SIAM J. SCI. COMPUT.
Vol. 37, No. 5, pp. S562–S580

A HYBRID LSMR ALGORITHM FOR LARGE-SCALE TIKHONOV REGULARIZATION*

JULIANNE CHUNG† AND KATRINA PALMER‡

Abstract. We develop a hybrid iterative approach for computing solutions to large-scale ill-posed inverse problems via Tikhonov regularization. We consider a hybrid LSMR algorithm, where Tikhonov regularization is applied to the LSMR subproblem rather than the original problem. We show that, contrary to standard hybrid methods, hybrid LSMR iterates are not equivalent to LSMR iterates on the directly regularized Tikhonov problem. Instead, hybrid LSMR leads to a different Krylov subspace problem. We show that hybrid LSMR shares many of the benefits of standard hybrid methods such as avoiding semiconvergence behavior. In addition, since the regularization parameter can be estimated during the iterative process, it does not need to be estimated a priori, making this approach attractive for large-scale problems. We consider various methods for selecting regularization parameters and discuss stopping criteria for hybrid LSMR, and we present results from image processing.

Key words. Tikhonov regularization, iterative methods, hybrid regularization, ill-posed inverse problems

AMS subject classifications. 65F20, 65F22, 65F30

DOI. 10.1137/140975024

1. Introduction.

We consider large-scale inverse problems of the form

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{n}, \tag{1.1}$$

where $\mathbf{b} \in \mathbb{R}^{m\times 1}$ are observed data, $\mathbf{A} \in \mathbb{R}^{m\times n}(m \geq n)$ models the forward process, $\mathbf{x}_{\text{true}} \in \mathbb{R}^{n\times 1}$ represents the desired parameters, and $\mathbf{n} \in \mathbb{R}^{m\times 1}$ is additive noise. We assume that the errors are independent and identically distributed (i.i.d.) from a Gaussian distribution with zero mean and standard deviation σ. Given $\mathbf{A}$ and $\mathbf{b}$, the goal is to compute an approximation to $\mathbf{x}_{\text{true}}$.

Linear inverse problems such as (1.1) arise in various scientific applications such as biomedical imaging, geoscience, astronomy, and security; cf. [15, 24, 30]. These problems often arise from the discretization of integral equations. It is well known that problems like (1.1) are challenging because of their ill-posed nature, whereby small errors in the data amplify during the inversion process, causing large errors in the solution. Regularization is typically used to stabilize the inversion process and can take many forms.

Tikhonov regularization estimates $\mathbf{x}_{\text{true}}$ from least squares (LS) problems of the form

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{L}\mathbf{x}\|_2^2, \tag{1.2}$$

where the regularization parameter λ controls the smoothness of the solution, and the regularization matrix $\mathbf{L}$ is often selected to be the identity matrix, or a finite

*Received by the editors June 30, 2014; accepted for publication (in revised form) July 1, 2015; published electronically October 29, 2015.
http://www.siam.org/journals/sisc/37-5/97502.html
†Department of Mathematics, Virginia Tech, Blacksburg, VA 24061 (jmchung@vt.edu).
‡Department of Mathematical Sciences, Appalachian State University, Boone, NC 28608 (palmerk@appstate.edu).

difference approximation of the derivatives of $\mathbf{x}$. Here, we consider standard form Tikhonov with $\mathbf{L} = \mathbf{I}$, and mention that if $\mathbf{L}$ is nonsingular, then a transformation to standard form is possible [22]. Selecting a good regularization parameter can be difficult, especially for large-scale problems.

An alternative to Tikhonov regularization is *iterative regularization*, in which a standard iterative method such as CGLS or LSQR is applied to the LS problem

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \tag{1.3}$$

and terminated early. Let $\mathbf{x}_k$ be the kth approximate solution. It is well known that iterative methods exhibit semiconvergence on ill-posed problems, with reconstruction errors $\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2 / \|\mathbf{x}_{\text{true}}\|_2$ decreasing initially but at some point beginning to increase as the small singular values of $\mathbf{A}$ start to amplify noise. This phenomenon can be seen in Figure 1(a), where reconstruction errors are plotted for each iteration. For iterative regularization, selecting a good stopping iteration is a crucial and nontrivial task.

Hybrid methods have been proposed as a means to overcome semiconvergence behavior and to avoid selecting a regularization parameter a priori. The basic idea of a hybrid approach is to use an iterative method to project the original problem onto Krylov subspaces of increasing dimension. At each iteration, regularization is imposed on the projected problem and the regularization parameter can be selected adaptively. Previous work on hybrid methods based on LSQR and Tikhonov regularization include [3, 5, 7, 8, 9, 13, 20, 26, 27, 32, 36]. Here, we propose a new hybrid approach for solving large-scale ill-posed inverse problems that combines Tikhonov regularization with the newly developed iterative LS solver called LSMR.

Recently Fong and Saunders [17] proposed the LSMR algorithm that (like LSQR) uses the Golub–Kahan (GK) bidiagonalization process to generate a Krylov subspace. At each iteration, the algorithm seeks a solution in the projected subspace that minimizes $\|\mathbf{A}^\top \mathbf{r}_k\|_2$, where $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$, whereas previous methods such as LSQR seek a solution that minimizes $\|\mathbf{r}_k\|_2$. Although both LSQR and LSMR converge to the same solution, it was shown in [17, 40, 1] that LSMR can have several advantages over LSQR. For problems where the regularization parameter λ is fixed a priori, Fong and Saunders provide an efficient implementation of LSMR for solving the standard form Tikhonov problem.

In this work, we propose a hybrid approach that combines LSMR and Tikhonov regularization, with the regularization parameter λ being estimated at each iteration and applied to the LSMR subproblem. Theoretical investigations motivate the development and provide insight for the new hybrid approach. Contrary to previously studied hybrid methods, we show that hybrid LSMR is not equivalent to applying LSMR on the Tikhonov problem, but instead leads to a different Krylov subspace problem. We investigate various strategies for selecting regularization parameters and stopping criteria for hybrid LSMR.

The paper is organized as follows. In section 2, we provide some background on regularization via spectral filtering and describe some iterative methods based on the GK process. We also describe previous work on hybrid methods for large-scale ill-posed inverse problems, providing some context for our work. Section 3 introduces hybrid LSMR. We provide some theoretical insight for the LSMR algorithm applied to ill-posed problems and develop methods for selecting regularization parameters and stopping criteria for the new hybrid method. Numerical results are presented in section 4, and conclusions can be found in section 5.

2. Background. In this section, we begin with an overview of Tikhonov regularization and describe various parameter selection techniques. We discuss iterative methods based on the GK process and provide an introduction to hybrid methods.

2.1. Tikhonov regularization and parameter selection. Regularization is used to obtain sound approximations to solutions of ill-posed inverse problems by filtering out the components that amplify the noise corresponding to small singular values. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ denote the singular value decomposition (SVD) of $\mathbf{A}$, where $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a diagonal matrix containing the singular values of $\mathbf{A}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, and the columns $\mathbf{u}_i$ of $\mathbf{U}$ and $\mathbf{v}_i$ of $\mathbf{V}$ contain the left and right singular vectors, respectively.

The standard form Tikhonov problem is

$$\min_{\mathbf{x}} \left\| \begin{pmatrix} \mathbf{A} \\ \lambda \mathbf{I} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|_2^2, \tag{2.1}$$

whose solution can be written as $\mathbf{x}_\lambda = (\mathbf{A}^\top\mathbf{A} + \lambda^2\mathbf{I})^{-1}\mathbf{A}^\top\mathbf{b}$. Given the SVD of $\mathbf{A}$, the Tikhonov solution can be written as a filtered SVD solution,

$$\mathbf{x}_\lambda = \sum_{i=1}^{n} \phi_i \frac{\mathbf{u}_i^\top \mathbf{b}}{\sigma_i} \mathbf{v}_i \tag{2.2}$$

$$= \mathbf{V}\mathbf{\Phi}\mathbf{\Sigma}^\dagger\mathbf{U}^\top\mathbf{b} \equiv \mathbf{A}_\lambda^\dagger\mathbf{b}, \tag{2.3}$$

where $\mathbf{\Phi} = \text{diag}(\phi_1, \ldots, \phi_n)$ and the filter factors are $\phi_i = \frac{\sigma_i^2}{\sigma_i^2+\lambda^2}$. The filter factors depend on the choice of λ and filter out components corresponding to small singular values.

Methods for selecting regularization parameters for Tikhonov regularization have been well studied in the literature [21]. We investigate three methods, namely, generalized cross validation (GCV), discrepancy principle (DP), and unbiased predictive risk estimator (UPRE). GCV is a predictive statistics-based method that does not require an estimate of the noise level. It assumes that a solution computed on a reduced set of data points should give a good estimate of the missing points [18]. The GCV function is given by

$$G_{\mathbf{A},\mathbf{b}}(\lambda) = \frac{n \left\| \mathbf{b} - \mathbf{A}\mathbf{x}_\lambda \right\|_2^2}{(\mathsf{trace}(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger))^2}, \tag{2.4}$$

and the desired regularization parameter λ is chosen to minimize (2.4).

