons for this. The first is that the factors σ_{k+1}/σ_k are basically fixed constants for severely ill-posed problems as k increases, and they are smaller than the counterparts for moderately ill-posed problems unless the degree α of its ill-posedness is far bigger than one and k small. The second is that the factor $\mathcal{O}(1)$ is smaller for severely ill-posed problems than the factor $1 + \mathcal{O}(e^{-2\alpha})$ for moderately ill-posed problems for the same k.

Remark 2.4 The situation is fundamentally different for mildly ill-posed problems: Firstly, we always have $|L_{i_0}^{(k)}(0)| > 1$ substantially for $\alpha \leq 1$ and any k, which is considerably bigger than $\mathcal{O}(1)$ for moderately ill-posed problems for the same k. Secondly, c_k defined by (2.11) is closer to one than that for moderately ill-posed problems for $k = 1, 2, \ldots, k_0$. Thirdly, for the same noise level $\|e\|$ and β, we see from the discrete Picard condition (1.4) and the definition of k_0 that k_0 is bigger for a mildly ill-posed problem than that for a moderately ill-posed problem. All of them show that $\mathcal{V}_k^s$ captures $\mathcal{V}_k$ *considerably better* for severely and moderately ill-posed problems than for mildly ill-posed problems for $k = 1, 2, \ldots, k_0$. In other words, our results illustrate that $\mathcal{V}_k^s$ contains more information on the other $n - k$ right singular vectors for mildly ill-posed problems, compared with severely and moderately ill-posed problems. The bigger k, the more it contains. Therefore, $\mathcal{V}_k^s$ captures $\mathcal{V}_k$ more effectively for severely and moderately ill-posed problems than mildly ill-posed problems. That is, $\mathcal{V}_k^s$ contains 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 before $k \leq k_0$ and LSQR has better regularizing effects 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 smaller than σ_{k_0+1} for $k \leq k_0$, 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 2.5 Relation (2.5) and c_k indicate that $\mathcal{V}_k^s$ captures $\mathcal{V}_k$ better for severely ill-posed problems than for moderately ill-posed problems. There are two reasons for this. First, the all the σ_{k+1}/σ_k are basically a fixed constant ρ^{-1} for severely ill-posed problems, which is smaller than those ratios for moderately ill-posed problems unless α is rather big and k small. Second, the quantities $|L_{i_0}^{(k)}(0)| = 1 + \mathcal{O}(e^{-2\alpha})$ for severely ill-posed problems are smaller than the corresponding $\mathcal{O}(1)$ for moderately ill-posed problems.

Let us investigate more and get insight into the regularization of LSQR. Define

$$(2.12) \qquad \gamma_k = \left\| A - P_{k+1} B_k Q_k^T \right\|,$$

which measures the quality of the rank k approximation $P_{k+1} B_k Q_k^T$ to A. Based on (2.5), we can derive the following estimate for γ_k.

THEOREM 2.2. *Assume that* (1.1) *is severely or moderately ill posed. Then*

$$(2.13) \qquad \sigma_{k+1} \leq \gamma_k \leq \sigma_{k+1} + \sigma_1 \| \sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s) \|.$$

Proof. Let $A_k = U_k \Sigma_k V_k^T$ 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 $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$.

Since $P_{k+1}B_kQ_k^T$ is of rank k, the lower bound in (2.13) 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

$$\left\|A - P_{k+1}B_kQ_k^T\right\| = \left\|A - AQ_kQ_k^T\right\|.$$

It is easily known that $\mathcal{V}_k^s = \mathcal{K}_k(A^TA, A^Tb) = span\{Q_k\}$ with Q_k having orthonormal columns. Then by the definition of $\|\sin\Theta(\mathcal{V}_k, \mathcal{V}_k^s)\|$ we obtain

$$\begin{aligned}
\left\|A - AQ_kQ_k^T\right\| &= \left\|(A - U_k\Sigma_kV_k^T + U_k\Sigma_kV_k^T)(I - Q_kQ_k^T)\right\| \\
&\leq \left\|(A - U_k\Sigma_kV_k^T)(I - Q_kQ_k^T)\right\| + \left\|U_k\Sigma_kV_k^T(I - Q_kQ_k^T)\right\| \\
&\leq \sigma_{k+1} + \|\Sigma_k\|\left\|V_k^T(I - Q_kQ_k^T)\right\| \\
&= \sigma_{k+1} + \sigma_1\|\sin\Theta(\mathcal{V}_k, \mathcal{V}_k^s)\|. \quad \square
\end{aligned}$$

