WINNOWED SPECTRAL REGULARIZATION OF INVERSE PROBLEMS

JULIANNE CHUNG†, GLENN EASLEY‡, AND DIANNE P. O’LEARY§

Abstract. Regularization is used in order to obtain a reasonable estimate of the solution to an ill-posed inverse problem. One common form of regularization is to use a filter to reduce the influence of components corresponding to small singular values, perhaps using a Tikhonov least squares formulation. In this work, we break the problem into subproblems with narrower bands of singular values using spectrally defined windows, and we regularize each subproblem individually. We show how to use standard parameter-choice methods, such as the discrepancy principle and generalized cross-validation, in a windowed regularization framework. A perturbation analysis gives sensitivity estimates. We demonstrate the effectiveness of our algorithms on deblurring images and on the backward heat equation.

Key words. ill-posed inverse problems, Tikhonov regularization, multiscale representations, locally windowed decompositions

AMS subject classifications. 65F20, 65F22

DOI. 10.1137/100809787

1. Introduction. Ill-posed inverse problems arise in a variety of scientific and engineering applications, often in the form of first kind integral equations,

\[ b_i \equiv b(s_i) = \int_{\Omega} a(s_i, t) x(t) \, dt + \epsilon_i, \quad i = 1, 2, \ldots, m. \]

These systems are often used to model data obtained after distortions and errors are introduced when measuring an unknown function \( x(t) \). The measured observations, \( b_i \), are made on a discrete mesh \( s_1, s_2, \ldots, s_m \), and \( \epsilon_i \) are random measurement errors with zero mean. The kernel, \( a \), models the distortion process of the system via impulse response functions. In this paper, we assume that these functions are known.

Discretizing (1.1) results in a linear system of equations

\[ b = Ax_{\text{true}} + \epsilon, \]

where \( b \in \mathbb{R}^m \) is the vector of observed data, \( x_{\text{true}} \in \mathbb{R}^n \) is an unknown vector containing values of \( x(t) \) on some mesh \( t_1, t_2, \ldots, t_n \), matrix \( A \in \mathbb{R}^{m \times n}, m \geq n \), is known, and \( \epsilon \in \mathbb{R}^m \) represents noise in the data. The goal in solving the problem is to compute an approximation of \( x_{\text{true}} \), given \( b \) and \( A \).

This is an ill-posed inverse problem, meaning that small perturbations in the data may result in large errors in the solution. Regularization can be used to filter out "hazardous" components (components that highly perturb the estimate) of the
solution in order to compute stable solutions. Standard filtering approaches, such as the truncated singular value decomposition and Tikhonov regularization, are spectral filtering methods that work in the frequency domain of the discretized kernel, defined by the singular value decomposition (SVD) of the matrix $A$. One of the key limitations of current filtering techniques is the use of a single regularization parameter to define the shape of the filter and ultimately determine the quality of the solution. As noted in [8], this global smoothing approach can severely oversmooth local features of the solution.

In this paper, we present a new windowed approach for spectral regularization, where varying amounts of regularization are applied to the observed data at different SVD-based frequency scales. The underlying idea of splitting a signal into different frequency bands is well established in the signal and image processing community [7, 15], but previous investigations in this line of inquiry for inverse problems have focused on partitioning the frequency space independent of the magnitudes of the singular values. Our approach uses operator-dependent windows defined by the singular values. The advantages of such an approach include separation of features in the image and better conditioning within each window or band of frequencies. In addition, we can get improved perturbation bounds compared to standard methods. Furthermore, by allowing varying amounts of regularization in each frequency window, this approach can be robust to high noise levels or colored noise. The cost of these improvements is the need to select multiple regularization parameters.

Some research has been done on selecting multiple regularization parameters. In Belge, Kilmer, and Miller [1], an extension of the L-curve to the L-hypersurface is used for selecting regularization parameters for a problem in which the coarse and fine wavelet coefficients of the desired solution are separately regularized. Regularization on the wavelet coefficients of the solution are also considered in [14]. In [8, 9], Modarresi and Golub use several regularization parameters, where regularization is applied to either overlapping or nonoverlapping subdomains. Their approach differs from ours by partitioning in the data space rather than in the frequency domain. Similar approaches are described in [16, 6]. However, none of these previous approaches considers decompositions in the SVD-based frequency domain of the operator.

The novelty of our approach is in applying regularization to different operator-dependent frequency windows, overlapping or nonoverlapping. Furthermore, we show how to extend standard parameter-choice methods to this windowed spectral framework, include a perturbation analysis for the new framework, and provide numerical results illustrating the improvement over standard methods. Another advantage of our approach is that we allow for arbitrary boundary conditions, which is an important feature for image processing applications. However, it is important to remark that the new approach is applicable only when the size or special structure of the model matrix permits computation of the SVD.

The paper is outlined as follows. In section 2, an SVD analysis of the problem is presented to establish notation. Standard regularization and regularization parameter selection methods are described. The general framework for multiscale regularization is then described in section 3. Operator-dependent decompositions are described, and methods for selecting regularization parameters in a windowed framework are investigated. In section 4, we derive perturbation bounds for the windowed approach to illustrate the improved conditioning that can be achieved. Numerical examples, including both one-dimensional (1-D) and two-dimensional (2-D) examples, are provided in section 5, investigating how the number of windows and the noise level affect the solutions. Concluding remarks and future directions are given in section 6.
2. The Tikhonov least squares formulation. One approach to solving (1.2) is to minimize the $l^2$-norm of the residual:

$$
\min_x \|b - Ax\|_2^2.
$$

It is well known that the solution to this problem can be written in terms of the SVD of $A$:

$$
A = U\Sigma V^T,
$$

where the $m \times n$ matrix $\Sigma = [\hat{\Sigma}]$ is diagonal with entries equal to the singular values

$$
\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.
$$

The singular vectors $u_i$ ($i = 1, \ldots, m$) and $v_i$ ($i = 1, \ldots, n$) are columns of the matrices $U$ and $V$, respectively. The singular vectors are orthonormal, so $U^T U = I_m$ and $V^T V = I_n$. For discrete ill-posed problems, we typically have the following properties:

- The singular values $\sigma_i > 0$ have a clusterpoint at 0 as $m,n \to \infty$.
- There is no noticeable gap in the singular values, and therefore the matrix $A$ should be considered to be full rank.
- The small singular values correspond to oscillatory singular vectors.

Substituting the SVD in (2.1), it can be shown that the (naive inverse) solution to the minimization problem is

$$
\hat{x} = V\begin{bmatrix} \hat{\Sigma}^{-1} & 0 \end{bmatrix} U^T b = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.
$$

From this expression, we see that if highly oscillatory noise is present in $b$, then it will be magnified by division by very small singular values and will greatly perturb the computed solution.

We will make two assumptions:

- The discretization is fine enough that problem (1.2) satisfies the discrete Picard condition: the sequence $|u_i^T b_{\text{true}}|$ decreases to 0 faster than the sequence $\sigma_i$.
- The noise components $\epsilon_j$, $j = 1, \ldots, m$, are uncorrelated, zero mean, and have identical but unknown variance.