On the other hand, DP and UPRE rely on a prior estimate of the noise level for the problem. For DP, λ is selected so that the residual norm, $\|\mathbf{A}\mathbf{x}_\lambda - \mathbf{b}\|_2^2$, is on the order of the noise in the data. That is,

$$\|\mathbf{A}\mathbf{x}_\lambda - \mathbf{b}\|_2^2 = \tau\epsilon, \tag{2.5}$$

where ϵ is an approximation of the expected value of $\|\mathbf{n}\|_2^2$ and τ is a user-defined "safety" parameter. For i.i.d. white Gaussian noise, $\epsilon \approx \sigma^2 n$ and, if not provided, noise variance σ^2 can be estimated from the data[1] [14, 25].

[1] The noise variance σ^2 should not be confused with the singular values σ_i.

For UPRE, the idea is to select the regularization parameter that minimizes the unbiased estimator for the expected value of the predictive risk [38]. That is, the UPRE regularization parameter minimizes the function

$$U_{\mathbf{A},\mathbf{b}}(\lambda) = \frac{1}{n}\left\|\mathbf{b} - \mathbf{A}\mathbf{x}_\lambda\right\|_2^2 + \frac{2\sigma^2}{n}\mathsf{trace}(\mathbf{A}\mathbf{A}_\lambda^\dagger) - \sigma^2\,. \tag{2.6}$$

If the SVD of $\mathbf{A}$ is available, functions (2.4)–(2.6) can be simplified for easy evaluation.

2.2. GK bidiagonalization. The GK bidiagonalization process is at the heart of several iterative algorithms for solving large, sparse LS systems and estimating singular values and vectors. Given a matrix $\mathbf{A}$ and vector $\mathbf{b}$, the GK process is an iterative procedure to transform matrix $(\mathbf{b}\ \ \mathbf{A})$ to upper-bidiagonal form $(\beta_1\mathbf{e}_1\ \ \mathbf{B}_k)$. With initialization $\beta_1 = \|\mathbf{b}\|_2$, $\mathbf{u}_1 = \mathbf{b}/\beta_1$, and $\alpha_1\mathbf{v}_1 = \mathbf{A}^\top\mathbf{u}_1$, the kth iteration of the GK process computes orthonormal vectors $\mathbf{u}_{k+1}$ and $\mathbf{v}_{k+1}$ such that

$$\begin{aligned}\beta_{k+1}\mathbf{u}_{k+1} &= \mathbf{A}\mathbf{v}_k - \alpha_k\mathbf{u}_k,\\ \alpha_{k+1}\mathbf{v}_{k+1} &= \mathbf{A}^\top\mathbf{u}_{k+1} - \beta_{k+1}\mathbf{v}_k.\end{aligned}$$

After k steps, we have matrices $\mathbf{V}_k = \begin{pmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_k\end{pmatrix} \in \mathbb{R}^{n\times k}$, $\mathbf{U}_k = \begin{pmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_k\end{pmatrix} \in \mathbb{R}^{m\times k}$, bidiagonal matrix

$$\mathbf{B}_k = \begin{pmatrix}\alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \\ & & & \beta_{k+1}\end{pmatrix} \in \mathbb{R}^{(k+1)\times k},$$

and $\mathbf{L}_{k+1} = \begin{pmatrix}\mathbf{B}_k & \alpha_{k+1}\mathbf{e}_{k+1}\end{pmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$ such that

$$\mathbf{A}\mathbf{V}_k = \mathbf{U}_{k+1}\mathbf{B}_k \quad \text{and} \quad \mathbf{A}^\top\mathbf{U}_{k+1} = \mathbf{V}_{k+1}\mathbf{L}_{k+1}^\top. \tag{2.7}$$

In LSQR, $\mathbf{y}_k$ is chosen at each iteration to minimize $\|\mathbf{r}_k\|_2$, and the solution is defined as $\mathbf{x}_k = \mathbf{V}_k\mathbf{y}_k$. From the relations above, the LS problem

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \tag{2.8}$$

can be approximated by the projected LS problem

$$\min_{\mathbf{x}\in R(\mathbf{V}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 = \min_{\mathbf{y}} \|\mathbf{B}_k\mathbf{y} - \beta_1\mathbf{e}_1\|_2\,. \tag{2.9}$$

An efficient implementation of LSQR is described in [33, 34].

It is known that LSQR is mathematically equivalent to the well-known CGLS method and can be interpreted as a minimum residual Krylov subspace method, where in exact arithmetic, $\mathbf{x}_k$ solves

$$\min_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A}^\top\mathbf{A},\mathbf{A}^\top\mathbf{b})} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2\,, \tag{2.10}$$

where $\mathcal{K}_k(\mathbf{A}^\top\mathbf{A}, \mathbf{A}^\top\mathbf{b}) = \text{span}\left\{\mathbf{A}^\top\mathbf{b}, (\mathbf{A}^\top\mathbf{A})\mathbf{A}^\top\mathbf{b}, \dots, (\mathbf{A}^\top\mathbf{A})^{k-1}\mathbf{A}^\top\mathbf{b}\right\}$ is a k-dimensional Krylov subspace [4].

For LSMR, $\mathbf{y}_k$ is chosen at each iteration to minimize $\left\|\mathbf{A}^\top \mathbf{r}_k\right\|_2$, and the solution is defined as $\mathbf{x}_k = \mathbf{V}_k \mathbf{y}_k$. That is, $\mathbf{y}_k$ is the solution to the subproblem

$$\begin{aligned} \min_{\mathbf{x}\in R(\mathbf{V}_k)} \left\|\mathbf{A}^\top(\mathbf{A}\mathbf{x}-\mathbf{b})\right\|_2 &= \min_{\mathbf{y}} \left\| \begin{pmatrix} \mathbf{B}_k^\top \mathbf{B}_k \\ \bar{\beta}_{k+1}\mathbf{e}_k^\top \end{pmatrix} \mathbf{y} - \bar{\beta}_1 \mathbf{e}_1 \right\|_2 \\ &\equiv \min_{\mathbf{y}} \left\| \hat{\mathbf{B}}_k \mathbf{y} - \bar{\beta}_1 \mathbf{e}_1 \right\|_2, \end{aligned} \tag{2.11}$$

where $\bar{\beta}_k \equiv \alpha_k \beta_k$. Again, relationships from the GK bidiagonalization process were used to simplify the terms. Efficient methods based on QR factorizations were investigated for solving (2.11) in [17], and there it was also proved that LSMR and LSQR converge to the solution.

It is worth mentioning that the LSMR solution can also be interpreted as a best approximation in a Krylov subspace, in that $\mathbf{x}_k$ solves

$$\min_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A}^\top\mathbf{A},\mathbf{A}^\top\mathbf{b})} \left\|\mathbf{A}^\top(\mathbf{A}\mathbf{x}-\mathbf{b})\right\|_2. \tag{2.12}$$

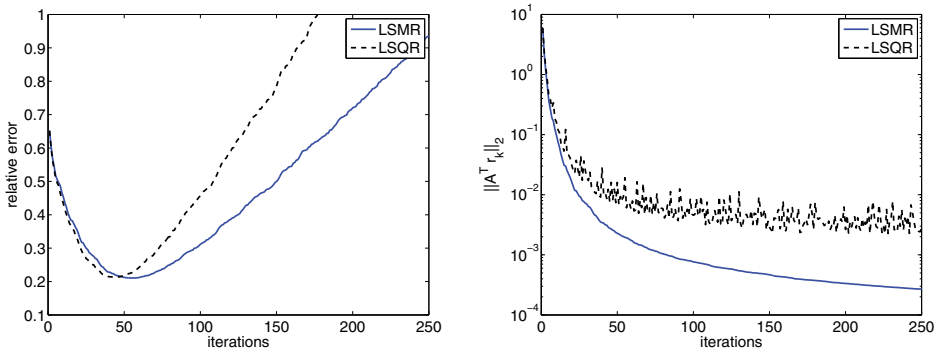
2.3. Hybrid methods. Hybrid methods [3, 32, 22] project the original LS problem (1.3) onto a small but growing subspace, using, for example, the GK process, and then use a direct regularization scheme such as Tikhonov regularization to solve the relatively small projected LS problem. Regularizing the projected problem at each iteration has the effect of stabilizing the convergence behavior, thereby reducing the risk of computing a poor solution with an imperfect stopping rule. Another benefit of hybrid methods is that the regularization parameter can be estimated during the iterative procedure rather than being required a priori. However, the additional cost of hybrid methods, compared to applying an iterative method such as LSQR or LSMR to (2.1) for fixed λ, is that the GK vectors $\mathbf{V}_k$ must be stored for computing the solution.

