

Technical Report

TR-2007-004

A Weighted GCV Method for Lanczos Hybrid Regularization

by

Julianne Chung, James Nagy, Dianne O'Leary

MATHEMATICS AND COMPUTER SCIENCE

EMORY UNIVERSITY

A WEIGHTED GCV METHOD FOR LANCZOS HYBRID REGULARIZATION

JULIANNE CHUNG*, JAMES G. NAGY†, AND DIANNE P. O'LEARY‡

Dedicated to Gene H. Golub on his 75th birthday

Abstract. Lanczos-hybrid regularization methods have been proposed as effective approaches for solving large-scale ill-posed inverse problems. Lanczos methods restrict the solution to lie in a Krylov subspace, but they are hindered by semi-convergence behavior, in that the quality of the solution first increases and then decreases. Hybrid methods apply a standard regularization technique, such as Tikhonov regularization, to the projected problem at each iteration. Thus, regularization in hybrid methods is achieved both by Krylov filtering and by appropriate choice of a regularization parameter at each iteration. In this paper we describe a weighted generalized cross validation (W-GCV) method for choosing the parameter. Using this method we demonstrate that the semi-convergence behavior of the Lanczos method can be overcome, making the solution less sensitive to the number of iterations.

Key words. generalized cross validation, ill-posed problems, iterative methods, Lanczos bidiagonalization, LSQR, regularization, Tikhonov

AMS Subject Classifications: 65F20, 65F30

1. Introduction. Linear systems that arise from large-scale inverse problems are typically written as

$$\mathbf{b} = A\mathbf{x} + \mathbf{e}, \tag{1.1}$$

where $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{x} \in \mathbb{R}^n$. The vector $\mathbf{e} \in \mathbb{R}^m$ represents unknown perturbations in the data (such as noise). Given A and \mathbf{b} , the aim is to compute an approximation of \mathbf{x} .

Inverse problems of the form (1.1) arise in many important applications, including image reconstruction, image deblurring, geophysics, parameter identification and inverse scattering; cf. [6, 15, 16, 29]. Typically these problems are *ill-posed*, meaning that noise in the data may give rise to significant errors in computed approximations of \mathbf{x} . The ill-posed nature of the problem is revealed by the singular values of A , which decay to and cluster at 0. Thus A is severely ill-conditioned, and *regularization* is used to compute stable approximations of \mathbf{x} [6, 12, 15, 29]. Regularization can take many forms; probably the most well known choice is Tikhonov regularization [12], which is equivalent to solving the least squares problem

$$\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} A \\ \lambda L \end{bmatrix} \mathbf{x} \right\|_2 \tag{1.2}$$

where L is a *regularization operator*, often chosen as the identity matrix or a discretization of a differentiation operator. The *regularization parameter* λ is a scalar,

*Department of Mathematics and Computer Science, Emory University. Email: jmchung@mathcs.emory.edu. Research supported in part by DOE Computational Sciences Graduate Research Fellowship.

†Department of Mathematics and Computer Science, Emory University. Email: nagy@mathcs.emory.edu. Research supported in part by the NSF under grant DMS-05-11454 and by an Emory University Research Committee grant.

‡Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland; and National Institute for Standards and Technology, Gaithersburg, Maryland. Email: oleary@cs.umd.edu. Research supported in part by National Science Foundation Grant CCF 0514213.

usually satisfying $\sigma_n \leq \lambda \leq \sigma_1$, where σ_n is the smallest singular value of A and σ_1 is the largest singular value of A .

No regularization method is effective without an appropriate choice of the regularization parameter. Various techniques can be used, such as the discrepancy principle, the L-curve, and generalized cross validation (GCV) [6, 15, 29]. There are advantages and disadvantages to each of these approaches [20], especially for large-scale problems. For example, to use the discrepancy principle, it is necessary to have information about the noise. In the case of GCV, efficient implementation for Tikhonov regularization requires computing the singular value decomposition (SVD) of the matrix A [10], which may be computationally impractical for large-scale problems. Some savings can be attained by using a bidiagonalization of A [5], or the iterative technique proposed by Golub and von Matt [11], but the cost can still be prohibitive for very large matrices. In the case of the L-curve, it may be necessary to solve (1.2) for several regularization parameters. This limitation can be partially alleviated by exploiting redundancies and additional information available in certain iterative methods [3, 8].

An alternative to Tikhonov regularization for large-scale problems is *iterative regularization*. In this case, an iterative method such as LSQR [27] is applied to the least squares problem,

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2. \quad (1.3)$$

When applied to ill-posed problems, iterative methods such as LSQR exhibit an interesting “semi-convergence” behavior. Specifically, the early iterations reconstruct information about the solution, while later iterations reconstruct information about the noise. This behavior can be observed (if the exact solution is known) by plotting the relative errors, $\|\mathbf{x}_k - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$, where \mathbf{x} is the exact solution and \mathbf{x}_k is the solution at the k th iteration. This is illustrated schematically in the left plot of Figure 1.1, where we plot the typical behavior of the relative error as the iteration proceeds. (Specific examples are detailed in later sections). If we terminate the iteration when the error is minimized, we obtain a regularized solution. Unfortunately the exact solution \mathbf{x} is not known in practical problems, so a plot of the relative errors cannot be used to find the optimal termination point. However, parameter selection methods such as the discrepancy principle, GCV and L-curve (see, for example, [15]) can be used to estimate this termination point. The difficulty is that these techniques are not perfect, and, as illustrated in the left plot in Figure 1.1, an imprecise estimate of the termination point can result in a solution whose relative error is significantly higher than the optimal.

The semi-convergence behavior of LSQR can be stabilized by using a *hybrid* method that combines an iterative Lanczos bidiagonalization algorithm with a direct regularization scheme, such as Tikhonov [1, 2, 13, 19, 20, 22, 26] or truncated SVD. The basic idea of this approach is to project the large-scale problem onto Krylov subspaces of small (but increasing) dimension. The projected problem can be solved cheaply using any direct regularization method. The potential benefits of this approach are illustrated in the right plot of Figure 1.1. Notice that, in contrast to the behavior of the relative errors for LSQR, the hybrid approach can effectively stabilize the iteration so that an imprecise (over) estimate of the stopping iteration does not have a deleterious effect on the computed solution.