Numerically, it has been extensively observed in the literature that the γ_k decay as fast as σ_{k+1} and, more precisely, $\gamma_k \approx \sigma_{k+1}$ for severely ill-posed problems; see, e.g., [3, 16, 17]. They mean that the $P_{k+1}B_kQ_k^T$ are very good rank k approximations to A. Recall that the TSVD method generates the best regularized solution $x_{k_0}^{TSVD} = A_{k_0}^\dagger b$. As a result, if $\gamma_{k_0} \approx \sigma_{k_0+1}$, the LSQR iterate $x^{(k_0)} = Q_{k_0}T_{k_0}^\dagger Q_{k_0+1}^T b$ is reasonably close to the TSVD solution $x_{k_0}^{TSVD}$ for σ_{k_0+1} is reasonably small. This means that LSQR has the full regularization and does not need any additional regularization to improve $x^{(k_0)}$. 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 strong theoretical supports. Compared to the observations, our (2.13) appears to be a considerable overestimate.

We next present some results on α_{k+1} appearing in (2.2). If $\alpha_{k+1} = 0$, the Lanczos bidiagonalization process terminates, and we have found k exact singular triples of A [26]. 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 α_{k+1} is possible. We aim to investigate how fast α_{k+1} decays. We first give a refinement of a result in [17].

THEOREM 2.3. *Let* $B_k = W_k\Theta_kS_k^T$ *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* $\tilde{U}_k = P_{k+1}W_k$ *and* $\tilde{V}_k = Q_kS_k$. *Then*

$$(2.14) \qquad\qquad\qquad A\tilde{V}_k - \tilde{U}_k\Theta_k = 0,$$

$$(2.15) \qquad\qquad\qquad \left\|A^T\tilde{U}_k - \tilde{V}_k\Theta_k^T\right\| = \alpha_{k+1}.$$

Proof. From (2.1) and $B_k = W_k\Theta_kS_k^T$, we obtain

$$A\tilde{V}_k = AQ_kS_k = P_{k+1}B_kS_k = \tilde{U}_k\Theta_k.$$

So (2.14) holds. From (2.2), we get

$$\begin{aligned}
A^T\tilde{U}_k &= A^TP_{k+1}W_k \\
&= Q_kB_k^TW_k + \alpha_{k+1}q_{k+1}e_{k+1}^TW_k \\
&= Q_kS_k\Theta_k^T + \alpha_{k+1}q_{k+1}e_{k+1}^TW_k \\
&= \tilde{V}_k\Theta_k^T + \alpha_{k+1}q_{k+1}e_{k+1}^TW_k.
\end{aligned}$$

Note that $\|q_{k+1}\| = \|e_{k+1}^T W_k\| = 1$. Then we get

$$(2.16) \qquad \left\| A^T \tilde{U}_k - \tilde{V}_k \Theta_k^T \right\| = \alpha_{k+1} \left\| q_{k+1} e_{k+1}^T W_k \right\| = \alpha_{k+1}. \qquad \Box$$

We remark that it is an inequality other than the equality in a result of [17] similar to (2.15).

In combination with the previous results and remarks, this theorem shows that once α_{k+1} becomes small for not big k, the k singular values of B_k may approximate the large singular values of A, and it is more possible that no small one appears for severely ill-posed problems and moderately ill-posed problems.

As our final result, we establish an intimate and interesting relationship between α_{k+1} and γ_k, showing how fast α_{k+1} decays.

THEOREM 2.4. *It holds that*

$$(2.17) \qquad \alpha_{k+1} \le \gamma_k.$$

Proof. With the notations as in Theorem 2.3, we have $P_{k+1} B_k Q_k^T = \tilde{U}_k \Theta_k \tilde{V}_k^T$. So, by (2.12), we have

$$\gamma_k = \left\| A - \tilde{U}_k \Theta_k \tilde{V}_k^T \right\|.$$

Note that $\tilde{U}_k^T \tilde{U}_k = I$. Therefore, from (2.16) we obtain

$$\begin{aligned}
\alpha_{k+1} &= \left\| A^T \tilde{U}_k - \tilde{V}_k \Theta_k^T \right\| \\
&= \left\| A^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k^T \right\| \\
&= \left\| A^T \tilde{U}_k \tilde{U}_k^T - \tilde{V}_k \Theta_k^T \tilde{U}_k^T \tilde{U}_k \tilde{U}_k^T \right\| \\
&= \left\| \left(A^T - \tilde{V}_k \Theta_k^T \tilde{U}_k^T \right) \tilde{U}_k \tilde{U}_k^T \right\| \\
&\le \left\| \left(A^T - \tilde{V}_k \Theta_k^T \tilde{U}_k^T \right) \right\| \left\| \tilde{U}_k \tilde{U}_k^T \right\| \\
&= \left\| A - \tilde{U}_k \Theta_k \tilde{V}_k^T \right\| = \gamma_k. \qquad \Box
\end{aligned}$$

The theorem indicates that α_{k+1} decays at least as fast as γ_k, which, in turn, means that α_{k+1} may decrease in the same rate as σ_{k+1}, as observed in [3, 16, 17] for severely ill-posed problems.