Previous hybrid methods have been studied in the context of minimizing the projected residual (2.9) and analyzing the singular values of $\mathbf{B}_k$ (also known as the Ritz values). For ill-posed problems where $\mathbf{A}$ is ill-conditioned, it is known that the singular values of $\mathbf{B}_k$ approximate singular values of $\mathbf{A}$ very well, depending on the relative spread of the singular values [19]. Thus, at early iterations, singular values of $\mathbf{B}_k$ converge quickly to the largest singular values of $\mathbf{A}$, resulting in good reconstructions. However, with more iterations, the singular values of $\mathbf{B}_k$ tend to approximate the largest and smallest singular values of $\mathbf{A}$, leading to an ill-conditioned matrix $\mathbf{B}_k$. Thus, it has been suggested to incorporate regularization such as TSVD or Tikhonov regularization to solve the projected subproblem [32, 3, 5, 20, 28]. At each iteration, rather than solving (2.9), we solve the Tikhonov problem

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{B}_k\mathbf{y} - \beta_1\mathbf{e}_1\|_2^2 + \lambda^2 \|\mathbf{y}\|_2^2, \tag{2.13}$$

where λ may be constant or may vary per iteration. We call this a *hybrid LSQR method*. There has been a lot of research on hybrid methods based on the GK bidiagonalization. For example, parameter selection methods within hybrid LSQR methods were studied in [27, 13, 2, 36], and hybrid LSQR methods for general form Tikhonov were considered in [26]. In [25], authors investigated how noise propagates to the projected problem and developed criteria based on the GK bidiagonalization for stopping the iterative method and for estimating the noise level in the original problem.

We propose a hybrid LSMR method in which Tikhonov regularization is used to solve the LSMR subproblem (2.11). Before giving details, we investigate properties of LSMR on ill-posed problems and motivate the need for a hybrid method.



(a) (b)

FIG. 1. *Comparison of LSMR and LSQR on an image deblurring problem where images are* 256×256. *Figure* (a) *provides relative errors,* $\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2 / \|\mathbf{x}_{\text{true}}\|_2$, *and Figure* (b) *presents values* $\left\|\mathbf{A}^\top \mathbf{r}_k\right\|_2$.

3. A hybrid LSMR algorithm. In this section we describe a new hybrid algorithm based on LSMR and Tikhonov regularization that is intended to solve large-scale problems where iterative methods are required. For example, iterative methods are desirable for problems where $\mathbf{A}$ may be extremely large and sparse, or for problems where $\mathbf{A}$ is not explicitly stored but can be accessed via function calls to compute matrix-vector products with $\mathbf{A}$ and $\mathbf{A}^\top$. In addition, similar to hybrid LSQR methods, we assume that solutions can be captured in relatively few iterations or that appropriate preconditioning can be used, so that k remains small and the GK vectors $\mathbf{V}_k$ can be stored. First we provide an example in section 3.1 that compares LSQR and LSMR for ill-posed problems. This illustration and the corresponding theory provide motivation and justification for the development of a hybrid LSMR approach. Then, in sections 3.2–3.4, we present the hybrid LSMR method, along with theoretical results and algorithmic details.

3.1. An illustration: LSMR for ill-posed problems. The goal of this section is to provide the reader with some insight into the performance of LSMR on ill-posed inverse problems and to motivate the need for hybrid LSMR. To do this, we consider a standard image deblurring example and compare the performance of LSMR versus LSQR, where no additional regularization (other than stopping criteria) is considered. Although specific details of the example are not included here, we remark that the behavior is representative of most ill-posed inverse problems.

In Figure 1(a), we provide relative reconstruction errors per iteration for LSQR and LSMR. Since we are dealing with ill-posed problems, both methods exhibit clear semiconvergence, as expected (cf. the introduction). However, note that LSMR exhibits delayed semiconvergence. That is, reconstruction errors for LSMR and LSQR initially decrease at similar rates, but after reaching an optimal value, LSMR errors increase at a slower rate. In Figure 1(b), values $\left\|\mathbf{A}^\top \mathbf{r}_k\right\|_2$ are presented for each algorithm. This figure resembles plots found in [17], where it is also observed that residual norm plots $\|\mathbf{r}_k\|_2$ for LSQR and LSMR are typically very similar.

To understand the difference between the LSQR and LSMR subproblems and to motivate the need for regularization on the LSMR subproblem, we analyze the singular values of $\hat{\mathbf{B}}_k = \begin{pmatrix} \mathbf{B}_k^\top \mathbf{B}_k \\ \bar{\beta}_{k+1} \mathbf{e}_k^\top \end{pmatrix}$ and compare them to the singular values of $\mathbf{B}_k$.

We make use of the following theorem from [6], originally studied in [37, 39].

THEOREM 3.1. *Let* $\hat{\mathbf{D}} = \mathbf{D} + \rho\mathbf{z}\mathbf{z}^\top$, *where* $\mathbf{D}$ *is diagonal and* $\|\mathbf{z}\|_2 = 1$. *Let* $d_1 \geq d_2 \geq \cdots \geq d_n$ *be the eigenvalues of* $\mathbf{D}$, *and let* $\hat{d}_1 \geq \hat{d}_2 \geq \cdots \geq \hat{d}_n$ *be the eigenvalues of* $\hat{\mathbf{D}}$. *Then*

$$\hat{d}_i = d_i + \rho\mu_i, \quad 1 \leq i \leq n,$$

where $\sum_{i=1}^n \mu_i = 1$ *and* $0 \leq \mu_i \leq 1$. *Moreover, if* $\rho > 0$, *then*

$$d_n \leq \hat{d}_n \leq \cdots \leq d_2 \leq \hat{d}_2 \leq d_1 \leq \hat{d}_1,$$

and if $\rho < 0$, *then*

$$\hat{d}_n \leq d_n \leq \cdots \leq \hat{d}_2 \leq d_2 \leq \hat{d}_1 \leq d_1.$$

Finally, if the d_i *are distinct and all of the elements of* $\mathbf{z}$ *are nonzero, then the eigenvalues of* $\hat{\mathbf{D}}$ *strictly separate those of* $\mathbf{D}$.

Consider the eigenvalues of symmetric matrices $\mathbf{B}_k^\top\mathbf{B}_k$ and $\hat{\mathbf{B}}_k^\top\hat{\mathbf{B}}_k$. Let

$$\mathbf{B}_k = \mathbf{P}\begin{pmatrix}\mathbf{S}\\ \mathbf{0}\end{pmatrix}\mathbf{Q}^\top$$

be the SVD of $\mathbf{B}_k$, where the singular values are $s_1 \geq s_2 \geq \cdots \geq s_k > 0$. Thus, the eigenvalues of $\mathbf{B}_k^\top\mathbf{B}_k$ are s_i^2, $i = 1, \ldots, k$. Matrix $\hat{\mathbf{B}}_k^\top\hat{\mathbf{B}}_k$ can be written as

$$\hat{\mathbf{B}}_k^\top\hat{\mathbf{B}}_k = \mathbf{B}_k^\top\mathbf{B}_k\mathbf{B}_k^\top\mathbf{B}_k + \bar{\beta}_{k+1}^2\mathbf{e}_k\mathbf{e}_k^\top \tag{3.1}$$

$$= \mathbf{Q}^\top(\mathbf{S}^2\mathbf{S}^2 + \bar{\beta}_{k+1}^2\mathbf{q}_k\mathbf{q}_k^\top)\mathbf{Q}, \tag{3.2}$$

where $\mathbf{q}_k$ is the kth column of $\mathbf{Q}$. Since similar matrices share the same eigenvalues and $\bar{\beta}_{k+1}^2 > 0$, by Theorem 3.1 we can conclude that

$$s_k^4 \leq \hat{s}_k^2 \leq \cdots \leq s_2^4 \leq \hat{s}_2^2 \leq s_1^4 \leq \hat{s}_1^2, \tag{3.3}$$

where $\hat{s}_i^2$ are the eigenvalues of $\hat{\mathbf{B}}_k^\top\hat{\mathbf{B}}_k$. Since $s_k^4 > 0$, we have the following interlacing property for the singular values of $\hat{\mathbf{B}}$ and the squares of the singular values of $\mathbf{B}_k$:

$$s_k^2 \leq \hat{s}_k \leq \cdots \leq s_2^2 \leq \hat{s}_2 \leq s_1^2 \leq \hat{s}_1. \tag{3.4}$$

Since the singular values of $\mathbf{B}_k$ tend to approximate the largest and smallest singular values of $\mathbf{A}$, (3.4) implies that the singular values of $\hat{\mathbf{B}}$ will approximate the *squares* of the largest and smallest singular values of $\mathbf{A}$. Thus, for ill-posed problems where many of the singular values are small, regularization may play a more crucial role. For the example above, we provide in Figure 2 the first 50 singular values of $\mathbf{A}$, along with the singular values of $\hat{\mathbf{B}}$ and $\mathbf{B}_k$, and the square of the singular values of $\mathbf{B}_k$ for iteration values k between 1 and 46 in increments of 5 iterations. This plot numerically verifies our analysis above.