Under these assumptions, we typically see the behavior indicated in Figure 2.1: the magnitude of the coefficients $|u_i^T b|$ (stars) eventually stabilizes and provides an estimate of the standard deviation of the noise. It is clear that these later coefficients (high index $i$) contribute mostly error to the naive inverse solution, and they should be suppressed or filtered. One approach to doing this, called Tikhonov regularization, imposes a specific one-parameter filter by solving the problem

$$
\min_x \left\{ \|b - Ax\|_2^2 + \lambda \|x\|_2^2 \right\},
$$

where $\lambda > 0$ is the regularization parameter.\(^1\) The solution to this Tikhonov problem is

$$
x_{\text{tik}}(\lambda) = \left( V\Phi_{\text{tik}}(\lambda) \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \end{bmatrix} U^T \right) b \equiv A_{\lambda}^{-1} b = \sum_{i=1}^{n} \frac{\phi_i u_i^T b}{\sigma_i} v_i,
$$

\(^1\)In the literature, the parameter that we call $\lambda$ is often called $\lambda^2$.  

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
where \( \Phi_{tik}(\lambda) = \text{diag}(\phi_1, \phi_2, \ldots, \phi_n) \), and the filter factors are given by

\[
\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda}
\]

Notice that the filter factors for large singular values are approximately equal to 1, while those corresponding to small singular values are close to zero, as desired. The designation large and small is determined solely by the regularization parameter \( \lambda \).

We define the optimal parameter to be the one that minimizes the mean squared error between the computed and true solution:

\[
MSE(\lambda) = \frac{1}{n} \| x_{tik}(\lambda) - x_{\text{true}} \|^2_2.
\]

This optimal parameter, of course, cannot be computed without knowledge of the true solution, so we need a way to estimate it.

There are a variety of methods that can guide the selection of a good regularization parameter [5]. Here we describe two common approaches. The first relies on an estimate of the noise variance. The variance might be known to the person who collected the data, or it might be estimated from the data. For example, it can be estimated by the median of the finest wavelet coefficients of \( b \) [2], with the regularization parameter computed as [10]

\[
\lambda = \frac{m \sigma^2}{\| b - \bar{b} \|^2_2}.
\]

where \( \sigma^2 \) is the estimate of the noise variance,\(^2\) \( \bar{b} \) is the mean of the observed data, and \( 1 \) is the vector with every entry equal to one.

\(^2\) The noise variance \( \sigma^2 \) should not be confused with the singular values \( \sigma_i \).
The second approach we consider is the generalized cross-validation (GCV) method from Golub, Heath, and Wahba [3], where $\lambda$ is computed to minimize

$$GCV(\lambda) = \frac{n \| (I - AA^\dagger_\lambda) b \|^2_2}{\left[ \text{trace}(I - AA^\dagger_\lambda) \right]^2}$$

$$= \frac{n \left( \sum_{i=1}^{n} \left( \frac{\lambda}{\sigma_i^2 + \lambda} \right)^2 c_i^2 + \sum_{i=n+1}^{m} c_i^2 \right)}{(m-n) + \sum_{i=1}^{n} \frac{\lambda}{\sigma_i^2 + \lambda}}$$

(2.7)

where $A^\dagger_\lambda$ is defined in (2.3) and $c_i = u_i^T b$.

As mentioned in the introduction, a severe limitation of Tikhonov regularization is its global, uniform smoothing. With only one parameter, the solution may be oversmoothed, especially in the case of large noise levels. We will improve upon the method with a windowed spectral decomposition, described in the next section.

3. Windowed spectral decomposition. In the windowed approach, the problem is partitioned based on the frequency space of the operator, and multiple regularization parameters allow local regularization in the frequency domain. The result is that each subproblem is better conditioned, and regularization can be tuned to each frequency subdomain. In this section we describe the mathematical formulation of the windowed approach, provide a variety of techniques for determining spectral windows, and extend standard approaches for choosing regularization parameters, described in section 2, to the windowed case.

To simplify notation, let $y = V^T x$ and $c = U^T b$. Then the original problem, $Ax \approx b$, can be written as

$$\Sigma y \approx c.$$  

(3.1)

Let $w^{(j)} \in \mathbb{R}^n$, $j = 1, 2, \ldots, p$, be nonnegative frequency weights such that

$$\sum_{j=1}^{p} w^{(j)} = 1 \quad \text{and} \quad \sum_{j=1}^{p} W^{(j)} = I,$$

(3.2)

where $W^{(j)} = \text{diag}(w^{(j)})$. Specific definitions of frequency weights will be given in section 3.1. The basic idea of the windowed approach is to divide the frequency space using weights or windows and allow the regularization parameter to change in each window. Let $\lambda^{(j)}$ be the regularization parameter for window $W^{(j)}$; then the reconstructed image can be computed as

$$x_{\text{win}} = \sum_{j=1}^{p} V (\Sigma^{(j)} \Sigma + \lambda^{(j)} I)^{-1} W^{(j)} \Sigma^{T} c,$$

(3.3)

or equivalently,

$$x_{\text{win}} = V \Phi_{\text{win}} |\Sigma^{-1}, 0 | c,$$

(3.4)

where

$$\Phi_{\text{win}} = \sum_{j=1}^{p} \Phi_{\text{tik}}(\lambda^{(j)}) W^{(j)}.$$
In the case $m = n$, windowed spectral Tikhonov regularization is equivalent to applying Tikhonov regularization to each of the weighted right-hand sides $W^{(j)} c$. That is, we solve $p$ least squares problems

$$
\min_y \left\{ \| W^{(j)} c - \Sigma y \|_2^2 + \lambda^{(j)} \| y \|_2^2 \right\},
$$

each with a different regularization parameter.

3.1. Defining the windows. In this section, we define some windows $w^{(j)}$ by using the SVD to partition the frequency space into intervals so that each subproblem is better conditioned.

To illustrate our ideas, we refer to the image deblurring problem of Figure 3.1, where the image on the left has been blurred with a $7 \times 7$ boxcar blur\(^3\) (reflexive boundary conditions), followed by the addition of random noise to obtain the blurred, noisy image on the right. The blurred-signal-to-noise ratio is

$$\text{BSNR} = 10 \cdot \log_{10} \left( \frac{\| b \|_2^2}{\| \epsilon \|_2^2} \right) = 25.$$ 

Many important blurring functions, including the boxcar blur, are separable, allowing the matrix $A$ to be expressed as a Kronecker product $A_1 \otimes A_2$. This means that the singular values of $A$ are the products of all singular values of $A_1$ with all singular values of $A_2$. We display these singular values in Figure 3.2, arranged so that each row corresponds to a particular singular value of $A_1$ and each column corresponds to one of $A_2$. For this particular example, the singular vectors are defined by the 2-D discrete cosine transform (DCT) [11], but our algorithm does not assume this.

The most generic (operator-independent) way to divide the frequency space is to use a dyadic subdivision, illustrated in the upper row of Figure 3.3 for $p = 5$ windows, where black and white represent components of $w_i$ equal to 0 and 1, respectively ($i = 1, \ldots, n$). We call these weights dyadic Shannon windows because of their relation to filters used in the classical Shannon sampling theorem [7]. The corresponding images in the bottom row partition the blurred airplane into the five subimages $UW^{(j)} c$; summing these images gives the original blurred image, and our goal is to regularize each of these images separately.

---

\(^3\)Boxcar blur is when each pixel in the image is replaced by the average of pixel values within a square neighborhood.
Comparing Figures 3.2 and 3.3, we see that the dyadic Shannon windows do not group singular values that are closest in magnitude. Next we define some Shannon windows that do this by setting $p + 1$ partition values $\tau^{(0)} \leq \cdots \leq \tau^{(p)}$, where $\tau^{(0)} < \sigma_{\min}$ and $\tau^{(p)} \geq \sigma_{\max}$. Define the $j$th weight vector $w^{(j)}$ to have entries

$$
    w^{(j)}_i = \begin{cases} 
    1 & \text{for } \tau^{(j-1)} < \sigma_i \leq \tau^{(j)}, \\
    0 & \text{otherwise.}
    \end{cases}
$$

FIG. 3.2. Singular values for the blur of the airplane (top) and four equally spaced level curves for the singular values (bottom). The figures on the left use a linear scale, while those on the right use a log scale.