3. Numerical experiments. In this section, we report numerical experiments to illustrate the 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 [21]. All the problems arise from the discretization of the first kind Fredholm integral equation

$$(3.1) \qquad \int_a^b K(s,t)x(t)dt = b(s), \quad c \le s \le d.$$

For each problem we use the codes of [21] 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. 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. We remind that, as far as ill-posed problem (1.1) is concerned, our primary goal consists in justifying the regularizing effects of iterative solvers, which are *unaffected by sizes* of ill-posed problems and only depends on the degree of ill-posedness. Therefore, for this purpose, as extensively done in the literature (see, e.g., [20, 22] and the references therein), it is enough to test not very large problems. Indeed, for n large, say, 1,0000 and more, we have observed completely the same behavior as that for n not large, e.g., $n = 1024$ used in this paper. A reason for using n not large is because such choice makes it practical to fully justify the regularization effects of LSQR by comparing it with the TSVD method, which suits only for small and/or medium sized problems for computational efficiency. All the computations are carried out in Matlab 7.8 with the machine precision $\epsilon_{\mathrm{mach}} = 2.22 \times 10^{-16}$ under the Microsoft Windows 7 64-bit system.

3.1. Severely ill-posed problems. We consider the following two severely ill-posed problems [21].

Example 1 This problem 'Shaw' arises from one-dimensional image restoration, and can be obtained by discretizing the first kind Fredholm integral equation (3.1) with $[-\frac{\pi}{2}, \frac{\pi}{2}]$ as both integration and domain 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 (3.1) with $[0, 1]$ as both integration and domain intervals. The kernel $K(s, t)$, the solution $x(t)$ and the right-hand side $b(s)$ are given by

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

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

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

In Figure 1, we display the curves of the sequences γ_k and α_{k+1} with $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$, respectively. They illustrate that the quantities γ_k decrease as fast as σ_{k+1} and both of them level off at the level of ϵ_{mach} for k no more than 20, and after that these quantities are purely round-offs and are reliable no more. Moreover, the curves of quantities α_{k+1} always lie below those of γ_k, which coincides with Theorem 2.4. We can see that the decaying curves with different noise levels are almost
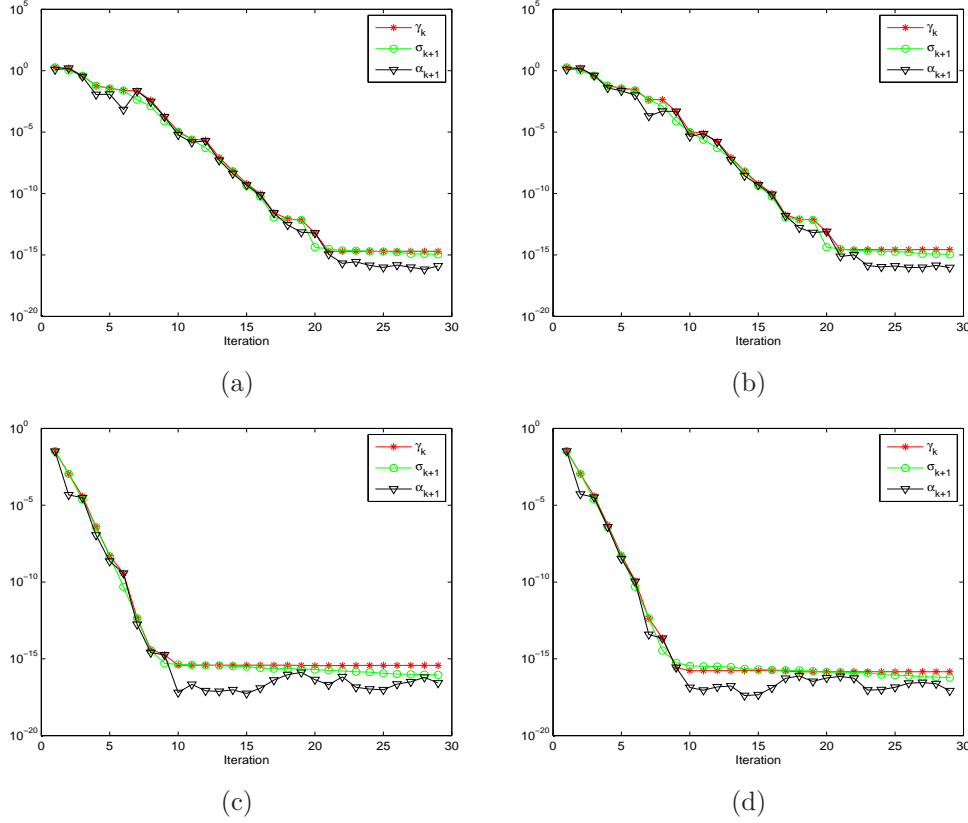
FIG. 1. *(a)-(b): Plots of decaying behavior of the sequences γ_k, σ_{k+1} and α_{k+1} for the problem Shaw with $\varepsilon = 10^{-2}$ (left) and $\varepsilon = 10^{-3}$ (right); (c)-(d): Plots of decaying behavior of the sequences γ_k and σ_{k+1} for the problem Wing with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right).*