3.2. Hybrid LSMR. Since LSMR exhibits semiconvergence behavior (as evident in Figure 1(a)) and $\hat{\mathbf{B}}$ becomes ill-conditioned with more GK iterations, we propose to solve the LSMR subproblem (2.11) using Tikhonov regularization, that is,

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \left\| \hat{\mathbf{B}}_k\mathbf{y} - \bar{\beta}_1\mathbf{e}_1 \right\|_2^2 + \lambda^2 \|\mathbf{y}\|_2^2, \tag{3.5}$$

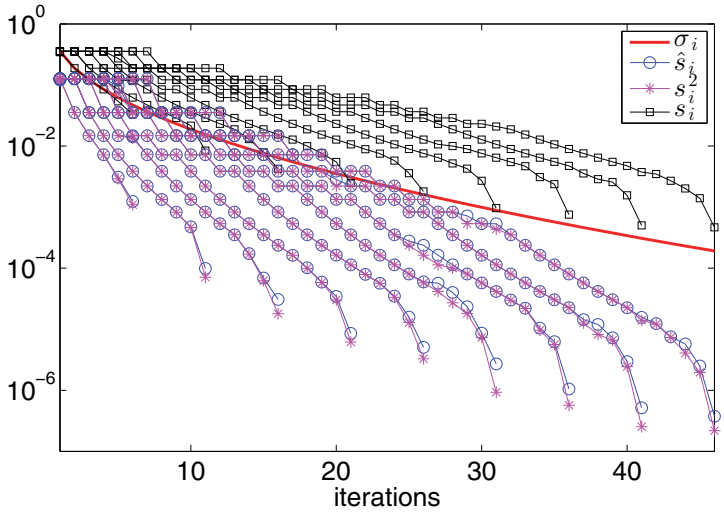


FIG. 2. *This plot compares the first* 46 *singular values of matrix* $\mathbf{A}(\sigma_i)$ *to the singular values of* $\hat{\mathbf{B}}_k(\hat{s}_i)$ *and the singular values of* $\mathbf{B}_k(s_i)$ *for iterations* k *between* 1 *and* 46 *in increments of* 5 *iterations. In addition, plots of* s_i^2 *illustrate the interlacing property described in section* 3.1.

where λ may be constant or vary per iteration. We call this a *hybrid LSMR* method, and the reconstruction is computed as $\mathbf{x}_k = \mathbf{V}_k\mathbf{y}_k$. Algorithmic details are presented shortly. First, we provide a theoretical analysis of hybrid LSMR.

By imposing Tikhonov regularization on the projected subproblem (i.e., solving (3.5) instead of (2.11)), we follow a "first project then regularize" approach. On the other hand, one could solve Tikhonov problem (2.1) using any iterative projection method for LS. This is considered a "first regularize then project" approach.

For previously studied hybrid methods based on LSQR, it has been shown that for fixed λ both approaches produce the same solution in exact arithmetic. More specifically, the solution after k iterations of LSQR on (2.1) is mathematically equivalent to the solution $\mathbf{V}_k\mathbf{y}_k$, where $\mathbf{y}_k$ solves (2.13); see [22, 20, 27, 34] and references therein. For large-scale problems where selecting a good regularization parameter a priori can be difficult, the latter case is preferred since standard methods for regularization parameter selection can be used on the projected problem, which is significantly smaller in dimension.

Contrary to previously studied hybrid methods, the solution after k iterations of LSMR applied to (2.1) is not equivalent to the solution after k iterations of hybrid LSMR. To see this, fix $\lambda > 0$ and let $\bar{\mathbf{r}} = \bar{\mathbf{b}} - \bar{\mathbf{A}}\mathbf{x}_k$ with $\bar{\mathbf{A}} = \begin{pmatrix}\mathbf{A}\\ \lambda\mathbf{I}\end{pmatrix}$, $\bar{\mathbf{b}} = \begin{pmatrix}\mathbf{b}\\ \mathbf{0}\end{pmatrix}$. Then, the solution after k iterations of LSMR applied to the Tikhonov problem $\min_{\mathbf{x}} \left\|\bar{\mathbf{A}}\mathbf{x} - \bar{\mathbf{b}}\right\|_2$ is given by $\mathbf{V}_k\mathbf{y}_k$, where

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \left\|\bar{\mathbf{A}}^\top\bar{\mathbf{r}}_k\right\|_2 = \arg\min_{\mathbf{y}} \left\|\hat{\mathbf{B}}_k\mathbf{y} - \bar{\beta}_1\mathbf{e}_1 + \lambda^2\begin{pmatrix}\mathbf{y}\\ 0\end{pmatrix}\right\|_2. \tag{3.6}$$

This solution is mathematically equivalent to the solution after k iterations of MINRES applied to normal equations

$$(\mathbf{A}^\top\mathbf{A} + \lambda^2\mathbf{I})\mathbf{x} = \mathbf{A}^\top\mathbf{b}\,. \tag{3.7}$$

However, it is not equivalent to the solution after k iterations of hybrid LSMR with λ fixed. To see this, compare subproblems (3.6) and (3.5). Such an equivalence can only be obtained in the case $\lambda = 0$.

Although hybrid LSMR for $\lambda > 0$ does not seem to have an equivalent "first regularize then project" approach, it can nevertheless be interpreted as a Krylov subspace projection method, thereby shedding light on the problem being solved.

THEOREM 3.2. *Fix $\lambda \geq 0$. Let $\mathbf{y}_k$ be the exact solution to the regularized subproblem*

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \left\| \hat{\mathbf{B}}_k \mathbf{y} - \bar{\beta}_1 \mathbf{e}_1 \right\|_2^2 + \lambda^2 \|\mathbf{y}\|_2^2 , \tag{3.8}$$

where $\hat{\mathbf{B}}_k, \mathbf{V}_k$ are derived from the original problem. Then the kth iterate of hybrid LSMR, $\mathbf{x}_k = \mathbf{V}_k \mathbf{y}_k$, solves the following problem:

$$\min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A}^\top \mathbf{A}, \mathbf{A}^\top \mathbf{b})} \left\| \mathbf{A}^\top \mathbf{A} \mathbf{x} - \mathbf{A}^\top \mathbf{b} \right\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 . \tag{3.9}$$

Proof. Projected problem (3.8) follows from derivations in section 2.2. ∎

3.3. Selecting regularization parameters for (3.5). In this section, we consider methods for selecting regularization parameters within hybrid LSMR. Since $\hat{\mathbf{B}}_k$ is small for small values of k, we can afford to use more expensive SVD-based regularization parameter selection schemes at each iteration. We consider three methods, DP, UPRE, and GCV, although any parameter selection scheme could be used.

Let $\hat{\mathbf{B}}_k = \hat{\mathbf{P}}_k \begin{pmatrix} \hat{\mathbf{S}}_k \\ \mathbf{0} \end{pmatrix} \hat{\mathbf{Q}}_k^\top$ be the SVD of $\hat{\mathbf{B}}_k$. Then the solution to (3.5) is

$$\mathbf{y}_k = \hat{\mathbf{B}}_{k,\lambda}^\dagger (\bar{\beta}_1 \mathbf{e}_1), \tag{3.10}$$

where $\hat{\mathbf{B}}_{k,\lambda}^\dagger$ is defined as in (2.3).

If a good estimate of the noise level for the original problem is available or can be estimated from the data, then DP or UPRE can be used to compute λ at each iteration. Otherwise, GCV can be used to select λ for the LSMR subproblem. Substituting $\mathbf{x}_\lambda = \mathbf{V}_k \mathbf{y}_k$ into (2.5), we can simplify the residual norm as

$$\|\mathbf{A}\mathbf{x}_\lambda - \mathbf{b}\|_2^2 = \|\mathbf{B}_k \mathbf{y}_k - \beta_1 \mathbf{e}_1\|_2^2 = \left\| \mathbf{B}_k \hat{\mathbf{B}}_{k,\lambda}^\dagger (\bar{\beta}_1 \mathbf{e}_1) - \beta_1 \mathbf{e}_1 \right\|_2^2 . \tag{3.11}$$

Since the singular vectors of $\mathbf{B}_k$ and $\hat{\mathbf{B}}_k$ are not directly related and their dimensions are relatively small, further simplifications are not computationally significant.