FIG. 3.3. Dyadic Shannon windows (top) and the corresponding subimages $UW^{(j)}e$ of the blurred airplane, $j = 1, \ldots, 5$. 
Fig. 3.4. Shannon windows constructed using linearly spaced thresholding of the singular values (top row), corresponding images (middle row), and Picard plots (bottom row). The solid line and the jagged line correspond to singular values and coefficients $|u_i^T b|$ represented in each window. The horizontal axis on the Picard plot for the fifth window is sampled at intervals of 100 due to the large number of singular values in that window.

If we choose the $p - 1$ linearly spaced (or log-spaced) values between the largest and smallest singular values, then for $p = 5$, our windows correspond to those displayed in Figure 3.2(c) (or (d)) for the boxcar blur example.

For linear-Shannon windows, Figure 3.4 displays the five frequency windows (top row) and the corresponding images (middle row). One of the nice features of using these Shannon windows, compared to dyadic Shannon windows, is the availability of the Picard plot (bottom row) for each of the windows. Similar to Figure 2.1, the solid line corresponds to singular values and the jagged line corresponds to coefficients $|u_i^T b|$ that are represented in each window. Here, the horizontal axis corresponds to index $i$. Simply gluing together the Picard plots will give the plot for the original image. This gives us a good understanding regarding the advantages of applying different amounts of regularization in each window. We quantify precisely some of these advantages in the next theorem.

**Theorem 3.1.** Let $\kappa(A) = \sigma_{\max}/\sigma_{\min}$ be the condition number of $A$. For Shannon windows, define

$$\kappa^{(j)} = \frac{\sigma^{(j)}_{\max}}{\sigma^{(j)}_{\min}},$$

where $\sigma^{(j)}_{\min}$ and $\sigma^{(j)}_{\max}$ are the minimum and maximum singular values in the $j$th window. Then for a linear partition of the singular values,

$$\kappa^{(j)} \leq \frac{1}{p} \kappa(A), \quad j = 1, 2, \ldots, p.$$
and for a logarithmic partition of the singular values,
\[ \kappa^{(j)} \leq (\kappa(A))^{\frac{j}{p}}, \quad j = 1, 2, \ldots, p. \]

Depending on the shape and spread of the spectrum of the problem, either the linear or the logarithmic partition may be the preferred choice. In either case, our upper bound on the spread of the singular values in each window explains why a single regularization parameter per window makes sense.

It is possible to use windows more general than Shannon windows. The Shannon windows are well localized in the frequency domain but have slow decay in the data domain. This means frequency tapering of these windows by regularization tends to create more pronounced Gibbs ringing in the data domain when compared to using windows that are less well localized in the frequency domain. Thus, we also suggest using alternative windows that have faster decay in the data domain and which will allow for nonuniform partitioning in the frequency domain. For example, we define cosine windows based on the midpoints
\[ \tau^{(j)}_m = \frac{\tau^{(j-1)}_m + \tau^{(j)}_m}{2}, \quad j = 2, \ldots, p - 1. \]

Then, the \( j \)th cosine window can be written as

\[
\begin{aligned}
w^{(j)}_i = \begin{cases} 
\cos^2 \left( \frac{\tau^{(j)} - \sigma_i}{\tau^{(j)}_m - \tau^{(j-1)}_m} \right) & \text{for } \tau^{(j-1)}_m < \sigma_i \leq \tau^{(j)}_m, \\
\cos^2 \left( \frac{\sigma_i - \tau^{(j)}_m}{\tau^{(j+1)}_m - \tau^{(j)}_m} \right) & \text{for } \tau^{(j)}_m < \sigma_i \leq \tau^{(j+1)}_m, \\
0 & \text{otherwise.}
\end{cases}
\end{aligned}
\]

A variety of definitions may be made for the first and \( p \)th windows. We choose

\[
\begin{aligned}
w^{(1)}_i = \begin{cases} 
1 & \text{for } \tau^{(0)} < \sigma_i \leq \tau^{(1)}_m, \\
\cos^2 \left( \frac{\sigma_i - \tau^{(1)}_m}{\tau^{(2)}_m - \tau^{(1)}_m} \right) & \text{for } \tau^{(1)}_m < \sigma_i \leq \tau^{(2)}_m, \\
0 & \text{otherwise,}
\end{cases}
\end{aligned}
\]

with a similar definition for the \( p \)th window. We illustrate these windows in Figure 3.5. Notice that the particular choice of overlap results in the important property that the window coefficients for each singular value sum to one, and that the nonsymmetric definition of the cosine window allows for variable spacing of the thresholds.

Figure 3.6 illustrates cosine windows with linearly spaced partitions, along with the corresponding images and Picard plots, for our airplane example. In contrast to the Shannon windows of Figure 3.4 where the Picard plots can just be pieced together, the cosine window plots should be appropriately summed to get the original Picard plot. To aid in visualizing this sum, the corresponding window components are displayed in the dotted line as a function of the sorted singular value index. The individual Picard plots also shed light on why the cosine windows make sense: although each window has a wider spread of singular values than in Shannon windowing, the Picard decay condition is much stronger for all windows except the last.

Note again that in contrast to the dyadic Shannon windowing, this newly proposed windowing depends on the operator and it preserves the Picard condition on the subproblems, an important tool in the choice of the regularization parameters. Note
Fig. 3.5. Cosine windows. The horizontal axis represents the singular values. For a particular interval, (a) compares the Shannon window, where the window components are either 0 or 1, to the cosine window. (b) illustrates the overlap of windows that allows the components to sum to 1 in the interval [0.2, 0.6] (unsymmetric spacing).

![Fig. 3.5](image)

Fig. 3.6. The cosine windows constructed using linearly spaced thresholding of the singular values (top row), corresponding images (middle row), and Picard plots (bottom row). In the Picard plots where the horizontal axis corresponds to index $i$, the solid line corresponds to singular values, the jagged line represents coefficients $|u_i^T b|$, and the dotted line denotes the corresponding cosine window components. The horizontal axis on the Picard plot for the fifth window is sampled at intervals of 100 due to the large number of singular values included in that window.

![Fig. 3.6](image)

also that although the blur in our airplane example had Kronecker structure which we used in drawing our figures, our method does not depend in any way on having such a structure.

In the extreme case, taking one window corresponds to standard Tikhonov regularization, and taking $n$ windows allows each pixel to have its own regularization parameter. The latter case is essentially equivalent to a Wiener filtering approach where each penalty parameter per pixel is determined by an approximate power spec-
trum of the unblurred image. We shall see in the next section that there is a tradeoff between the flexibility of multiple windows and the ability to determine the regularization parameters effectively.

3.2. Selecting regularization parameters in a windowed framework. This section describes how standard approaches for selecting regularization parameters, described in section 2, can be extended to the case of windowed spectral regularization.