A disadvantage of the hybrid approach is that at each iteration we must choose a new regularization parameter for the projected problem. Although this is not computationally expensive, in order for the approach to be viable for practical problems, we

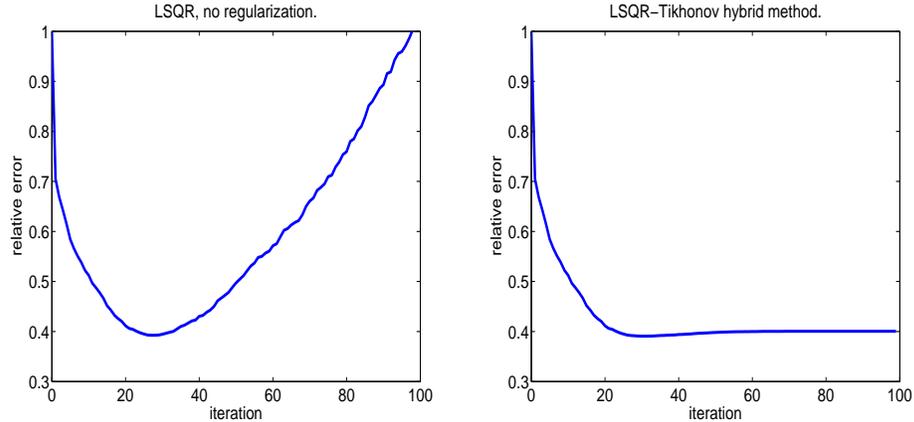


FIG. 1.1. The left plot illustrates semi-convergence behavior of LSQR for an ill-posed problem; a regularized solution is computed by terminating the iteration when the relative error is small, but the error can be very large if this termination point is over-estimated. The right plot illustrates how this semi-convergence behavior can be stabilized with the LSQR-Tikhonov hybrid method.

must choose good parameters. Optimal choices for the parameter at each iteration result in convergence behavior similar to that illustrated in the right plot of Figure 1.1. However, our computational experience indicates that such optimal behavior cannot be expected when using parameter selection methods such as the discrepancy principle, GCV and the L-curve (see also [20]).

In this paper we consider a Lanczos-hybrid method, using Tikhonov regularization for the projected problem, and we use GCV to choose regularization parameters for the projected problem. We show that GCV has a strong tendency to over-estimate the regularization parameter, but that a weighted GCV (W-GCV) method can be very effective.

An outline for the rest of the paper is as follows. In Section 2 we review Tikhonov regularization, the GCV method, and SVD based implementations. In Section 3 we describe the Lanczos-hybrid method, with Tikhonov regularization for the projected problem. We illustrate the deficiencies of using GCV with the method in Section 4. We show that although it is efficient, it generally provides an estimate that is too large. This can seriously degrade the overall convergence behavior. In Section 5, we describe the W-GCV method and show how it is related to the standard GCV. Numerical experiments are provided that illustrate the effectiveness of the W-GCV method on the test problems, and some concluding remarks are given in Section 6.

2. Tikhonov Regularization and GCV. To establish notation used in the paper, we briefly review Tikhonov regularization and GCV. In particular, we show that by using the SVD of the matrix A , we can recast the Tikhonov problem as a filtering method. In addition, the SVD allows us to put the GCV function into a computationally convenient form. Although this SVD approach is impractical for large-scale problems, it is both an extremely useful tool for problems of small dimension and an important component of the Lanczos-hybrid method.

Tikhonov regularization requires solving the minimization problem given in (1.2). For ease of notation, we take L to be the identity matrix. Let $A = U\Sigma V^T$ denote the SVD of A , where the columns \mathbf{u}_i of U and \mathbf{v}_i of V contain, respectively, the left and right singular vectors of A , and $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ is a diagonal matrix

containing the singular values of A , with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. Replacing A by its SVD and after a little algebraic manipulation, we can express the solution of (1.2) by

$$\mathbf{x}_\lambda = \sum_{i=1}^n \phi_i \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i, \quad (2.1)$$

where choosing the *filter factors* $\phi_i = 1$ gives the solution to (1.3). The Tikhonov *filter factors* ϕ_i are given by $\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \in [0, 1]$; cf. [15]. The regularization parameter, λ , plays a crucial role in the quality of the solution. For example, if λ is too large, the filter factors damp (or, equivalently, filter out) too many of the components in the SVD expansion (2.1), and the corresponding solution is *over-smoothed*. On the other hand, if λ is too small, the filter factors damp too few components, and the corresponding solution is *under-smoothed*.

As mentioned in the introduction, a variety of parameter choice methods can be used to determine λ . We choose to use the GCV method, which is a predictive statistics-based method that does not require a priori estimates of the error norm. The basic idea of GCV is that a good choice of λ should predict missing values of the data. That is, if an arbitrary element of the observed data is left out, then the corresponding regularized solution should be able to predict the missing observation fairly well [15]. We leave out each data value b_j in turn and seek the value of λ that minimizes the prediction errors, measured by the GCV function

$$G(\lambda) = \frac{n \|(I - AA_\lambda^\dagger) \mathbf{b}\|_2^2}{\left[\text{trace}(I - AA_\lambda^\dagger) \right]^2}, \quad (2.2)$$

where $A_\lambda^\dagger = (A^T A + \lambda^2 I)^{-1} A^T$ represents the pseudo-inverse matrix that gives the regularized solution, $\mathbf{x}_\lambda = A_\lambda^\dagger \mathbf{b}$. Replacing A with its SVD, (2.2) can be rewritten as

$$G(\lambda) = \frac{n \sum_{i=1}^n \left(\frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i^2 + \lambda^2} \right)^2}{\left(\sum_{i=1}^n \frac{1}{\sigma_i^2 + \lambda^2} \right)^2}, \quad (2.3)$$

which is a computationally convenient form to evaluate, thus making GCV easily used with standard minimization algorithms.