For UPRE and GCV, λ_k is selected at each iteration to minimize the functions

$$U_{\hat{\mathbf{B}}_k, \bar{\beta}_1 \mathbf{e}_1}(\lambda) = \frac{1}{k} \left\| (\mathbf{I} - \hat{\mathbf{B}}_k \hat{\mathbf{B}}_{k,\lambda}^\dagger) \bar{\beta}_1 \mathbf{e}_1 \right\|_2^2 + \frac{2\sigma^2}{k} \mathsf{trace}(\hat{\mathbf{B}}_k \hat{\mathbf{B}}_{k,\lambda}^\dagger) - \sigma^2 \tag{3.12}$$

and

$$G_{\hat{\mathbf{B}}_k, \bar{\beta}_1 \mathbf{e}_1}(\lambda) = \frac{k \left\| (\mathbf{I} - \hat{\mathbf{B}}_k \hat{\mathbf{B}}_{k,\lambda}^\dagger) \bar{\beta}_1 \mathbf{e}_1 \right\|_2^2}{\left(\mathsf{trace}(\mathbf{I} - \hat{\mathbf{B}}_k \hat{\mathbf{B}}_{k,\lambda}^\dagger) \right)^2}, \tag{3.13}$$

respectively. Using the SVD of $\hat{\mathbf{B}}_k$, we can simplify these functions as

(3.14)
$$U_{\hat{\mathbf{B}}_k,\bar{\beta}_1\mathbf{e}_1}(\lambda) = \frac{\bar{\beta}_1^2}{k}\left(\sum_{i=1}^{k}\left(\frac{\lambda^2}{\hat{s}_i^2+\lambda^2}[\hat{\mathbf{P}}_k^\top\mathbf{e}_1]_i\right)^2 + ([\hat{\mathbf{P}}_k^\top\mathbf{e}_1]_{k+1})^2\right) + \frac{2\sigma^2}{k}\sum_{i=1}^{k}\frac{\hat{s}_i^2}{\hat{s}_i^2+\lambda^2} - \sigma^2$$

and

$$G_{\hat{\mathbf{B}}_k,\bar{\beta}_1\mathbf{e}_1}(\lambda) = \frac{k\bar{\beta}_1^2\left(\sum_{i=1}^{k}\left(\frac{\lambda^2}{\hat{s}_i^2+\lambda^2}[\hat{\mathbf{P}}_k^\top\mathbf{e}_1]_i\right)^2 + ([\hat{\mathbf{P}}_k^\top\mathbf{e}_1]_{k+1})^2\right)}{\left(1+\sum_{i=1}^{k}\frac{\lambda^2}{\hat{s}_i^2+\lambda^2}\right)^2}, \tag{3.15}$$

where $[\hat{\mathbf{P}}_k^\top\mathbf{e}_1]_i$ denotes the ith component of vector $\hat{\mathbf{P}}_k^\top\mathbf{e}_1$ and $\hat{s}_i$ is the ith diagonal element of $\hat{\mathbf{S}}_k$.

We remark that the above residual norm and UPRE and GCV functions can be represented solely in terms of the SVD of $\mathbf{B}_k$. However, since the number of iterations remains moderately small, forming $\hat{\mathbf{B}}_k$ and obtaining its SVD is not a significant computation and results in simpler implementations. Furthermore, in hybrid LSMR we propose to delay regularization until $\hat{\mathbf{B}}_k$ more fully captures the ill-conditioning of $\mathbf{A}$ and the noise level estimate for the original problem becomes relevant.

3.4. Stopping criteria for the GK process. Another practical issue is determining a stopping iteration for the GK process. Standard methods described in [17] were not developed with ill-posed inverse problems in mind and would take quite a bit of tweaking and parameter tuning to work. Stopping criteria based on the GK bidiagonalization for ill-posed problems [25] could also be used in hybrid LSMR, but our numerical experience revealed that determining the noise-revealing iteration, as described in [25, (4.9)], was extremely sensitive to the user-defined parameters. Hence, we do not consider it here. Instead we develop a GCV function for selecting the stopping criteria that is similar to the derivation in [13], but for a different problem. Consider the problem

$$\min_{\mathbf{x}} \left\| \hat{\mathbf{A}}\mathbf{x} - \hat{\mathbf{b}} \right\|_2, \tag{3.16}$$

where $\hat{\mathbf{A}} = \mathbf{A}^\top\mathbf{A}$ and $\hat{\mathbf{b}} = \mathbf{A}^\top\mathbf{b}$. Then the GCV function to determine a stopping iteration, k, can be written as

$$G(k) = \frac{n\left\|(\mathbf{I} - \hat{\mathbf{A}}\hat{\mathbf{A}}_{k,\lambda}^\dagger)\hat{\mathbf{b}}\right\|_2^2}{(\mathsf{trace}(\mathbf{I} - \hat{\mathbf{A}}\hat{\mathbf{A}}_{k,\lambda}^\dagger))^2}, \tag{3.17}$$

where the solution at the kth iteration is given by

$$\mathbf{x}_k = \mathbf{V}_k\mathbf{y}_k \tag{3.18}$$

$$= \underbrace{\mathbf{V}_k(\hat{\mathbf{B}}_k^\top\hat{\mathbf{B}}_k + \lambda_k^2\mathbf{I})^{-1}\hat{\mathbf{B}}_k^\top\mathbf{V}_{k+1}^\top}_{\hat{\mathbf{A}}_{k,\lambda}^\dagger}\hat{\mathbf{b}}. \tag{3.19}$$

With the SVD of $\hat{\mathbf{B}}_k$, the GCV function (3.17) can be simplified as

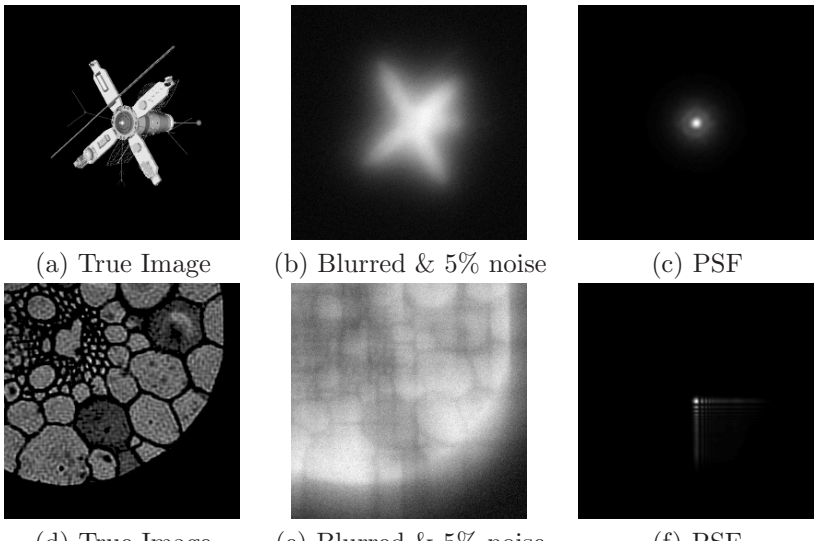
$$G(k) = \frac{n\bar{\beta}_1^2 \left(\sum_{i=1}^{k} \left(\frac{\lambda^2}{\hat{s}_i^2 + \lambda^2} [\hat{\mathbf{P}}_k^\top \mathbf{e}_1]_i \right)^2 + \left([\hat{\mathbf{P}}_k^\top \mathbf{e}_1]_{k+1} \right)^2 \right)}{\left(n - \sum_{i=1}^{k} \frac{\hat{s}_i^2}{\hat{s}_i^2 + \lambda^2} \right)^2}. \tag{3.20}$$

Since there is rarely one approach that is perfect, we propose to use a combination of stopping criteria. The hybrid LSMR method is terminated if a maximum number of iterations is attained, the GCV function (3.20) attains a minimum or flattens out, or tolerances on the residual are attained.

4. Numerical results. We test hybrid LSMR on two problems from image processing, namely, image deblurring and superresolution. In both examples, we compare LSMR and LSQR, along with their hybrid counterparts. For hybrid LSMR, we compare various methods for finding the regularization parameter. Full reorthogonalization of the GK vectors was used for Example 1 only.

Example 1: Image deblurring. This example uses two image deblurring problems from RestoreTools [31]: satellite and grain, both of which are 256×256 pixels. The satellite image is a computer simulation of a field experiment showing a satellite as taken from a ground-based telescope, and therefore represents an example of atmospheric blurring. We consider three different noise levels: 10%, 5%, and 1%, corresponding to $\frac{\|\mathbf{n}\|_2}{\|\mathbf{A}\mathbf{x}_{\text{true}}\|_2} = 0.1$, 0.05, and 0.01, respectively. The true image, blurred and noisy (5%) image, and point spread function (PSF) for satellite and grain are shown in Figure 3.