The windowed approach requires computing the reconstruction in (3.3), and the problem is most easily understood in the transformed coordinates (3.1) defined by the SVD. For the special case of Shannon windows, the regularized problem can be written as

\[
\begin{bmatrix}
\Sigma \\
L
\end{bmatrix} y \approx \begin{bmatrix} c \\
o \end{bmatrix},
\]

or more precisely as

\[
\begin{bmatrix}
\Sigma^{(1)} \\
\vdots \\
\Sigma^{(p)} \\
\sqrt{\lambda^{(1)}} I \\
\vdots \\
\sqrt{\lambda^{(p)}} I
\end{bmatrix} \begin{bmatrix}
y^{(1)} \\
\vdots \\
y^{(p)}
\end{bmatrix} \approx \begin{bmatrix} c^{(1)} \\
\vdots \\
c^{(p)} \\
o \\
\vdots \\
o
\end{bmatrix},
\]

where \(c^{(j)}\) is the subvector of \(c\) corresponding to the singular values in the \(j\)th window. Therefore, the solution to (3.9) can be found by solving \(p\) separate minimization problems,

\[
\min_{y^{(j)}} \| \Sigma^{(j)} y^{(j)} - c^{(j)} \|^2_2 + \lambda^{(j)} \| y^{(j)} \|^2_2.
\]

A similar approach for regularization was described in Modarresi and Golub [8], but a key difference is that we are applying Tikhonov regularization in the frequency domain, rather than directly in the data space. Thus, we are able to take advantage of the frequency-based separations to give a multiscaled decomposition that separates features in the image.

Now, we describe, as examples, two approaches for selecting the regularization parameters \(\lambda^{(j)}\). First, we extend the variance estimation idea of (2.5). For Shannon windows (the nonoverlapping case), let \(\sigma^2\) be the estimate of the noise variance computed using vector \(c\). We define

\[
\lambda^{(j)} = \frac{n^{(j)} \sigma^2}{\| c^{(j)} - \bar{c}^{(j)} \|^2_2},
\]

where \(n^{(j)}\) is the number of singular values in the \(j\)th window and \(\bar{c}^{(j)}\) is the mean of \(c^{(j)}\). We might also use this parameter for overlapping windows. For example, the regularization parameter computed for the Shannon window shown in Figure 3.5(a) could be used for the corresponding cosine window. By decoupling the problem, we can estimate regularization parameters independently for each window and perform reconstructions in parallel.
Second, we extend the GCV method to the windowed case, using the original derivation of the GCV function [3], but in the spectral domain. The derivation is provided in the appendix, and the result is stated here.

**Theorem 3.2.** Let $w^{(j)}_i$ be the $i$th entry of window $w^{(j)}$ and $\lambda$ be the vector of regularization parameters, $\lambda^{(j)}$. The GCV function for windowed spectral Tikhonov regularization can be written as

$$GCV(\lambda) = \frac{1}{m} \left[ \sum_{i=1}^{n} \left( \sum_{j=1}^{p} \frac{1}{\gamma^{(j)}} \left( \frac{\gamma^{(j)}}{p} - \hat{\gamma}^{(j)} + 1 - \frac{\sigma_i^2 w^{(j)}_i}{\sigma_i^2 + \lambda^{(j)}} \right) \right)^2 \sigma_i^2 \right] + \sum_{i=n+1}^{m} \left[ \sum_{j=1}^{p} \frac{1}{\gamma^{(j)}} \left( \frac{\gamma^{(j)}}{p} - \hat{\gamma}^{(j)} + 1 \right) \right]^2 \sigma_i^2 \right],$$

(3.12)

where

$$\gamma^{(j)} = \frac{1}{m} \left( (m-n) + \sum_{i=1}^{n} \frac{\lambda^{(j)}}{\sigma_i^2 + \lambda^{(j)}} \right)$$

and

$$\hat{\gamma}^{(j)} = \frac{1}{m} \left( (m-n) + \sum_{i=1}^{n} \left( 1 - \frac{\sigma_i^2 w^{(j)}_i}{\sigma_i^2 + \lambda^{(j)}} \right) \right).$$

Notice that this is a function of $p$ variables, so optimization could become difficult. A good initial guess for the optimization would be beneficial and can be obtained from parameters for the corresponding nonoverlapping Shannon windows.

For Shannon windows, we determine the $p$ parameters by minimizing the GCV function for each of the $p$ problems in (3.10). From (2.6), the $j$th function to be minimized is

$$GCV(\lambda) = \frac{\sum_{\sigma_i \in D^{(j)}} \left( \frac{\lambda}{\sigma_i^2 + \lambda} \right)^2 \sigma_i^2}{\left( \sum_{\sigma_i \in D^{(j)}} \frac{\lambda}{\sigma_i^2 + \lambda} \right)^2},$$

(3.13)

where $D^{(j)}$ denotes the set of singular values represented in window $j$. These one-parameter GCV functions are easy to evaluate and can be minimized using standard optimization routines. The possibility of decoupling the multiparameter GCV function into 1-D GCV functions was also suggested in [9], and there is potential parallelism in this formulation.

As in (2.4), we define the (generally uncomputable) optimal parameters to be the minimizer of the mean squared error

$$MSE(\lambda) = \left\| \sum_{j=1}^{p} W^{(j)} \Phi_{tik}(\lambda^{(j)}) \Sigma^{-1} U^T b - V^T x_{true} \right\|_2^2.$$

(3.14)
For Shannon windows, (3.14) simplifies to \( p \) one-parameter minimization problems
\[
\lambda^{(j)} = \arg \min_{\lambda} \| W^{(j)} \Phi_{\text{tik}}(\lambda) \Sigma^{-1} U^T b - W^{(j)} V^T x_{\text{true}} \|^2_2.
\]

Other regularization parameter selection approaches could be used, such as the residual periodogram approach \[13\], where the goal would be to find the set of parameters that maximize the quality of the periodogram. However, this would also require one to solve a \( p \)-dimensional optimization problem for the overlapping case.

We conclude our airplane example by showing in Figure 3.7 the absolute error images compared to the true image (with inverted colormap where white corresponds to zero error) of using GCV to determine regularization parameters in the case of standard Tikhonov, dyadic Shannon windows from Figure 3.3, linearly spaced Shannon windows from Figure 3.4, and linearly spaced cosine windows from Figure 3.6. Five windows were used in all windowed approaches, and the computed SNR for the corresponding reconstructed image is provided in each case. More systematic numerical results are presented in section 5.

![Absolute error images and SNR of reconstructions for the airplane example using GCV to determine regularization parameters.](image)

**Fig. 3.7.** Absolute error images (where white corresponds to zero error) and SNR of reconstructions for the airplane example using GCV to determine regularization parameters.

4. **Perturbation analysis.** In this section, we consider the sensitivity of the solution to perturbations in the data and compute bounds for windowed Tikhonov regularization. For ease of presentation, we first derive bounds for the Shannon windows and then extend the analysis to overlapping windows, such as the cosine windows.