3. Lanczos-Hybrid Methods. Using GCV to determine the Tikhonov regularization parameter can be quite effective, but the minimization function (2.3) requires that the SVD of the matrix A be computed, and this is not feasible when A is too big. This leads us to Lanczos-hybrid methods, which make computing the SVD of the operator feasible by projecting the problem onto a subspace of small dimension. As described in Section 1 and illustrated in Figure 1.1, hybrid methods can be an effective way to stabilize the semi-convergent behavior that is characteristic of iterative methods like Lanczos when applied to ill-posed problems. Using an iterative method like Lanczos bidiagonalization (LBD) in combination with a direct method like Tikhonov regularization on the projected problem, we can hope to efficiently solve large-scale, ill-posed inverse problems. In this section, we provide some background on the Lanczos-hybrid methods.

Given a matrix A and vector \mathbf{b} , the k th-iteration of Lanczos bidiagonalization ($k = 1, \dots, n$) computes an $m \times (k + 1)$ matrix W_k , an $n \times k$ matrix Y_k , an $n \times 1$ vector \mathbf{y}_{k+1} , and a $(k + 1) \times k$ bidiagonal matrix B_k such that

$$\begin{aligned} A^T W_k &= Y_k B_k^T + \alpha_{k+1} \mathbf{y}_{k+1} \mathbf{e}_{k+1}^T \\ AY_k &= W_k B_k. \end{aligned}$$

where \mathbf{e}_{k+1} denotes the $(k + 1)$ st unit vector. Matrices W_k and Y_k have orthonormal columns, and the first column of W_k is $\mathbf{b}/\|\mathbf{b}\|$.

Given these relations, we approximate the least squares problem

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

by the *projected* LS problem

$$\begin{aligned} \min_{\mathbf{x} \in R(Y_k)} \|\mathbf{b} - A\mathbf{x}\|_2 &= \min_{\mathbf{f}} \|W_k^T \mathbf{b} - B_k \mathbf{f}\|_2 \\ &= \min_{\mathbf{f}} \|\beta \mathbf{e}_1 - B_k \mathbf{f}\|_2 \end{aligned} \quad (3.1)$$

where $\beta = \|\mathbf{b}\|$, and choose our approximate solution as $\mathbf{x}_k = Y_k \mathbf{f}$. Thus each iteration of the LBD method requires solving a least squares problem involving a bidiagonal matrix B_k .

The singular values of B_k for small values of k tend to approximate the largest and smallest singular values of A . Since the original problem is ill-posed, B_k may become very ill-conditioned. Therefore, regularization must be used to compute

$$\mathbf{f}_\lambda = \beta B_{k,\lambda}^\dagger \mathbf{e}_1$$

as described in Section 2. Notice that since the dimension of B_k is very small compared to A , we can afford to use SVD-based filtering methods to solve for \mathbf{f}_λ and SVD-based parameter choice methods to find λ at each iteration. O’Leary and Simmons [26] proposed using Tikhonov regularization to solve the projected problem, and Björck [1] suggested using truncated SVD (TSVD) with GCV to choose the regularization parameters. A variety of existing methods can be implemented. For a comparative study, see Kilmer and O’Leary [20]. Björck [1] also suggested using GCV as a way to determine an appropriate stopping iteration.

Next we illustrate how well this method works.

4. Experimental Results using Lanczos-Hybrid Methods and GCV. To illustrate the behavior of the GCV and W-GCV methods in Lanczos-hybrid methods, we use six problems.

4.1. Test Problems. The first problem is the *satellite* image deblurring problem that can be found in ‘RestoreTools’: an object oriented MATLAB package for image restoration [23]. In this problem, the goal is to reconstruct an image \mathbf{x} , given the blurred image represented by a vector \mathbf{b} of length 65,536, along with some information about the blurring operator, represented by a $65,536 \times 65,536$ matrix A .

The other five test problems are taken from the ‘Regularization Tools’ package [14]. In each case we generate a 256×256 matrix A , true solution vector \mathbf{x} , and (noise free) observation vector \mathbf{b} .

- **Phillips** is Phillips' "famous" test problem, obtained by discretizing the first kind Fredholm integral equation $b(s) = \int_{-6}^6 a(s, t)x(t)dt$, where

$$\begin{aligned} a(s, t) &= \rho(s - t) \\ x(t) &= \rho(t) \\ b(s) &= (6 - |s|) \left(1 + \frac{1}{2} \cos\left(\frac{\pi s}{3}\right) \right) + \frac{9}{2\pi} \sin\left(\frac{\pi|s|}{3}\right), \end{aligned}$$

and

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

- **Shaw** is a one-dimensional image restoration problem. A and \mathbf{x} are obtained by discretizing, on the interval $-\frac{\pi}{2} \leq s, t \leq \frac{\pi}{2}$, the functions

$$\begin{aligned} a(s, t) &= (\cos(s) + \cos(t)) \left(\frac{\sin(u)}{u} \right)^2, \quad u = \pi(\sin(s) + \sin(t)) \\ x(t) &= 2 \exp(-6(t - 0.8)^2) + \exp(-2(t + 0.5)^2), \end{aligned}$$

and $\mathbf{b} = A\mathbf{x}$.

- **Deriv2** constructs A , \mathbf{b} and \mathbf{x} by discretizing a first kind Fredholm integral equation, $b(s) = \int_0^1 a(s, t)x(t)dt$, $0 \leq s \leq 1$, where the kernel $a(s, t)$ is given by the Green's function for the second derivative:

$$a(s, t) = \begin{cases} s(t - 1) & , \quad s < t \\ t(s - 1) & , \quad s \geq t \end{cases}$$

There are several choices for x and b ; in this paper, we use $x(t) = t$ and $b(s) = (s^3 - s)/6$.