First, we compare hybrid LSMR and hybrid LSQR when the optimal regularization parameter is used at each iteration. Relative error plots as a function of iteration are displayed in Figure 4. The optimal regularization parameter corresponds



(a) True Image (b) Blurred & 5% noise (c) PSF

(d) True Image (e) Blurred & 5% noise (f) PSF

FIG. 3. *Image deblurring example. Figures* (a)–(c) *correspond to the satellite example, and Figures* (d)–(f) *correspond to the grain example.*

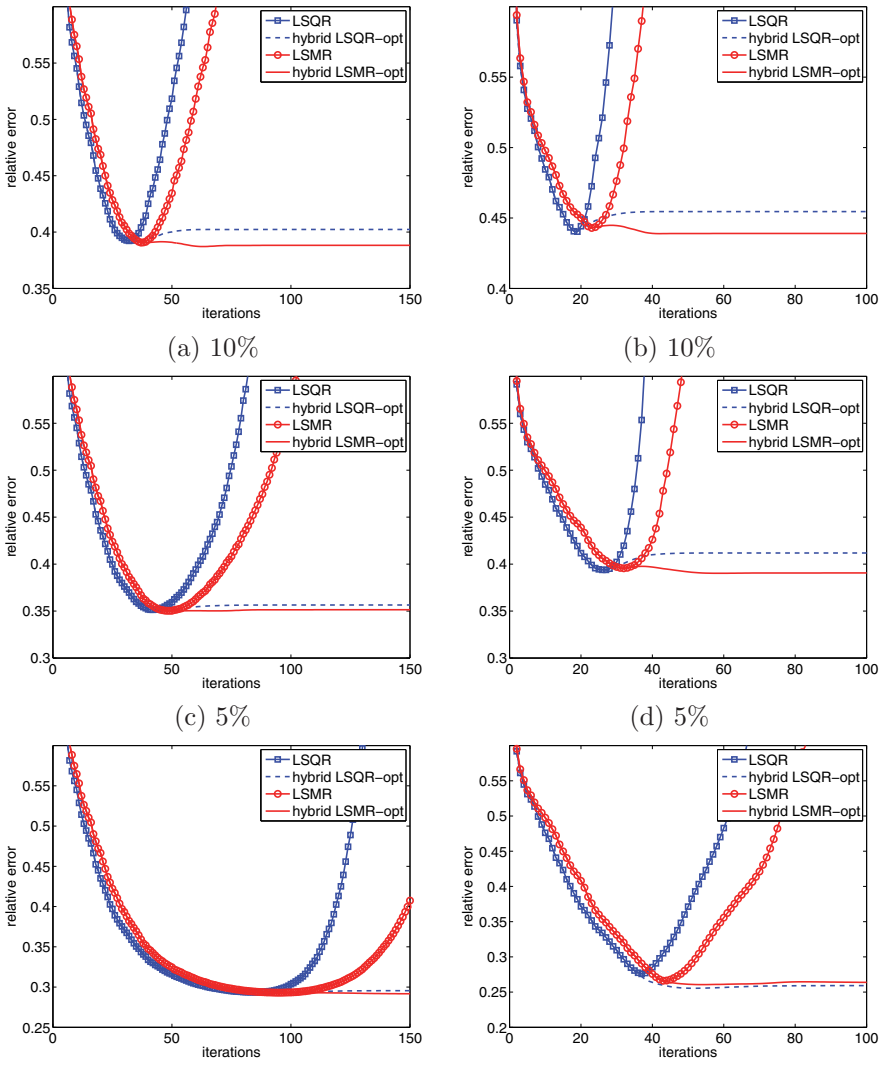


FIG. 4. *Comparison of LSQR and LSMR for satellite (column 1) and grain (column 2) at different noise levels. Results for hybrid LSMR-opt and hybrid LSQR-opt correspond to hybrid methods, where the optimal regularization parameter was used at each iteration.*

to the parameter λ_k that minimizes the error norm, $\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2$, at each iteration. This approach is not practical in real applications because the true solution is not available; however, we notice that if such a parameter can be obtained, then reconstructions for hybrid LSMR-opt provide slightly smaller relative errors than hybrid LSQR-opt, especially for problems with high noise levels. These plots demonstrate the competitiveness of hybrid LSMR. Relative error plots for LSQR and LSMR are included in Figure 4 for reference. They provide additional examples of the delayed semiconvergence behavior exhibited by LSMR on ill-posed problems.

Next, we compare various methods for selecting regularization parameters for hybrid LSMR, as discussed in section 3.3. In Figure 5 we display relative errors for

LSMR (where $\lambda_k = 0$ for all k) and hybrid LSMR with different parameter selection techniques at three different noise levels. For both DP and UPRE, we estimated the noise variance using the highest frequency of the wavelet transform of the observed image [14]. We set $\tau = 1$ in (2.5) for DP. Results corresponding to the satellite example are displayed in the first column. We observe that DP worked very well for this example, while GCV had a little trouble estimating a good regularization parameter. Nevertheless, it did very well for smaller noise levels (1–5%). UPRE performed very poorly for this example, resulting in very large reconstruction errors. For clarity of presentation, these results are not provided. Results corresponding to the grain example can be found in the second column. For this example DP did not

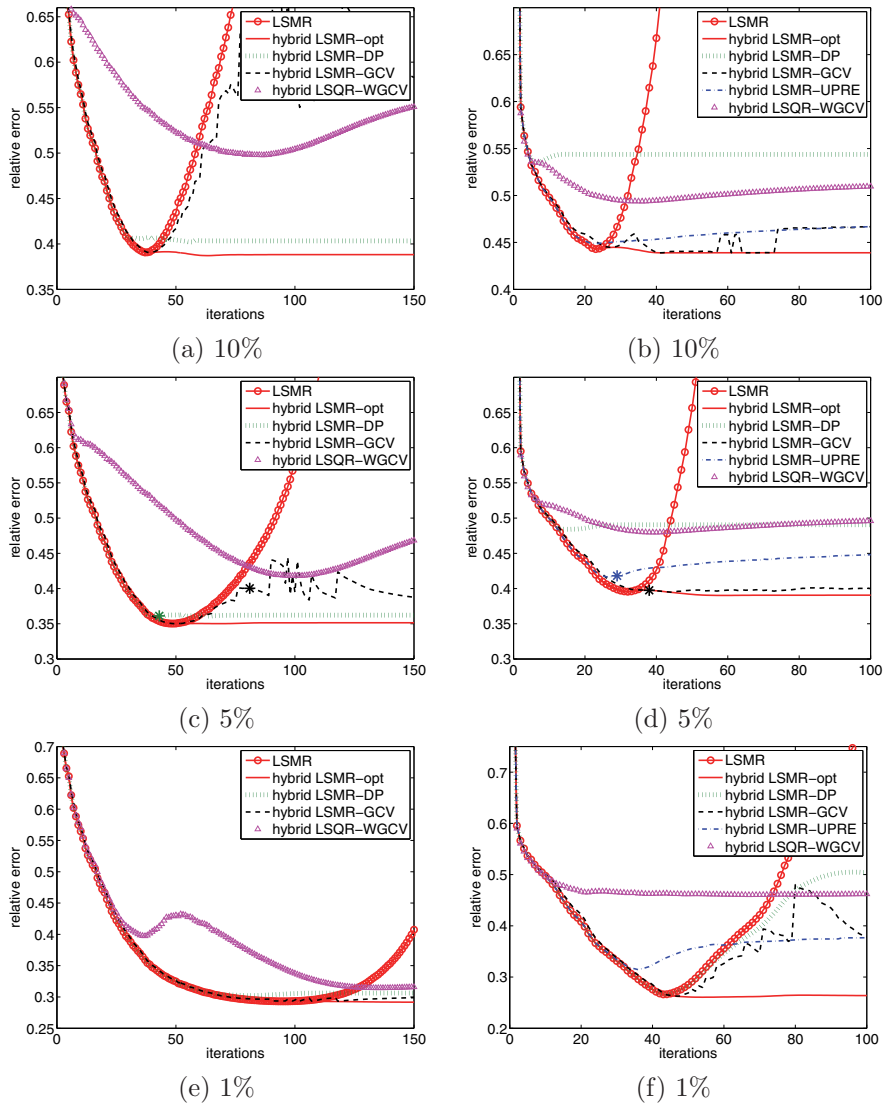


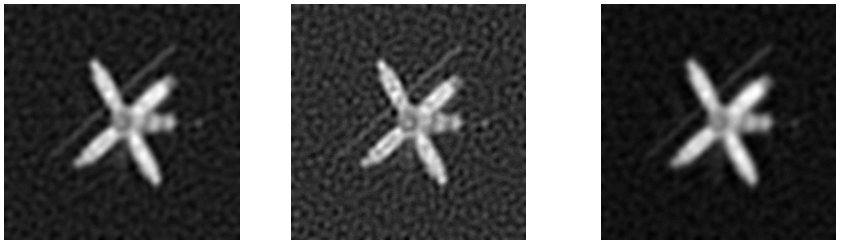
FIG. 5. *Comparison of regularization parameter selection methods for hybrid LSMR for satellite (column 1) and grain (column 2) at different noise levels. Results for hybrid LSQR using WGCV are provided for comparison.*

do as well, but GCV and UPRE did well, especially for large noise levels. In each figure, we provide the relative errors for hybrid LSQR using weighted GCV (WGCV) to automatically select the regularization parameter [13]. We do not claim that this is the best hybrid LSQR approach, but provide it only for comparison. We emphasize that parameter selection methods are problem dependent and no method is perfect [22, 27].