To begin, assume two nonoverlapping Shannon windows with frequency weights \( W^{(1)} \) and \( W^{(2)} \). Let \( \mathcal{D}^{(1)} \) and \( \mathcal{D}^{(2)} \) denote the sets of singular values represented in the two windows, with corresponding regularization parameters \( \lambda^{(1)} \) and \( \lambda^{(2)} \). Using this partition, (3.1) can be formulated as
\[
\begin{bmatrix}
\Sigma^{(1)} \\
\Sigma^{(2)}
\end{bmatrix}
\begin{bmatrix}
y^{(1)} \\
y^{(2)}
\end{bmatrix}
\approx
\begin{bmatrix}
c^{(1)} \\
c^{(2)}
\end{bmatrix},
\]
and the \( i \)th component of the Tikhonov solution vector in the SVD-based frequency domain is

\[
y_i = \left( \frac{\sigma_i}{\sigma_i^2 + \lambda} \right) c_i,
\]

where

\[
\lambda = \begin{cases} 
\lambda^{(1)} & \text{for } \sigma_i \in \mathcal{D}^{(1)} , \\
\lambda^{(2)} & \text{for } \sigma_i \in \mathcal{D}^{(2)} .
\end{cases}
\]

Now the perturbed problem can be written as

\[
\begin{bmatrix} \Sigma^{(1)} \\ \Sigma^{(2)} \end{bmatrix} \begin{bmatrix} \tilde{y}^{(1)} \\ \tilde{y}^{(2)} \end{bmatrix} \approx \begin{bmatrix} c^{(1)} \\ c^{(2)} \end{bmatrix} + \begin{bmatrix} \delta c^{(1)} \\ \delta c^{(2)} \end{bmatrix},
\]

where \( \delta c \) are additional errors in the data. The \( i \)th component of the solution vector for the perturbed problem is

\[
\tilde{y}_i = \left( \frac{\sigma_i}{\sigma_i^2 + \lambda} \right) (c_i + \delta c_i),
\]

with \( \lambda \) as in (4.2). Using (4.1) and (4.3), the componentwise difference can be written as

\[
|y_i - \tilde{y}_i|^2 = \left( \frac{\sigma_i}{\sigma_i^2 + \lambda} \right)^2 |\delta c_i|^2.
\]

When \( \lambda > 0 \), the first term in (4.4), taken as a function of \( \sigma \), has its maximum for \( \sigma = \sqrt{\lambda} \), and the maximum value is \( 1/(4\lambda) \). The maximum value over all \( i \) can be written as

\[
f = \begin{cases} 
\frac{1}{4\lambda} & \text{if } \sigma_{\min} \leq \sqrt{\lambda} \text{ and } \sqrt{\lambda} \leq \sigma_{\max} , \\
\left( \frac{\sigma_{\min}}{\sigma_{\min} + \lambda} \right)^2 & \text{if } \sigma_{\min} > \sqrt{\lambda} , \\
\left( \frac{\sigma_{\max}}{\sigma_{\max} + \lambda} \right)^2 & \text{if } \sigma_{\max} < \sqrt{\lambda} .
\end{cases}
\]

Now, considering each set of singular values \( \mathcal{D}^{(1)} \) and \( \mathcal{D}^{(2)} \) separately and defining \( \sigma_{\max}^{(1)} = \max_{\mathcal{D}^{(1)}} \sigma_i \) and \( \sigma_{\min}^{(1)} = \min_{\mathcal{D}^{(1)}} \sigma_i \) (similarly for \( \sigma_{\max}^{(2)} \) and \( \sigma_{\min}^{(2)} \)), we get the following bounds:

\[
\| y^{(1)} - \tilde{y}^{(1)} \|_2^2 \leq f^{(1)} \| \delta c^{(1)} \|_2^2 ,
\]

\[
\| y^{(2)} - \tilde{y}^{(2)} \|_2^2 \leq f^{(2)} \| \delta c^{(2)} \|_2^2 ,
\]

where \( f^{(1)} \) is from (4.5) with \( \sigma_{\max}^{(1)}, \sigma_{\min}^{(1)} \) and \( \lambda^{(1)} \) (similarly for \( f^{(2)} \)). Thus, we can bound the relative change in the solution as

\[
\frac{\| y - \tilde{y} \|_2^2}{\| y \|_2^2} = \frac{\| y^{(1)} - \tilde{y}^{(1)} \|_2^2 + \| y^{(2)} - \tilde{y}^{(2)} \|_2^2}{\| y \|_2^2} \leq \left( f^{(1)} \| \delta c^{(1)} \|_2^2 + f^{(2)} \| \delta c^{(2)} \|_2^2 \right) \frac{1}{\| y \|_2^2} \leq \sigma_{\max}^2 \max(f^{(1)}, f^{(2)}) \| \delta c \|_2^2 \| c \|_2^2 .
\]
where the last inequality uses the property that
\[ ||c||_2^2 \leq ||\Sigma||_2^2 ||y||_2^2.\]

Later in this section, we extend this result to multiple windows, but first we make the following observations:

- If we take the maximum over all \( i \) in (4.5), then we get the same bound as standard Tikhonov with a single parameter, \( \lambda \).
- If we argue that, for independent and identically distributed normal noise, \( E(||\delta c^{(1)}||_2^2) = \frac{1}{2}E(||\delta c||_2^2) \), then the windowed approach with two windows with \( |D^{(1)}| = |D^{(2)}| \) gives, on average, half the bound that Tikhonov gives for similar parameters.
- Since the bound for Tikhonov depends only on \( \lambda \), the optimal bound can be attained when the optimal choice for \( \lambda \) is used. However, in addition to the now two regularization parameters, the bound in the windowed case also depends on the division of singular values. To obtain a good bound, it is necessary to consider the distribution of singular values and to understand how the two parameters behave relative to the optimal Tikhonov parameter.

In the following theorem, we extend the perturbation result for Shannon windows to multiple windows.

**Theorem 4.1.** Let \( D^{(j)} \) denote the nonoverlapping set of singular values contained in window \( j \), for \( j = 1, \ldots, p \). Define \( \sigma_{\text{min}}^{(j)} \) and \( \sigma_{\text{max}}^{(j)} \) to be the minimum and maximum singular values in \( D^{(j)} \), respectively. Given regularization parameters \( \lambda^{(j)} \) for each window \( D^{(j)} \), the relative change in the solution \( y \) when the data is perturbed by \( \delta c \) is bounded by

\[
\frac{||y - \tilde{y}||_2^2}{||y||_2^2} \leq \sigma_{\text{max}}^{(j)} \max_j (f^{(j)}(\lambda^{(j)})) \frac{||\delta c||_2^2}{||c||_2^2},
\]

where, analogous to (4.5),

\[
f^{(j)} = \begin{cases} 
\frac{1}{\lambda^{(j)}} & \text{if } \sigma_{\text{min}}^{(j)} \leq \sqrt{\lambda^{(j)}} \text{ and } \sqrt{\lambda^{(j)}} \leq \sigma_{\text{max}}^{(j)}, \\
\left( \frac{\sigma_{\text{min}}^{(j)}}{\sigma_{\text{max}}^{(j)} + \lambda^{(j)}} \right)^2 & \text{if } \sigma_{\text{min}}^{(j)} > \sqrt{\lambda^{(j)}}, \\
\left( \frac{\sigma_{\text{max}}^{(j)}}{\sigma_{\text{max}}^{(j)} + \lambda^{(j)}} \right)^2 & \text{if } \sigma_{\text{max}}^{(j)} < \sqrt{\lambda^{(j)}}.
\end{cases}
\]

Next we consider an extension to the case of overlapping windows. For simplicity of illustration, we once again use two windows, which overlap in a given interval of singular values. Since nonoverlapping regions were already considered above, here we consider only regions of overlap. That is, assume \( w_i^{(1)} \) and \( w_i^{(2)} \) are the values of the \( i \)th component of windows \( w^{(1)} \) and \( w^{(2)} \), respectively, and \( 0 < w_i^{(1)}, w_i^{(2)} < 1 \). Then the \( i \)th component of the Tikhonov solution can be written as

\[
y_i = \left( \frac{w_i^{(1)} \sigma_i}{\sigma_i^2 + \lambda^{(1)}} + w_i^{(2)} \frac{\sigma_i}{\sigma_i^2 + \lambda^{(2)}} \right) c_i
\]

and the componentwise difference is

\[
|y_i - \tilde{y}_i|^2 = \left( \frac{w_i^{(1)} \sigma_i}{\sigma_i^2 + \lambda^{(1)}} + w_i^{(2)} \frac{\sigma_i}{\sigma_i^2 + \lambda^{(2)}} \right)^2 |\delta c_i|^2.
\]
Thus, a tight bound for the overlapping region can be obtained by taking the maximum over all \( i \):

\[
\frac{\| y - \tilde{y} \|_2^2}{\| y \|_2^2} \leq \sigma_{\text{max}}^2 \max_{i} \left( w_i^{(1)} \frac{\sigma_i}{\sigma_i^2 + \lambda^{(1)}} + w_i^{(2)} \frac{\sigma_i}{\sigma_i^2 + \lambda^{(2)}} \right)^2 \frac{\| \delta c \|_2^2}{\| c \|_2^2}.
\]

We remark that the above bound heavily depends on the choice of windows. Even with a simple functional representation of the windows, it is not easy to get an explicit formula.