- **Baart** constructs A , \mathbf{b} and \mathbf{x} by discretizing the first kind Fredholm integral equation $b(s) = \int_0^\pi a(s, t)x(t)dt$, $0 \leq s \leq \frac{\pi}{2}$, where

$$\begin{aligned} a(s, t) &= \exp(s \cos t) \\ x(t) &= \sin t \\ b(s) &= \frac{2 \sinh s}{s}. \end{aligned}$$

- **Heat** is an inverse heat equation using the Volterra integral equation of the first kind on $[0, 1]$ with kernel $a(s, t) = k(s - t)$, where

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

The vector \mathbf{x} does not have a simple functional representation, but rather is constructed directly as a discrete vector; see [14] for details. The right-hand side \mathbf{b} is produced as $\mathbf{b} = A * \mathbf{x}$.

In order to simulate noisy data, as modeled by equation (1.1), for each test problem, we generate a noise vector \mathbf{e} whose entries are chosen from a normal distribution with mean 0 and variance 1, and scaled so that

$$\frac{\|\mathbf{e}\|_2}{\|A\mathbf{x}\|_2} = 0.1 .$$

4.2. What Goes Wrong in Using GCV-Lanczos?. We solved these test problems with the Lanczos-based hybrid method, using GCV to choose the Tikhonov regularization parameter λ at each iteration. The results are shown in Figure 4.1. In all of our examples, LSQR, which is essentially LBD with no regularization, exhibits semi-convergent behavior, as we expect. If we could choose ‘optimal’ regularization parameters at each iteration, then Lanczos-hybrid methods would be excellent at stabilizing the regularized solution, as shown with the dashed line. However, in realistic situations, we do not know the optimal solution, so this is impossible. On the *Phillips*, *Shaw* and *Deriv2* problems, the performance of standard GCV, though slightly worse than optimal, is acceptable. For the other three problems, the convergence behavior for GCV is significantly worse than optimal.

A major concern is the possibility that rounding errors in the computation of the matrices W_k , Y_k and B_k are causing the poor behavior. Björck, Grimme and Van Dooren [2] showed that in some cases reorthogonalization may be necessary for better performance, and Larsen [22] considered partial reorthogonalization. However, in our tests GCV still had difficulty even after reorthogonalization. Another option is to use a different regularization method such as TSVD or exponential filtering, but we found little to no improvement in the solution. In addition, we delayed regularization until after $k > k_{min}$ to wait until B_k captures the ill-conditioning of A , but that attempt proved futile as well.

Our conclusion is that the poor behavior is caused by the suboptimal parameter chosen by GCV, so in the next section we propose replacing it by a weighted-GCV method which shows much better behavior.

5. Weighted-GCV for Lanczos-Tikhonov methods. The standard GCV method is a popular parameter choice method used in a variety of applications; however, as we have just seen, the method may not perform well for certain classes of problems. Here we consider a weighted-GCV function for Lanczos-hybrid methods using Tikhonov regularization.

Other studies in statistical nonparametric modeling and function approximation noted that in practical applications, GCV occasionally chose Tikhonov parameters too small, thereby under-smoothing the solution [4, 7, 21, 25, 28]. Thus, the investigators developed a concept that we call weighted-GCV. In contrast, we observed over-smoothing difficulties when using GCV in Lanczos-hybrid methods, which motivated us to use a different range of weights in the W-GCV method.

Instead of the Tikhonov GCV function defined in (2.2), we consider the weighted-GCV function

$$G(\omega, \lambda) = \frac{n \|(I - AA_\lambda^\dagger)\mathbf{b}\|^2}{\left[\text{trace}(I - \omega AA_\lambda^\dagger)\right]^2}.$$

Notice the function’s dependency on a new parameter ω in the denominator trace term. Choosing $\omega = 1$ gives the standard GCV function (2.2). If we choose $\omega > 1$, we obtain smoother solutions, while $\omega < 1$ results in less smooth solutions. The obvious question here is how to choose a good value for ω . For smoothing spline applications, Kim and Gu empirically found that standard GCV consistently produced regularization parameters that were too small, while choosing ω in the range of 1.2-1.4 worked well [21]. In our problems, though, the GCV regularization parameter is chosen too large, and thus we choose a parameter ω in the range $0 < \omega \leq 1$. In addition, rather than using a user-defined parameter choice for ω as in previous