Finally, reconstructions for satellite and grain at 5% noise are provided in Figure 6. For both data sets, we provide the "best" LSMR reconstruction, corresponding to minimal reconstruction error. These use knowledge of the true image and correspond to 48 and 31 iterations for satellite and grain, respectively. Stopping criteria for hybrid LSMR were described in section 3.4. For satellite, hybrid LSMR-GCV and hybrid LSMR-DP terminated at 80 and 42 iterations, respectively. These iterations are marked with stars in Figure 5(c). The hybrid LSMR-GCV reconstruction in 6(b) contains noise "freckles" [23], but seems to be a sharper reconstruction than 6(a) and (c). For grain, hybrid LSMR-GCV terminated at 37 iterations and hybrid LSMR-UPRE terminated at 28 iterations, corresponding to the stars in Figure 5(d).

Example 2: Superresolution imaging. Superresolution imaging, or image fusion, is the process of combining a set of low-resolution images into one high-resolution image by using advanced mathematical algorithms and software tools. Superresolution imaging has gained increasing popularity over the last few years, as it allows scientists to obtain images with higher spatial resolution, without having to sacrifice signal-to-noise ratio and dynamic range [35, 16, 10].

Suppose we are given a set of low-resolution images $\mathbf{b}_1, \ldots, \mathbf{b}_K$, each containing different information of the same scene and represented in vector form. Then



(a) LSMR (b) hybrid LSMR-GCV (c) hybrid LSMR-DP

(d) LSMR (e) hybrid LSMR-GCV (f) hybrid LSMR-UPRE

FIG. 6. *Reconstructed images for the image deblurring examples for* 5% *noise. LSMR reconstructions in* (a) *and* (d) *correspond to optimal stopping iteration* 80 *for satellite and* 31 *for grain. These iterations are best in the sense that they correspond to minimal reconstruction error. Hybrid LSMR reconstructions for satellite correspond to* 80 *iterations for hybrid LSMR-GCV* (b) *and* 42 *iterations of hybrid LSMR-DP* (c)*, and hybrid LSMR reconstructions for grain correspond to* 37 *iterations of hybrid LSMR-GCV* (e) *and* 28 *iterations of hybrid LSMR-UPRE* (f).

superresolution imaging can be modeled as (1.1), where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_K \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_K \end{bmatrix}. \tag{4.1}$$

Each matrix $\mathbf{A}_i$ represents some linear transformation (e.g., rotation or linear shift) of the high-resolution image, followed by a restriction operation that takes a high-resolution image to a low-resolution image. Given $\mathbf{A}$ and $\mathbf{b}$, the goal of superresolution imaging is to reconstruct an approximation of the vectorized version of the true high-resolution image, $\mathbf{x}_{\text{true}}$.

For this example, the high-resolution image was 128×128 pixels and is shown in the top left corner of Figure 11. We generated 32 low-resolution images of size 32×32 pixels, and each image is slightly rotated from the others. White noise was added to each image, where the noise standard deviation was $\sigma = 0.0293$. Three observed low-resolution images are shown in Figure 7.

Each sparse matrix $\mathbf{A}_i$ was generated as described in [10], where bilinear interpolation was used to define the interpolation matrix and a Kronecker product was used to represent the restriction operation. For this example, we assume that all of the parameters defining $\mathbf{A}$ (e.g., rotation parameters) are known. If this is not the case, then such knowledge may be estimated using image registration techniques [29], or a separable nonlinear LS framework can be considered for simultaneous estimation of both the parameters and the high-resolution image [10, 12, 11]. In this example, we investigate the use of hybrid LSMR for solving the linear superresolution problem. We remark that this approach could also be used to solve the linear subproblem within a nonlinear optimization scheme.

First we compare the relative error plots for LSQR, LSMR, and their hybrid counterparts in Figure 8. As with previous examples, we see that LSMR exhibits slightly slower semiconvergence than LSQR, and hybrid LSMR-opt provides slightly smaller relative errors than hybrid LSQR-opt, demonstrating the competitiveness of hybrid LSMR.

Then we compare regularization parameter selection methods for hybrid LSMR. Both DP and UPRE used the estimated noise variance described above, where $\tau = 1$ for DP. Results are presented in Figure 9. We observe that UPRE tends to underestimate the regularization parameter, while DP seems to overestimate the parameter. Nevertheless, these two approaches can still produce reasonable reconstructions. Standard GCV had a difficult time in estimating regularization parameters for this problem, but WGCV worked well with $\omega = 1.05$ [13]. For reference, we provide errors corresponding to hybrid LSQR-WGCV.

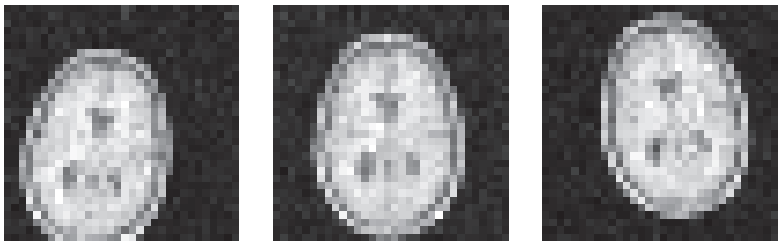


FIG. 7. *Three low-resolution images that were used for the superresolution imaging example.*

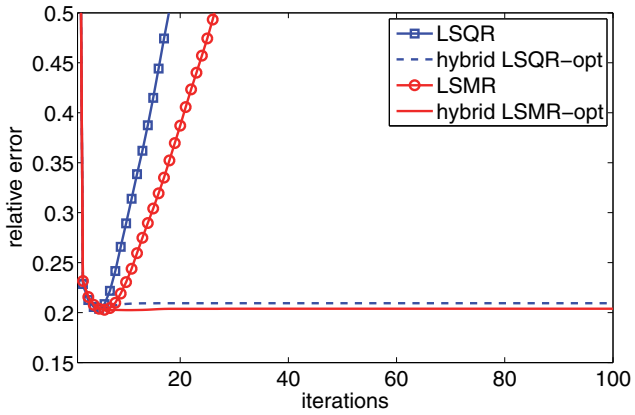


FIG. 8. *Comparison of LSQR and LSMR for superresolution image reconstruction. Results for hybrid LSQR-opt and hybrid LSMR-opt correspond to hybrid methods, where the optimal regularization parameter was used at each iteration.*

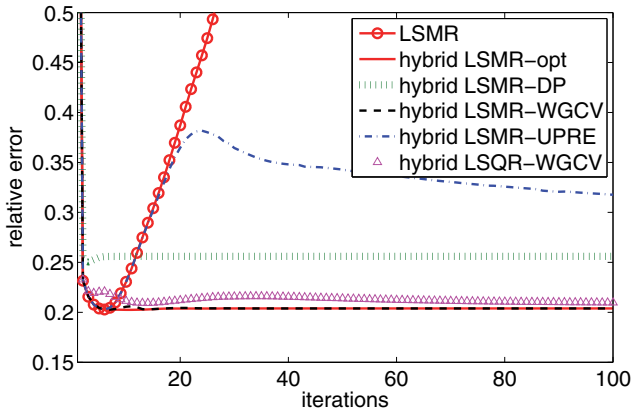


FIG. 9. *Comparison of regularization parameter selection methods within hybrid LSMR for the superresolution example. Results for hybrid LSQR using WGCV are provided for comparison.*

Both hybrid LSMR-DP and hybrid LSMR-UPRE use the estimated noise standard deviation for the original problem. In Figure 10, we consider the effect of using the estimated noise standard deviation, $\sigma_{est} = 0.0393$, versus the true standard deviation, $\sigma = 0.0293$. For DP, it seems that having an accurate estimate of the noise level can significantly improve the quality of the results. Similar plots were observed for different choices of τ. For UPRE, we did not encounter a significant difference. Future work includes investigating how the noise level for the original problem is related to the noise level of the subproblems, since noise in the subproblem for both LSQR and LSMR cannot be modeled as i.i.d. Gaussian.

For hybrid LSMR, we used a combination of stopping criteria, as described in section 3.4. The stopping iteration for each approach, along with the computed regularization parameter at that iteration, is provided in Table 1. Image reconstructions corresponding to stopping iterations k in Table 1 are provided in Figure 11.