Notice that by using the fact that \( w_i^{(1)} + w_i^{(2)} = 1 \) and by properties of convexity, we can get the bound

\[
\frac{\| y - \tilde{y} \|_2^2}{\| y \|_2^2} \leq \sigma_{\text{max}}^2 \max_{i} \left( w_i^{(1)} f^{(1)}(1) + w_i^{(2)} f^{(2)}(1) \right) \frac{\| \delta c \|_2^2}{\| c \|_2^2}.
\]

However, taking the maximum of the convex combination in the above bound actually gives the bound in (4.10) for Shannon windows. Thus, the perturbation bound for overlapping windows is at least as good as for the Shannon windows.

In the following theorem, we extend and summarize the result for overlapping windows.

**Theorem 4.2.** Assume \( p \) general windows where there are at most two windows which overlap for each singular value. Let \( D^{(j)} \) denote the set of singular values that are represented in window \( j \), and let \( \lambda^{(j)} \) be the corresponding regularization parameter. Then for each singular value, \( \sigma_i \), which is in at most two sets \( D^{(A)} \) and \( D^{(B)} \), the componentwise difference can be written as

\[
\left( \frac{w_i^{(A)}}{\sigma_i^2 + \lambda^{(A)}} + \frac{w_i^{(B)}}{\sigma_i^2 + \lambda^{(B)}} \right)^2 \| \delta c \|_2^2.
\]

Taking the maximum over all \( i \), the relative difference is bounded as

\[
\frac{\| y - \tilde{y} \|_2^2}{\| y \|_2^2} \leq \sigma_{\text{max}}^2 \max_{i} \left( \frac{w_i^{(A)}}{\sigma_i^2 + \lambda^{(A)}} + \frac{w_i^{(B)}}{\sigma_i^2 + \lambda^{(B)}} \right)^2 \frac{\| \delta c \|_2^2}{\| c \|_2^2}.
\]

5. **Numerical results.** In this section, we present some numerical results illustrating the benefits of regularizing ill-posed inverse problems in a windowed spectral framework. Two test problems are considered here: a 1-D “heat” example and an image deblurring problem. A variety of window functions can be used for defining frequency windows, but we limit our study to these four: Shannon windows and cosine windows, with both linear and logarithmic spacing.

5.1. 1-D example. The 1-D example we consider is an inverse heat equation, generated using the MATLAB package Regularization Tools [4]. The problem is a Volterra integral equation of the first kind, whose model is given in (1.1), with \( \Omega = [0, 1] \) and kernel \( a(s, t) = k(s - t) \), where

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

The discretization of this problem results in an \( n \times n \) matrix \( A \), true solution vector \( x_{\text{true}} \) defined in [4], and (noise-free) observation vector \( b_{\text{true}} \), computed as \( b_{\text{true}} = \)
Table 5.1
SNR and relative error comparisons for the 1-D example.

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR</th>
<th>Relative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VE</td>
<td>GCV</td>
</tr>
<tr>
<td>Tikhonov</td>
<td>22.87</td>
<td>22.90</td>
</tr>
<tr>
<td>Linear-Shannon</td>
<td>29.66</td>
<td>23.57</td>
</tr>
<tr>
<td>log-Shannon</td>
<td>20.07</td>
<td>22.92</td>
</tr>
<tr>
<td>Linear-cosine</td>
<td>29.77</td>
<td>23.63</td>
</tr>
<tr>
<td>log-cosine</td>
<td>18.08</td>
<td>28.68</td>
</tr>
</tbody>
</table>

Ax_{true}. The observed, noisy data \( b = b_{true} + \varepsilon \) is constructed by including a randomly generated vector of samples from a normal distribution, with zero mean. In this example, we take \( n = 512 \) with a BSNR of 10 dB, which makes the standard deviation approximately .001.

We use \( p = 2 \) windows, dividing the singular values in half using either a linear or a log scale. Table 5.1 shows the resulting SNRs

\[
\text{SNR}(x) = 10 \cdot \log_{10} \left( \frac{\|x_{true}\|_2^2}{\|x - x_{true}\|_2^2} \right)
\]

and relative errors \( \|x_{true} - x\|_2^2 / \|x_{true}\|_2^2 \) for various parameter selection methods. Results that are better than the standard Tikhonov approach are noted in bold. We observe that for the optimal regularization parameter, the windowed approach is at least as good as the standard Tikhonov approach in all cases, since choosing both parameters equal to the optimal Tikhonov parameter gives the standard Tikhonov approach.

This example illustrates the potential benefits of using windowed spectral regularization. The optimal choice of the regularization parameter cannot be computed without knowing the true solution, but in all cases except for log-Shannon windows, significant improvement was obtained relative to standard Tikhonov, using parameters chosen by variance estimation (VE), found in (2.5), or GCV. Thus, using only one regularization parameter, even the best parameter, unnecessarily limits the reconstruction accuracy that one can achieve. The linear-Shannon and linear-cosine window methods are consistently better than the standard Tikhonov method, and when using the new GCV function, all of our windowing methods perform better. The best result using the new GCV function was obtained by using log-cosine windows, and the best result using the VE approach was obtained using the linear-cosine windows. In addition, notice that the log-cosine windowing method performed significantly better when the optimal parameter was found. Figure 5.1(a) displays a portion of this reconstruction, comparing it with standard GCV Tikhonov. It is evident that the windowed approach gives a better approximation to the true solution. The noticeable difference in the computed filter factors is shown in Figure 5.1(b).

5.2. 2-D example. We return to our image deblurring application, this time reconstructing the true image (“Elaine”) in Figure 5.2(a) from the blurred, noisy image in Figure 5.2(b). We used the blur shown in Figure 5.2(c) which combines a 21 × 21 boxcar blur and a nonsymmetric Gaussian blur. We assumed reflexive boundary conditions and added Gaussian noise to make BSNR = 25dB. Results using \( p = 2 \) windows can be found in Table 5.2. The linear-Shannon, linear-cosine, and log-cosine windows perform better than standard Tikhonov. The log-Shannon windows did not
Fig. 5.1. Results for 1-D heat example: (a) compares the reconstruction for standard Tikhonov to windowed Tikhonov, and (b) contains the computed filter factors as a function of the singular values.

Fig. 5.2. 2-D image deblurring example: The goal is to reconstruct an approximation of the true image in (a), given the blurred, noisy image in (b). We use a spatially invariant blur which combines a boxcar blur and a nonsymmetric Gaussian blur. The blur applied to a point source is shown in (c).

perform much better than Tikhonov, and we omit it in our tabulated results. Absolute error images (with inverted colormap) along with the SNR of the corresponding reconstructed image can be found in Figure 5.3. Notice that the windowed approaches tend to localize the errors better than the standard Tikhonov approach. As in the 1-D problem, the linear-Shannon and linear-cosine windowing methods with the GCV choice of parameters work well, as does the log-cosine windowing method.