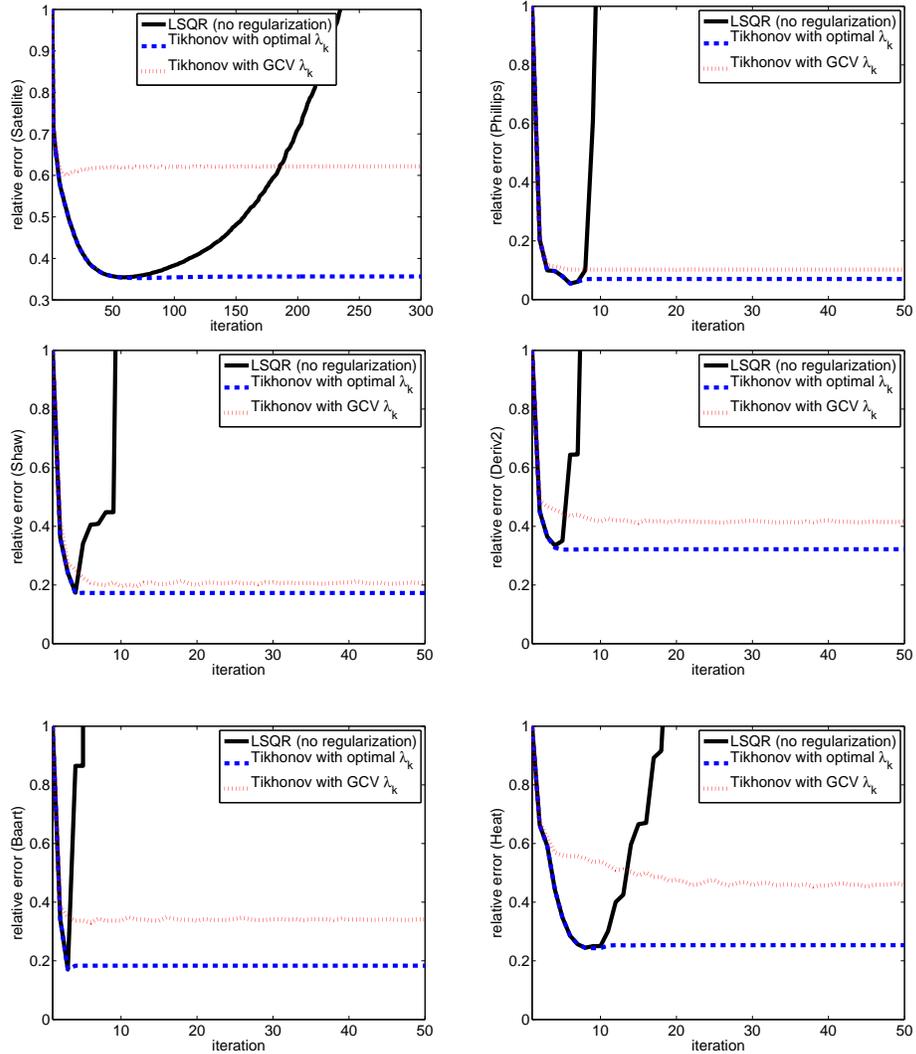


FIG. 4.1. These plots show the relative error at each iteration of LSQR and the Lanczos-hybrid method. Upper left: Satellite. Upper right: Regtools-Phillips. Middle left: Regtools-Shaw. Middle right: Regtools-Deriv2. Bottom left: Regtools-Baart. Bottom right: Regtools-Heat. The standard GCV method chooses regularization parameters that are too large at each iteration, which causes poor convergence behavior.

papers, we propose a more automated approach that is also versatile and can be used on a variety of problems.

5.1. Interpretations of the W-GCV method. In this section, we consider the W-GCV method and look at various theoretical aspects of the method. By looking at different interpretations of the W-GCV method, we hope to shed some light on the role of the new parameter, ω .

As mentioned in Section 2, the standard GCV method is a “leave-one-out” prediction method. In fact, in leaving out the j th observation, the derivation seeks to

minimize the prediction error, given by

$$\sum_{i=1, i \neq j}^n (b_i - [A\mathbf{x}]_i)^2 + \lambda^2 \|\mathbf{x}\|_2^2.$$

If we define the matrix

$$E_j = \text{diag}(1, 1, \dots, 1, 0, 1, \dots, 1),$$

where 0 is the j^{th} entry, then the above minimization is equivalent to

$$\min_{\mathbf{x}} \|E_j(\mathbf{b} - A\mathbf{x})\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2.$$

We can derive the W-GCV method in a similar manner, but we instead use a weighted “leave-one-out” philosophy. More specifically, consider the case $0 < \omega < 1$. Then define the matrix

$$F_j = \text{diag}(1, 1, \dots, 1, \sqrt{1-\omega}, 1, \dots, 1),$$

where $\sqrt{1-\omega}$ is the j^{th} diagonal entry of F_j . By using the W-GCV method, we seek a solution to the minimization problem,

$$\min_{\mathbf{x}} \|F_j(\mathbf{b} - A\mathbf{x})\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2.$$

In this problem, the j^{th} observation is still present but has been down-weighted by the factor $\sqrt{1-\omega}$; thus it is completely left out when $\omega = 1$. A derivation of the W-GCV method follows immediately from the derivation of the GCV method found in [9].

By introducing a new parameter in the trace term of the GCV function, we not only introduce a new weighted prediction approach, but also change the interpretation of the function we are minimizing. We consider the special case of Tikhonov regularization and look at how the GCV function is altered algebraically and graphically with the new parameter. Using the SVD expansion of A , it can be shown that the trace term in the standard GCV function is given by

$$\begin{aligned} \text{trace}(I - AA_F^\dagger) &= n - \sum_{i=1}^n \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \\ &= \sum_{i=1}^n (1 - \phi_i). \end{aligned}$$

In contrast, the trace term for the W-GCV function is given by:

$$\begin{aligned} \text{trace}(I - \omega AA_F^\dagger) &= n - \omega \sum_{i=1}^n \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \\ &= \sum_{i=1}^n (1 - \phi_i) + (1 - \omega) \sum_{i=1}^n \phi_i. \end{aligned}$$

Thus, depending on the value of ω , we are either adding or subtracting a multiple of the sum of the filter factors to the original trace term. The graph of the GCV

function also undergoes changes as ω is changed from 1. If $\omega > 1$, we introduce a pole in the W-GCV function whenever $\sum_{i=1}^n \phi_i = \frac{n}{\omega}$, which could cause difficulties when attempting to find the minimum of the W-GCV function. Fortunately, in our case, $0 < \omega \leq 1$. Note that larger values of ω result in larger computed regularization parameters, and smaller values of ω result in smaller values of λ .

5.2. Choosing ω . For many ill-posed problems, a good value of ω is crucial for the success of Lanczos-hybrid methods. In this section we consider how different values of ω may affect convergence behavior and present an adaptive approach for finding a good value for ω .

Consider the test problem *Heat*, whose convergence graph with Tikhonov regularization and the standard GCV method is given in the bottom right corner of Figure 4.1. To illustrate the effects of using the W-GCV function with Lanczos-hybrid methods, we present results using different values of ω . More specifically, for each trial we fix a value for ω that is used at each step of the iterative process. The results are shown in Figure 5.1.

For this particular example, it is evident that $\omega = 0.2$ is a good value for the new parameter. However, finding a good ω in this way is not possible since the true solution is generally not available. Hence, we introduce an automated, adaptive approach that in our experience produces adequate results.