the same. 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 good rank k approximations to A with the approximate accuracy σ_{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_{true} until iteration k_0, and they are deteriorated after that iteration. At iteration k_0, $x^{(k_0)}$ only captures the k_0 dominant SVD components of A and suppress the other $(n-k_0)$ SVD components, so that it is a best possible regularized solution. As a result, the pure LSQR has the full regularization for severely ill-posed problems. We will give a more direct justification on these assertions in Section 3.3.

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

3.2. Moderately ill-posed problems. We now consider the following two moderately ill-posed problems [21].

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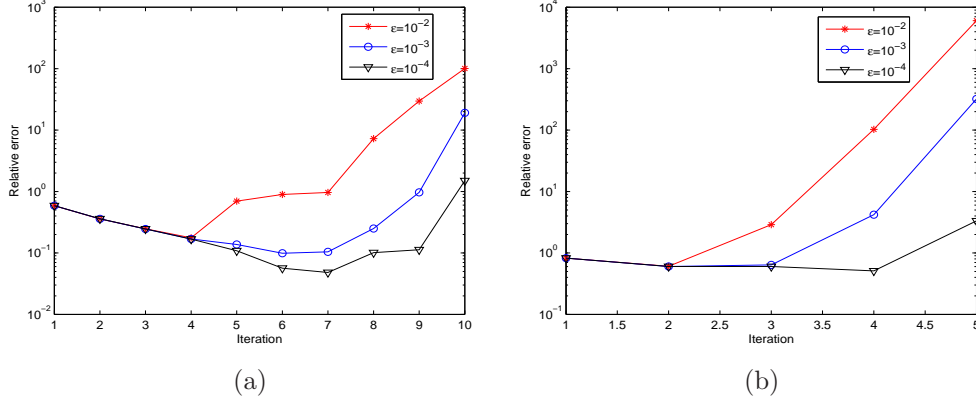
FIG. 2. *The relative errors $\left\| x^{(k)} - x_{true} \right\| / \| x_{true} \|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for the problems Shaw (left) and Wing (right).*

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

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

Example 4   This problem is the famous Phillips' test problem. It can be obtained by discretizing the first kind Fredholm integral equation (3.1) with $[-6,6]$ as both integration and domain intervals. The kernel $K(s,t)$, the solution $x(t)$ and the right-hand side $b(s)$ are given by