To demonstrate the robustness of the algorithm, we repeated the experiment with 100 different noise realizations (BSNR = 25dB), recording the resulting SNRs for the
Table 5.2
SNR and relative error comparisons for the 2-D example (BSNR = 25dB) with p = 2.

<table>
<thead>
<tr>
<th>Window Type</th>
<th>SNR</th>
<th>VE</th>
<th>GCV</th>
<th>Optimal</th>
<th>SNR</th>
<th>VE</th>
<th>GCV</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tikhonov</td>
<td>23.44</td>
<td>21.77</td>
<td>25.06</td>
<td>6.73e-02</td>
<td>8.16e-02</td>
<td>5.58e-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyadic-Shannon</td>
<td>12.27</td>
<td>22.99</td>
<td>25.32</td>
<td>2.44e-01</td>
<td>7.11e-02</td>
<td>5.42e-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear-Shannon</td>
<td>22.48</td>
<td>24.91</td>
<td>25.31</td>
<td>7.52e-02</td>
<td>5.68e-02</td>
<td>5.42e-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear-cosine</td>
<td>22.85</td>
<td>24.81</td>
<td>25.33</td>
<td>7.20e-02</td>
<td>5.75e-02</td>
<td>5.42e-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-cosine</td>
<td>24.66</td>
<td>25.00</td>
<td>25.43</td>
<td>5.85e-02</td>
<td>5.62e-02</td>
<td>5.35e-02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5.3. Absolute error images (where white corresponds to zero error) and SNR of reconstructions using different windows and GCV-chosen regularization parameters.

100 reconstructions. In Figure 5.4 we display the results using histograms, where the horizontal axis corresponds to SNR in dB. The histograms show that results from the SVD-based windows are consistently superior to those from both Tikhonov and dyadic Shannon windows.

Next we study the effect of varying \( p \), the number of windows. Table 5.3 presents the results for linear-Shannon windows. We can see that, given the optimal regularization parameters, more windows result in better solutions, but when we have to estimate the parameters from the data, using VE or GCV, then there is a point of diminishing returns; eventually there is not enough discriminating information in each windowed image to determine good parameters.

In the next study, we investigate the effects of varying the noise level. We use \( p = 4 \) windows and linear-Shannon windows. Results for varying noise levels are presented in Table 5.4. The improvement that can be obtained with the windowed approach is more significant in problems with high noise level.

Finally, we provide an example where noise in the image is not Gaussian. That is, we consider noise that has a frequency distribution in which some frequencies are
represented more prominently than others. This is often referred to as colored noise in the signal processing literature [12]. We generated colored noise by applying a low pass filter to standard Gaussian noise and scaling it to the appropriate standard deviation. Then we computed reconstructions with $p = 2$. Results are presented in Table 5.5. An assumption of Gaussian noise is not valid for this problem, but it is a good approximation in separate frequency windows. Therefore, we expect that our windowed approach will provide an even greater improvement over standard regularization parameter choice methods that are based on a Gaussian assumption,

Fig. 5.4. Histograms of SNR for reconstructions for 100 noise realizations, using different windows and GCV-chosen regularization parameters.
and the experiment confirms this. In Table 5.2, with Gaussian noise, the largest improvement in SNR for our windowed method is 1.22dB for VE and 3.23dB for GCV; in contrast, for colored noise, the results in Table 5.5 show an improvement of 4.63dB for VE and 5.87dB for GCV.

6. Conclusions and discussion. In this paper, we developed a windowed approach for spectral regularization of ill-posed inverse problems. By partitioning the image in the frequency domain defined by the SVD and by allowing multiple regularization parameters for different windows of singular values, higher reconstruction accuracy can be obtained. Methods for selecting regularization parameters in a windowed framework, using VE and GCV, were developed and tested. A perturbation analysis demonstrates the better conditioning of the regularized problem in the windowed approach. Numerical results have demonstrated the value of the windowed regularization approach, particularly for high-noise data and colored noise data, since noise in different windows can be estimated separately.

Our work improves spectral methods by allowing better reconstruction of features,
and any spectral regularization approach can be extended to the windowed framework. Future work includes adding noise filtering schemes to the windowed images similar to that proposed in the ForWaRD algorithm of [10] and extensions to problems where the SVD cannot be easily computed. Also, there is great potential for the development of better windows, alternatives to Tikhonov filtering in each window (e.g., by a spline filter), and the use of other parameter choice methods.

Appendix. Derivation of GCV for the windowed case. In this section, we present a detailed derivation of the GCV function for the windowed case, providing a proof for Theorem 3.2. The model is given by \( \Sigma y \approx z \), and regularization parameters \( \lambda^{(j)} \), corresponding to windows \( W^{(j)} \), are found in vector \( \lambda \). Notice that in the SVD-based frequency domain, the windowed regularized solution can be written as

\[
y_{\text{win}} = \sum_{j=1}^{p} (\Sigma^T \Sigma + \lambda^{(j)} I)^{-1} W^{(j)} \Sigma^T z.
\]

The original derivation of the GCV function follows a “leave-one-out” approach, and it is known that difficulties arise when the system matrix is diagonal [3], which is the case here. Thus, let \( C \) be the unitary matrix that diagonalizes the circulants. Furthermore, assume that all \( n \) singular values are nonzero. Then, following standard practice [3], we consider the new system

\[
C \Sigma y \approx C z.
\]

Let’s define the new system matrix

\[
A = C \Sigma.
\]

Then the corresponding windowed regularized solution can be written as

\[
(A.1) \quad y_{\text{win}} = \sum_{j=1}^{p} (A^* A + \lambda^{(j)} I)^{-1} W^{(j)} A^* C z.
\]

Let \( y^{(k)}(\lambda) \) be the estimate of \( y_{\text{win}} \) with the \( k \)th data point missing. Then the Allen PRESS estimates [3] of \( \lambda^{(j)} \) minimize

\[
(A.2) \quad P(\lambda) = \frac{1}{m} \sum_{k=1}^{m} (Cz_k - [Ay^{(k)}(\lambda)]_k)^2,
\]

where \([Cz]_k\) and \([Ay^{(k)}(\lambda)]_k\) are the \( k \)th components of vectors \( Cz \) and \( Ay^{(k)}(\lambda) \), respectively.

It will be convenient to have the following notation. Let

\[
A(\lambda^{(j)}) = A(A^* A + \lambda^{(j)} I)^{-1} A^*;
\]

and let \( B(\lambda^{(j)}) \) be the diagonal matrix with entries \( 1 - \hat{a}_{kk}^{(j)} \), where \( \hat{a}_{kk}^{(j)} \) are the diagonal entries of matrix \( A(\lambda^{(j)}) \). Similarly, let

\[
\hat{A}(\lambda^{(j)}) = A(A^* A + \lambda^{(j)} I)^{-1} W^{(j)} A^*;
\]

and let \( \hat{B}(\lambda^{(j)}) \) be the diagonal matrix with entries \( 1 - \hat{a}_{kk}^{(j)} \), where \( \hat{a}_{kk}^{(j)} \) are the diagonal entries of matrix \( \hat{A}(\lambda^{(j)}) \). As the first step in deriving the windowed GCV function, we state and prove the following theorem for the windowed case.
Theorem A.1. For windowed spectral regularization, the Allen PRESS estimates of $\lambda^{(j)}$ minimize

\[ P(\lambda) = \frac{1}{m} \left\| \left( I - \sum_{j=1}^{p} B(\lambda^{(j)})^{-1} \hat{A}(\lambda^{(j)}) + \sum_{j=1}^{p} B(\lambda^{(j)})^{-1} (I - \hat{B}(\lambda^{(j)})) \right) Cz \right\|_2^2. \]

Proof. Define the $m \times m$ matrix

\[ E_k = \text{diag}(1, 1, \ldots, 1, 0, 1, \ldots, 1), \]

where $0$ is the $k$th entry, so that

\[ y^{(k)}(\lambda) = \sum_{j=1}^{p} \left( A^* E_k^T E_k A + \lambda^{(j)} I \right)^{-1} W^{(j)} A^* E_k^T E_k Cz. \]

From this definition, two important properties follow:

- $E_k = I - e_k e_k^T$;
- $E_k^T E_k = E_k$.