Recall from Section 3 that at each iteration of the Lanczos-hybrid method, we solve the projected LS problem (3.1) using Tikhonov regularization. Since the early iterations of LBD do not capture the ill-conditioning of the problem, we expect that little or no regularization is needed to solve the projected LS problem. Let $\lambda_{k,opt}$ denote the optimal regularization parameter at the k^{th} iteration. Then, we can assume that for small k , $\lambda_{k,opt}$ should satisfy

$$0 \leq \lambda_{k,opt} \leq \sigma_{min}(B_k).$$

If at iteration k , we assume that we know $\lambda_{k,opt}$, then we can find ω by minimizing the GCV function with respect to ω . That is, solving

$$\left. \frac{\partial}{\partial \lambda} [G(\omega, \lambda)] \right|_{\lambda=\lambda_{k,opt}} = 0.$$

Since we do not know $\lambda_{k,opt}$, we instead find $\hat{\omega}_k$ corresponding to $\lambda_{k,opt} = \sigma_{min}(B_k)$. In later iterations, this approach fails because $\sigma_{min}(B_k)$ becomes nearly zero due to ill-conditioning. For these iterations, a better approach is to adaptively take $\omega_k = \text{mean}\{\hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_k\}$. By averaging the previously computed ω values, we are essentially using the earlier well-conditioned components of our problem to help stabilize the harmful effects of the smaller singular values. There are two disadvantages to this approach. First, it over-smooths the solutions at early iterations, since it uses a rather large value of λ for a well-conditioned problem. Since these solutions are discarded, this is not a significant difficulty. Second, it undersmooths values for large k , so semi-convergence will eventually reappear. However, in practice we will also be using a method like GCV to choose a stopping iteration, so k will not be allowed to grow too large; this is discussed in the following subsection.

5.3. Stopping Criteria. The next practical issue to consider is an approach to determine an appropriate stopping iteration. Björck [1] suggested using GCV for

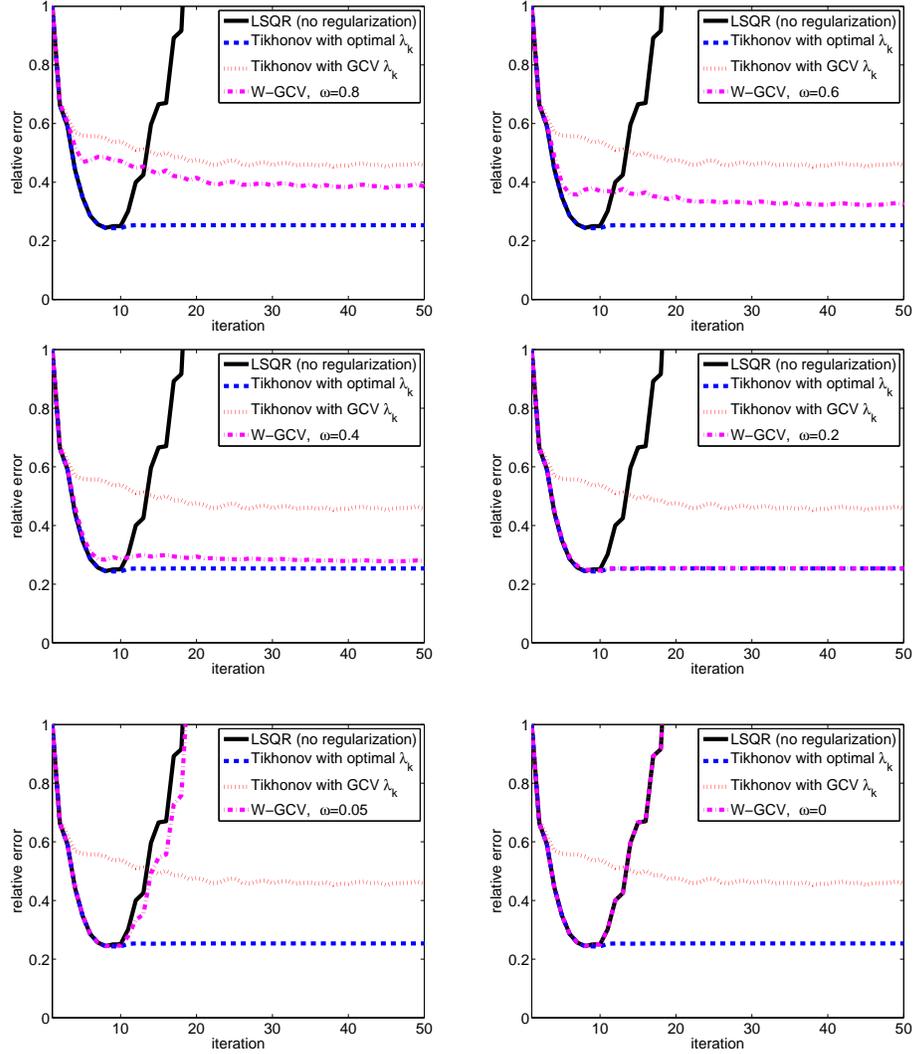


FIG. 5.1. This figure shows plots of the relative errors for LSQR and the Lanczos-hybrid method for the Heat example from Regtools. The various plots show how the convergence behavior changes when regularization parameters are chosen using the W-GCV method with different values of ω . Note that $\omega = 1$ is equivalent to using standard GCV, and $\omega = 0$ is equivalent to using no regularization.

this purpose, when TSVD is used to solve the projected problem. However, Björck, Grimme and van Dooren [2] showed that modifications of the algorithm were needed to make the approach effective for practical problems. Specifically, they proposed a fairly complicated scheme based on implicitly restarting the iterations.

In this section we describe stopping criteria that use the weighted GCV function with Tikhonov regularization to solve the projected problem. Implicit restarts are not needed in our implementation.