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

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

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

From Figure 3, we see that γ_k decreases as fast as σ_{k+1}, and α_{k+1} decays as fast as γ_k. However, slightly different from severely ill-posed problems, we can observe that the γ_k may not be so close to the σ_{k+1}, as reflected by the thick rope formed by three lines. By comparing the behavior of γ_k for severely and moderately ill-posed problem, we come to the conclusion that the k-step Lanczos bidiagonalization may generate more accurate rank k approximation for severely ill-posed problems than for moderately ill-posed problems, namely, the rank k approximation $P_{k+1} B_k Q_k^T$ may be more accurate for severely ill-posed problems than for moderately ill-posed problems. Nonetheless, we have seen that, for the test moderately ill-posed problems, all the γ_k are still excellent approximations to the σ_{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.
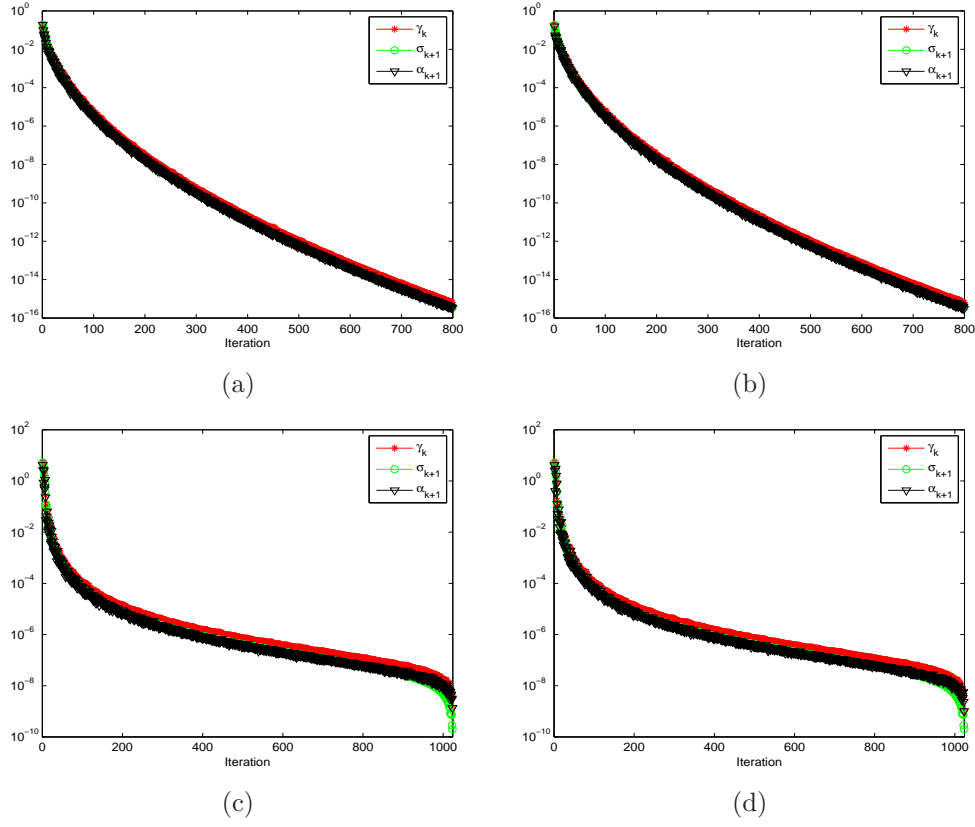
(a)

(b)

(c)

(d)

FIG. 3. *(a)-(b): Plots of decaying behavior of the sequences γ_k, σ_{k+1} and α_{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 γ_k and σ_{k+1} for the problem Phillips with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right).*
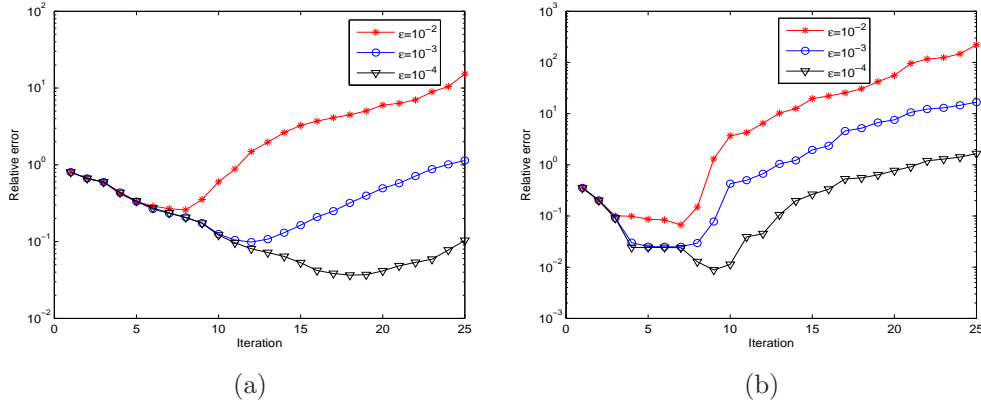

(a)

(b)

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

3.3. 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. We show that LSQR has the full regularization and no additional regularization is needed, which is based on the observation that at semi-convergence the regularized solution by LSQR is as accurate as that obtained by the hybrid LSQR for each problem.