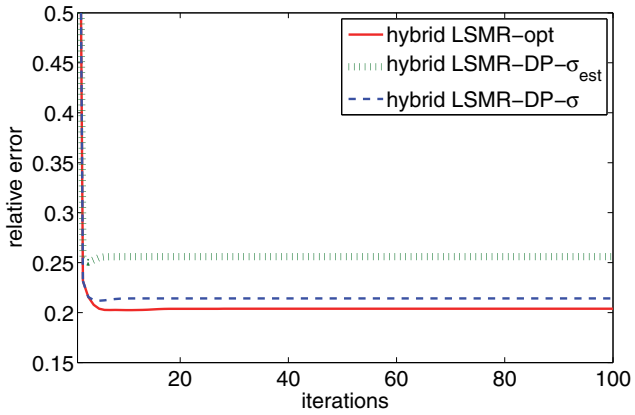


FIG. 10. *This figure compares the performance of hybrid LSMR-DP, using the true and estimated noise standard deviation.*

TABLE 1

Computed stopping iteration using the stopping criteria described in section 3.4*, and computed regularization parameters for the superresolution example.*

	opt	DP-σ_{est}	DP-σ	GCV	WGCV	UPRE
Stopping iteration, k	100	9	17	100	11	24
Parameter, λ_k	3.19e-2	2.65e-1	8.86e-2	1.13e-4	2.40e-2	2.82e-3

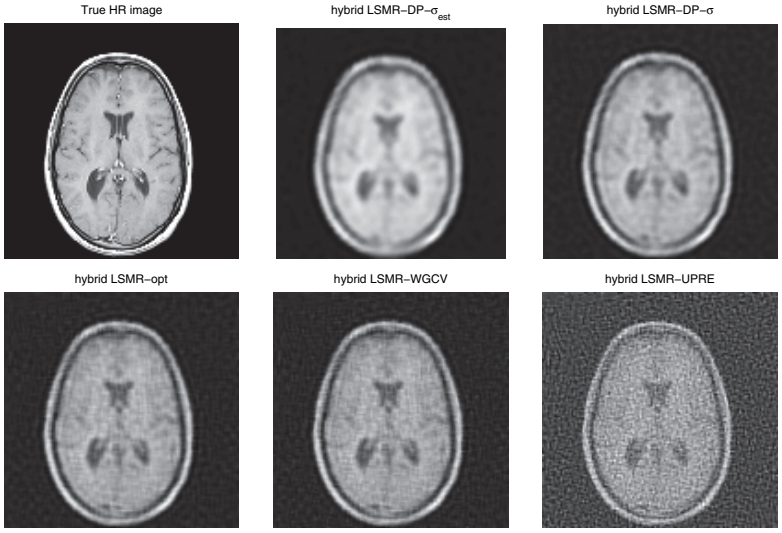


FIG. 11. *Reconstructions of the high-resolution image for various parameter selection methods, along with the true high-resolution (HR) image for the super-resolution example.*

5. Conclusions. We developed, implemented, and tested hybrid LSMR for solving large-scale ill-posed inverse problems. We showed that contrary to standard hybrid methods, hybrid LSMR is not equivalent to LSMR on the original Tikhonov-regularized problem. Rather, we showed that hybrid LSMR leads to a Krylov subspace problem for Tikhonov regularization applied to the normal equations. We motivated

the need for a hybrid LSMR approach by investigating the singular values of the LSMR subproblem matrix $\hat{\mathbf{B}}_k$, which approximates singular values of $\mathbf{A}^\top\mathbf{A}$ rather than $\mathbf{A}$. We also provided error plots for LSMR, which revealed delayed semiconvergence relative to LSQR for ill-posed problems. Another benefit is that hybrid LSMR can produce lower relative errors than hybrid LSQR, especially for problems with high noise levels. Methods for selecting regularization parameters and stopping criteria for hybrid LSMR were investigated, and numerical results were provided.

Acknowledgment. We are grateful to the referees for helpful and detailed comments to improve the manuscript.

REFERENCES

[1] J. Baglama, L. Reichel, and D. Richmond, *An augmented LSQR method*, Numer. Algorithms, 64 (2013), pp. 263–293.
[2] F. S. V. Bazán and L. S. Borges, *GKB-FP: An algorithm for large-scale discrete ill-posed problems*, BIT, 50 (2010), pp. 481–507.
[3] Å. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659–670.
[4] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[5] Å. Björck, E. Grimme, and P. van Dooren, *An implicit shift bidiagonalization algorithm for ill-posed systems of linear equations*, BIT, 34 (1994), pp. 510–534.
[6] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen, *Rank-one modification of the symmetric eigenproblem*, Numer. Math., 31 (1978), pp. 31–48.
[7] D. Calvetti, G. H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
[8] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
[9] D. Calvetti and L. Reichel, *Tikhonov regularization with a solution constraint*, SIAM J. Sci. Comput., 26 (2004), pp. 224–239.
[10] J. Chung, E. Haber, and J. G. Nagy, *Numerical methods for coupled super-resolution*, Inverse Problems, 22 (2006), pp. 1261–1272.
[11] J. Chung, S. Knepper, and J. G. Nagy, *Large-scale inverse problems in imaging*, in Handbook of Mathematical Methods in Imaging, O. Scherzer, ed., Springer, New York, 2011, pp. 43–86.
[12] J. Chung and J. G. Nagy, *An efficient iterative approach for large-scale separable nonlinear inverse problems*, SIAM J. Scientific Computing, 31 (2010), pp. 4654–4674.
[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] D. L Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.
[15] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Springer, New York, 2000.
[16] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, *Advances and challenges in super-resolution*, Int. J. Imaging Systems Tech., 14 (2004), pp. 47–57.
[17] D. C. L. Fong and M. Saunders, *LSMR: An iterative algorithm for sparse least-squares problems*, SIAM J. Scientific Computing, 33 (2011), pp. 2950–2971.
[18] G. H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.
[19] G. H. Golub, F. T. Luk, and M. L. Overton, *A block Lanczos method for computing the singular values and corresponding singular vectors of a matrix*, ACM Trans. Math. Software, 7 (1981), pp. 149–169.
[20] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surv. Math. Ind., 3 (1993), pp. 253–315.
[21] P. C. Hansen, *Rank-deficient and Discrete Ill-posed Problems*, SIAM, Philadelphia, 1997.
[22] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM Monogr. Math. Model. Comput., SIAM, Philadelphia, 2010.
[23] P. C. Hansen and T. K. Jensen, *Noise propagation in regularizing iterations for image deblurring*, Electron. Trans. Numer. Anal., 31 (2008), pp. 204–220.

[24] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra and Filtering*, Fundam. Algorithms, SIAM, Philadelphia, 2006.
[25] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT, 49 (2009), pp. 669–696.
[26] M. E. Kilmer, P. C. Hansen, and M. I. Español, *A projection-based approach to general-form Tikhonov regularization*, SIAM J. Sci. Comput., 29 (2007), pp. 315–330.
[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] R. M. Larsen, *Lanczos Bidiagonalization with Partial Reorthogonalization*, Ph.D. thesis, Department of Computer Science, University of Aarhus, Aarhus, Denmark, 1998.
[29] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, Oxford, 2003.
[30] J. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*, Comput. Sci. Eng. 10, SIAM, Philadelphia, 2012.
[31] J. G. Nagy, K. Palmer, and L. Perrone, *Iterative methods for image deblurring: A MATLAB object oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93.
[32] D. P. O'Leary and J. A. Simmons, *A bidiagonalization-regularization procedure for large-scale discretizations of ill-posed problems*, SIAM J. Sci. Stat. Comp., 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. Software, 8 (1982), pp. 43–71.
[34] C. C. Paige and M. A. Saunders, *Algorithm* 583*, LSQR: Sparse linear equations and least-squares problems*, ACM Trans. Math. Software, 8 (1982), pp. 195–209.
[35] S. C. Park, M. K. Park, and M. G Kang, *Super-resolution image reconstruction: A technical overview*, IEEE Signal Processing Mag., 20 (2003), pp. 21–36.
[36] R. A. Renaut, I. Hnětynková, and J. Mead, *Regularization parameter estimation for large-scale Tikhonov regularization using a priori information*, Comput. Statist. Data Anal., 54 (2010), pp. 3430–3445.
[37] R. C. Thompson, *The behavior of eigenvalues and singular values under perturbations of restricted rank*, Linear Algebra Appl., 13 (1976), pp. 69–78.
[38] C. R. Vogel, *Computational Methods for Inverse Problems*, Front. Appl. Math. 23, SIAM, Philadelphia, 2002.
[39] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Monogr. Numer. Anal. 87, Oxford University Press, Oxford, 1965.
[40] C. Zhao, T. Huang, X. Zhao, and L. Deng, *Two new efficient iterative regularization methods for image restoration problems*, Abstr. App. Anal., 2013 (2013), 129652.