Each iteration of the Lanczos-hybrid method chooses a regularization parameter,

λ_k , satisfying

$$\lambda_k = \operatorname{argmin}_{\lambda} G(\omega, \lambda).$$

In the ideal situation where the convergence behavior of the Lanczos-hybrid method is perfectly stabilized, we expect λ_k to converge to a fixed value corresponding to an appropriate regularization parameter for the original problem (1.2). In this case the values of $G(\omega, \lambda_k)$ will also converge to a fixed value. Therefore, we choose to terminate the iterations when these values change very little; for example,

$$\left| \frac{G(\omega, \lambda_{k+1}) - G(\omega, \lambda_k)}{G(\omega, \lambda_1)} \right| < \text{tol},$$

for some prescribed tolerance.

However, as remarked in the previous subsection, it may be impossible to completely stabilize the iterations for realistic problems, resulting in slight semi-convergent behavior of the iterations. In this case, the GCV values $G(\omega, \lambda_k)$ will begin to increase. Thus, we implement a second stopping criteria to stop at iteration k_0 satisfying

$$k_0 = \operatorname{argmin}_k G(\omega, \lambda_k).$$

In the next section we present some numerical experiments with an implementation of the Lanczos-hybrid method that uses this approach.

5.4. Numerical Results. We now illustrate the effectiveness of using the W-GCV method in Lanczos-hybrid methods with Tikhonov regularization. We implement the adaptive method presented in Section 5.2 for choosing ω and provide numerical results for each of the test problems. The resulting convergence curves are displayed in Figure 5.2.

In all of the test problems, choosing ω adaptively provides nearly optimal convergence behavior. The results for the *Phillips* and *Shaw* problems are excellent with the adaptive W-GCV approach. The *Satellite*, *Baart* and *Heat* examples exhibit a slowed convergence, but achieve much better results than with the standard GCV. In addition, preconditioning could be used here to accelerate convergence. It is interesting to note that the *Deriv2* example converges, but eventually exhibits small signs of semi-convergent behavior. Nevertheless, the results are still better than the standard GCV, and, moreover, if combined with the stopping criteria described in the previous section, the results are quite good. To illustrate, in Table 5.1 we report the iteration at which our code detected a minimum of $G(\omega, \lambda_k)$.

TABLE 5.1

Results of using $G(\omega, \lambda_k)$ to determine a stopping iteration. The numbers reported in this table are the iteration index at which our Lanczos-hybrid code detected a minimum of $G(\omega, \lambda_k)$.

Problem	<i>Satellite</i>	<i>Phillips</i>	<i>Shaw</i>	<i>Deriv2</i>	<i>Baart</i>	<i>Heat</i>
Stopping Iteration	249	12	12	20	9	26

Comparing the results in Table 5.1 with the convergence history plots shown in Figure 5.2, we see that our approach to choosing a stopping iteration is very effective. Although the scheme does not perform as well on the *Baart* example, the results are still quite good considering the difficulty of this problem. (Observe that with no regularization, semi-convergence happens very quickly, and we should therefore

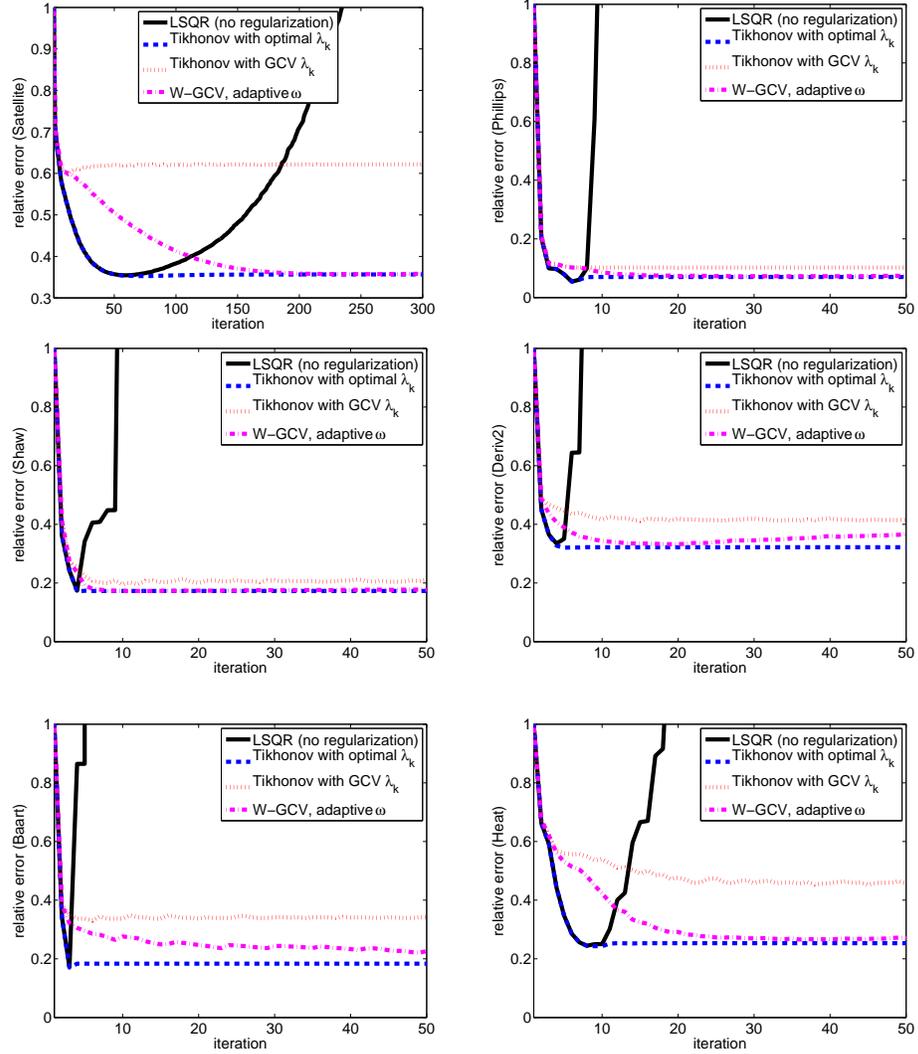


FIG. 5.2. These plots show the relative error at each iteration of LSQR and the Lanczos-hybrid method. Upper left: *Satellite*. Upper right: *Regtools-Phillips*. Middle left: *Regtools-Shaw*. Middle right: *Regtools-Deriv2*. Bottom left: *Regtools-Baart*. Bottom right: *Regtools-Heat*. The standard GCV method chooses regularization parameters that are too large at each iteration, which cause poor convergence behavior. However, the W-GCV method, with our adaptive approach to choose ω , produces near optimal convergence behavior.

expect difficulties in stabilizing the iterations.) These results show that our W-GCV method performs better than standard GCV, and that we are able to determine an appropriate stopping iteration on a wide class of problems.