In the sequel, we only report the results for the noise level $\varepsilon = 10^{-3}$. Results for other ε 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 simply stabilizes the regularized solutions with the minimum error. This means that the pure LSQR itself has already found a best possible regularized solution at semi-convergence and no additional regularization is needed. So it has the full regularization. Our 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 $x_{reg} = \arg\min_k \|x^{(k)} - x_{true}\|$ for the pure LSQR. Figure 5 (c) and Figures 6 (c)-(d) show that the regularized solutions are generally very 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 fact that the regularization of LSQR and its hybrid variants is unsuitable for the ill-posed problems with discontinuous solutions. For such kind of problems, more reasonable regularization is Total Variation Regularization, which takes the form $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \lambda^2 \|Lx\|_1^2$ with $L \neq I$ some $p \times n$ matrix and $\|\cdot\|_1$ the 1-norm [22].

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 to improve the regularized solution by LSQR at semi-convergence.

Example 5 The problem 'deriv2' is mildly ill-posed, which is obtained by discretizing the first kind Fredholm integral equation (3.1) with $[0, 1]$ as both integration and domain 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. As we have seen, 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 $\mathcal{K}_6(A^T A, A^T b)$ to construct
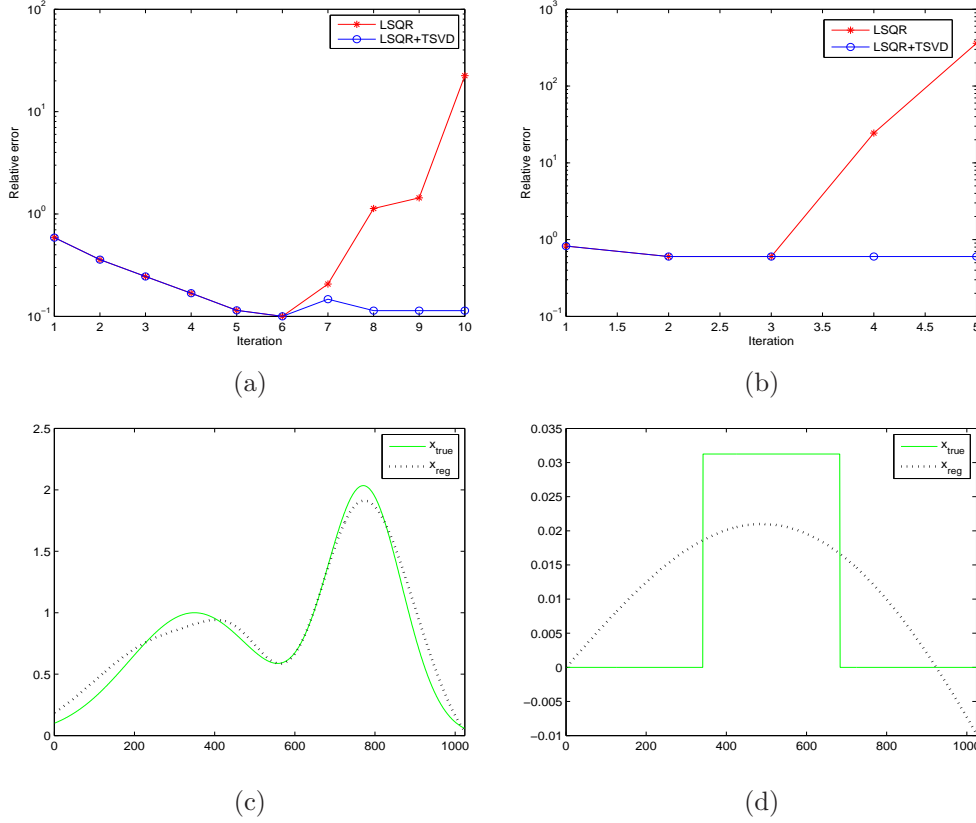
FIG. 5. *(a)-(b): The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to LSQR and LSQR with additional TSVD regularization for $\varepsilon = 10^{-3}$; (c)-(d): The regularized solutions x_{reg} for the pure LSQR for the problems Shaw (left) and Wing (right).*

a best possible regularized solution. We also choose $x_{reg} = \arg\min_k \|x^{(k)} - x_{true}\|$ 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}.

4. Conclusions. For large-scale discrete ill-posed problems, LSQR and CGLS are commonly used methods. These methods have regularizing effects and exhibit semi-convergence. However, if a small Ritz value appears before the methods capture all the needed dominant SVD components, the methods have 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 additional regularization needed.

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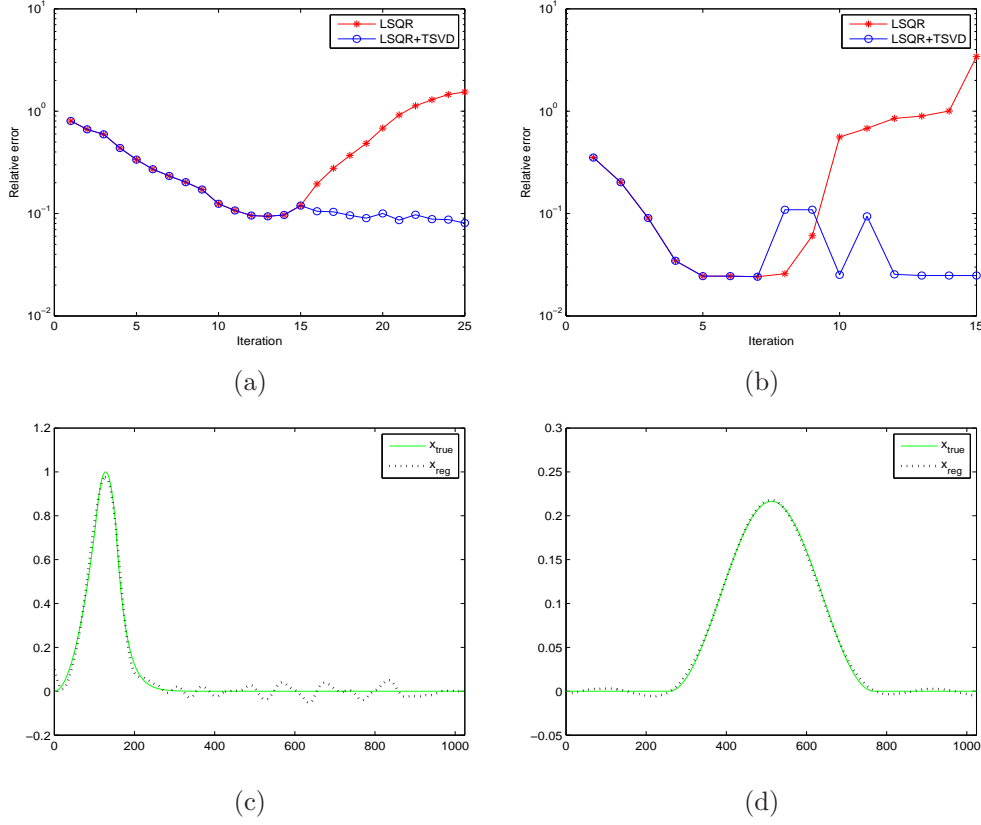
FIG. 6. *(a)-(b): The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ obtained by the pure LSQR and LSQR with the additional TSVD regularization for $\varepsilon = 10^{-3}$; (c)-(d): The regularized solutions x_{reg} for the pure LSQR for the problems Heat (left) and Phillips (right).*
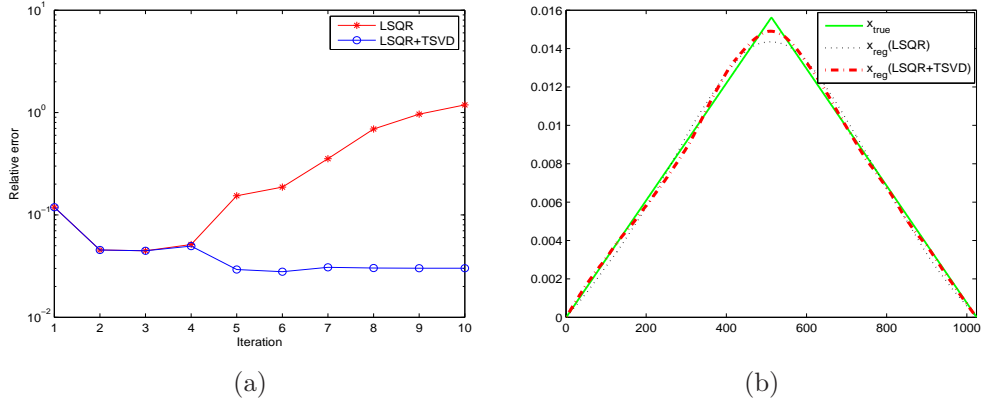

FIG. 7. *The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ and the regularized solution x_{reg} with respect to LSQR and LSQR with the additional TSVD regularization for the problem Deriv2 and $\varepsilon = 10^{-3}$.*

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 problems, compatible with our assertion. Together with the observations [3, 16, 17], it appears that the excellent performances of LSQR on severely and moderately ill-posed problems generally hold.

As for future work, it is more appealing to derive an accurate estimate for $\|\Delta_k\|$ other than $\|\Delta_k\|_F$, as it plays a crucial role in analyzing the accuracy γ_k of the rank k approximation, generated by Lanczos bidiagonalization, to A. Accurate bounds for γ_k are the core of completely understanding the regularizing effects of LSQR, but our bound (2.12) for γ_k is conservative and is expected to be improved on substantially. 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 more research is needed.