We also remark that the *Satellite* example is a much larger problem than the other examples, and so more iterations are needed. However, the Lanczos-hybrid method can easily incorporate standard preconditioning techniques to accelerate convergence. For the *Satellite* image deblurring example, we used a Kronecker product based preconditioner [17, 18, 24] implemented in RestoreTools [23]. In this case, the

Lanczos-hybrid method, with W-GCV, detects a minimum of $G(\omega, \lambda_k)$ in only 3 iterations. The corresponding solution has relative error 0.3749, which compares very well with relative error of 0.3573 achieved at iteration 249 when using no preconditioning.

6. Concluding Remarks. In this paper, we have considered using a weighted GCV method in Lanczos-hybrid methods for solving large scale ill-posed problems. The W-GCV method requires choosing yet another parameter, so we proposed and implemented an adaptive, automatic approach for choosing this parameter. We demonstrated through a variety of test problems that our approach was effective in stabilizing semi-convergence behavior.

The MATLAB implementations used to generate the results presented in this paper can be obtained from <http://www.mathcs.emory.edu/~nagy/WGCV>.

Several open questions remain. With the ability to obtain near optimal solutions, Lanczos-hybrid methods should have a significant impact on many applications. Recently, Kilmer, Hansen and Español [19] suggested a projection-based algorithm that can be implemented for more general regularization operators. We can treat this iterative method as a hybrid method and apply W-GCV. In addition, we would like to see how well W-GCV works in combination with truncated-SVD and Lanczos-hybrid methods. Finally, work remains to be done on alternative ways to determine the new parameter in the W-GCV method.

REFERENCES

- [1] Å. BJÖRCK, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659–670.
- [2] 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.
- [3] D. CALVETTI, G. H. GOLUB, AND L. REICHEL, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
- [4] D. CUMMINS, T. FILLOON, AND D. NYCHKA, *Confidence intervals for non-parametric curve estimates: Toward more uniform pointwise coverage*, J. Am. Stat. Assoc., 96 (2001), pp. 233–246.
- [5] L. ELDEÉN, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
- [6] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 2000.
- [7] J. FRIEDMAN AND B. SILVERMAN, *Flexible parsimonious smoothing and additive modeling*, Technometrics, 31 (1989), p. ???
- [8] A. FROMMER AND P. MAASS, *Fast CG-based methods for Tikhonov-Phillips regularization*, SIAM J. Sci. Comput., 20 (1999), pp. 1831–1850.
- [9] 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.
- [10] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations, third edition*, Johns Hopkins Press, 1996.
- [11] G. H. GOLUB AND U. VON MATT, *Quadratically constrained least squares and quadratic problems*, Numer. Math., 59 (1991), pp. 561–580.
- [12] C. W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind*, Pitman, Boston, 1984.
- [13] M. HANKE, *On lanczos based methods for the regularization of discrete ill-posed problems*, BIT, 41 (2001), pp. 1008–1018.
- [14] P. C. HANSEN, *Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems*, Numerical Algorithms, 6 (1994), pp. 1–35.
- [15] ———, *Rank-deficient and discrete ill-posed problems*, SIAM, Philadelphia, PA, 1997.
- [16] P. C. HANSEN, J. G. NAGY, AND D. P. O'LEARY, *Deblurring Images: Matrices, Spectra and Filtering*, SIAM, Philadelphia, PA, 2006.
- [17] J. KAMM AND J. G. NAGY, *Kronecker product and SVD approximations in image restoration*, Linear Algebra Appl., 284 (1998), pp. 177–192.

- [18] ———, *Optimal Kronecker product approximation of block Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 155–172.
- [19] M. E. KILMER, P. C. HANSEN, AND M. I. ESPAÑOL, *A projection-based approach to general-form Tikhonov regularization*, SISC, submitted (2006).
- [20] 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.
- [21] Y. KIM AND C. GU, *Smoothing spline Gaussian regression: More scalable computation via efficient approximation*, J. Roy. Stat. Soc., 66 (2004), pp. 337–356.
- [22] R. M. LARSEN, *Lanczos bidiagonalization with partial reorthogonalization*, (1998).
- [23] J. NAGY, K. PALMER, AND L. PERRONE, *Iterative methods for image deblurring: A matlab object oriented approach*, Numerical Algorithms, 36 (2004), pp. 73–93.
- [24] J. G. NAGY, M. K. NG, AND L. PERRONE, *Kronecker product approximation for image restoration with reflexive boundary conditions*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 829–841.
- [25] D. NYCHKA, B. BAILEY, S. ELLNER, P. HAALAND, AND M. O’CONNEL, *FUNFITS: Data analysis and statistical tools for estimating functions*, in Case Studies in Environmental Statistics, Springer, New York, 1998, pp. 159–179.
- [26] 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.
- [27] 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.
- [28] R. VIO, P. MA, W. ZHONG, J. NAGY, L. TENORIO, AND W. WAMSTEKER, *Estimation of regularization parameters in multiple-image deblurring*, Astronomy and Astrophysics, 423 (2004), pp. 1179–1186.
- [29] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, PA, 2002.