REFERENCES

[1] F. Bauer and M. A. Lukas, *Comparing parameter choice for regularization of ill-posed problems.* Math. Comput. Simul., 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] E. H. Bergou, S Gratton, and J. Tshimanga, *The exact condition number of the truncated singular value solution of a linear ill-posed problem.* SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1073–1085.
[5] Å. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations.* BIT Numer. Math., 28 (1988), pp. 659–670.
[6] Å. Björck, *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, PA, 1996.
[7] D. Calvetti, G. H. Golub and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization.* BIT Numer. Math., 39 (1999), pp. 603–619.
[8] D. Calvetti, P. C. Hansen and L. Reichel, *L-curve curvature bounds via Lanczos bidiagonalization.* Electron. Trans. Numer. Anal., 14 (2002), pp. 20–35.
[9] 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.
[10] D. Calvetti, B. Lewis and L. Reichel, *GMRES, L-curve, and discrete ill-posed problems.* BIT Numer. Math., 42 (2002), pp. 44–65.
[11] D. Calvetti, S. Morigi, L. Reichel and F. Sgallari, *Tikhonov regularization and the L-curve for large discrete ill-posed problems.* J. Comput. Appl. Math., 123 (2000), pp. 423–446.
[12] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems.* BIT Numer. Math., 43 (2003), pp. 263–283.
[13] J. Chung, 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.
[14] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems.* Kluwer, Dordrecht, The Netherlands, 1996.
[15] S. Gazzola and P. Novati, *Multi-parameter Arnoldi-Tikhonov methods.* Electron. Trans. Numer. Anal., 40 (2013), pp. 452–475.
[16] S. Gazzola, *Regularization techniques based on Krylov methods for ill-posed linear systems.* Ph. D. thesis, Dept. of Mathematics, University of Padua, Italy, 2014.
[17] S. Gazzola, P. Novati and M. R. Russo, *On Krylov projection methods and Tikhonov regularization.* Electron. Trans. Numer. Anal., 44 (2015), pp. 83–123.
[18] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems.* Longman, Essex, 1995.

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

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

[21] P. C. HANSEN, *Regularization tools version 4.0 for Matlab 7.3*. Numer. Algor., 46 (2007), pp. 189–194.

[22] P. C. HANSEN, *Discrete Inverse Problems: Insight and Algorithms*. SIAM, Philadelphia, PA, 2010.

[23] 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.

[24] B. HOFMANN, *Regularization for Applied Inverse and Ill-posed Problems*. Teubner, Stuttgart, Germany, 1986.

[25] T. K. JENSEN AND P. C. HANSEN, *Iterative regularization with minimum-residual methods*. BIT Numer. Math., 47 (2007), pp. 103–120.

[26] 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.

[27] 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.

[28] B. LEWIS AND L. REICHEL, *Arnoldi-Tikhonov regularization methods*. J. Comput. Appl. Math., 226 (2009), pp. 92–102.

[29] Z. C. LI, H. T. HUANG AND Y. WEI, *Ill-conditioning of the truncated singular value decomposition, Tikhonov regularization and their applications to numerical partial differential equations*. Numer. Linear Algebra Appl., 18 (2011), pp. 205–221.

[30] A. NEUMAN, L. REICHEL AND H. SADOK, *Algorithms for range restricted iterative methods for linear dicrete ill-posed problems*. Numer. Algor., 59 (2012), pp 325–331.

[31] P. NOVATI AND M. R. RUSSO, *A GCV based Arnoldi-Tikhonov regularization method*. BIT Numer. Math., 54 (2014), pp. 501–521.

[32] 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.

[33] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: an algorithm for sparse linear equations and sparse least squares*. ACM Trans. Math. Soft., 8 (1982), pp. 43–71.

[34] B. N. PARLETT, *The Symmetric Eigenvalue Problem*. SIAM, Philedaphia, PA, 1998.

[35] L. REICHEL AND G. RODRIGUEZ, *Old and new parameter choice rules for discrete ill-posed problems*. Numer. Algor., 63 (2013), pp. 65–87.

[36] G. W. STEWART, *Matrix Algorithms. Volume II: Eigensystems*. SIAM, Philadelphia, PA, 2001.

[37] C. R. VOGEL, *Computational Methods for Inverse Problems*. Frontiers in Appl. Math., SIAM, Philadelphia, PA, 2002.