Using the first property,

\[ A^* E_k^T E_k A + \lambda^{(j)} I = (A^* A + \lambda^{(j)} I) - a_k a_k^*, \]

where $a_k^* = e_k^T A$. Then, using the Sherman–Morrison–Woodbury formula and the second property, $y^{(k)}(\lambda)$ in (A.4) can be written as

\[ y^{(k)}(\lambda) = \sum_{j=1}^{p} \left[ (A^* A + \lambda^{(j)} I)^{-1} + \frac{(A^* A + \lambda^{(j)} I)^{-1} a_k a_k^* (A^* A + \lambda^{(j)} I)^{-1}}{1 - a_k^* (A^* A + \lambda^{(j)} I)^{-1} a_k} \right] W^{(j)} A^* E_k Cz. \]

Since the $k$th component of $A y^{(k)}(\lambda)$ can be written as $e_k^T A y^{(k)}(\lambda)$, we see that

\[ [A y^{(k)}(\lambda)]_k = \sum_{j=1}^{p} \left( \frac{1}{1 - \hat{a}_{kk}^{(j)}} \right) e_k^T \hat{A}(\lambda^{(j)}) E_k Cz, \]

with some simple algebra. Then, using the first property above, the $k$th term in (A.2) can be written as

\[ [Cz]_k - [A y^{(k)}(\lambda)]_k = e_k^T Cz - \sum_{j=1}^{p} \left( \frac{1}{1 - \hat{a}_{kk}^{(j)}} \right) e_k^T \hat{A}(\lambda^{(j)}) E_k Cz \]

\[ = e_k^T Cz - \sum_{j=1}^{p} \left( \frac{1}{1 - \hat{a}_{kk}^{(j)}} \right) e_k^T \hat{A}(\lambda^{(j)}) Cz + \sum_{j=1}^{p} \frac{\hat{a}_{kk}^{(j)}}{1 - \hat{a}_{kk}^{(j)}} e_k^T Cz \]

\[ = e_k^T \left( I - \sum_{j=1}^{p} \left( \frac{1}{1 - \hat{a}_{kk}^{(j)}} \right) \hat{A}(\lambda^{(j)}) + \sum_{j=1}^{p} \frac{1 - (1 - \hat{a}_{kk}^{(j)})}{1 - \hat{a}_{kk}^{(j)}} I \right) Cz. \]

Thus, (A.2) is equivalent to (A.3).
Next, we proceed with the proof of Theorem 3.2, resulting in the GCV function for the general windowed case.

Proof. From Theorem A.1, the PRESS estimates for \( \lambda^{(j)} \) in the windowed case should minimize (A.3). A more computationally convenient form can be obtained by using the fact that \( A = C \Sigma \), so that

\[
\hat{A}(\lambda^{(j)}) = A(A^* A + \lambda^{(j)} I)^{-1} A^* = C \Sigma (\Sigma^T C^* \Sigma + \lambda^{(j)} I)^{-1} \Sigma^T C^* = C \Sigma (\Sigma^T \Sigma + \lambda^{(j)} I)^{-1} \Sigma^T C^*.
\]

Since \( D^{(j)} = \Sigma (\Sigma^T \Sigma + \lambda^{(j)} I)^{-1} \Sigma^T \) is diagonal, \( \hat{A}(\lambda^{(j)}) \) must be circulant. Furthermore, \( I - \hat{A}(\lambda^{(j)}) = I - C D^{(j)} = C(I - D^{(j)}) C^* \).

Now, consider \( B(\lambda^{(j)}) \), the diagonal matrix in (A.3) with entries equal to those of \( I - \hat{A}(\lambda^{(j)}) \). Determining these seems impossible until we use these two facts:

- The matrix \( I - \hat{A}(\lambda^{(j)}) \) is circulant, so all of the diagonal entries are equal, and thus \( B(\lambda^{(j)}) = \gamma^{(j)} I \).
- We know the eigenvalues of \( I - \hat{A}(\lambda^{(j)}) \), and therefore we know the trace of this matrix, which is the same as the trace of \( I - D^{(j)} \). Therefore, \( \gamma^{(j)} \) is the average value of the eigenvalues,

\[
\gamma^{(j)} = \frac{1}{m} \text{trace}(I - \hat{A}(\lambda^{(j)})) = \frac{1}{m} \left( (m - n) + \sum_{i=1}^{n} \frac{\lambda^{(j)}}{\sigma_i^2 + \lambda^{(j)}} \right).
\]

Similarly, if \( D^{(j)} = \Sigma (\Sigma^T \Sigma + \lambda^{(j)} I)^{-1} W^{(j)} \Sigma^T \), then \( \hat{A}(\lambda^{(j)}) = C D^{(j)} C^* \) must be circulant, and \( B(\lambda^{(j)}) = \hat{\gamma}^{(j)} I \), where

\[
\hat{\gamma}^{(j)} = \frac{1}{m} \text{trace}(I - \hat{A}(\lambda^{(j)})) = \frac{1}{m} \left( (m - n) + \sum_{i=1}^{n} \left( 1 - \frac{\sigma_i^2 \gamma^{(j)}}{\sigma_i^2 + \lambda^{(j)}} \right) \right).
\]

Thus, (A.3) can be further simplified,

\[
\text{GCV}(\lambda) = P(\lambda) = \frac{1}{m} \left\| \left( I - \sum_{j=1}^{p} \frac{1}{\gamma^{(j)}} \hat{A}(\lambda^{(j)}) + \sum_{j=1}^{p} \frac{1 - \hat{\gamma}^{(j)}}{\gamma^{(j)}} I \right) C z \right\|^2 = \frac{1}{m} \left\| C \left( I - \sum_{j=1}^{p} \frac{1}{\gamma^{(j)}} \hat{D}^{(j)} + \sum_{j=1}^{p} \frac{1 - \hat{\gamma}^{(j)}}{\gamma^{(j)}} I \right) C^* C z \right\|^2 = \frac{1}{m} \left\| \left( I - \sum_{j=1}^{p} \frac{1}{\gamma^{(j)}} \hat{D}^{(j)} + \sum_{j=1}^{p} \frac{1 - \hat{\gamma}^{(j)}}{\gamma^{(j)}} I \right) z \right\|^2 = \frac{1}{m} \left\| \sum_{i=1}^{n} \left( \left( \sum_{j=1}^{p} \frac{1 - \hat{\gamma}^{(j)}}{\gamma^{(j)}} \right) z_i - \sum_{j=1}^{p} \frac{1 - \hat{\gamma}^{(j)}}{\gamma^{(j)}} \frac{\sigma_i^2 \gamma^{(j)} z_i}{\sigma_i^2 + \lambda^{(j)}} \right)^2 \right. \\
+ \sum_{i=n+1}^{m} \left( 1 + \sum_{j=1}^{p} \frac{1 - \hat{\gamma}^{(j)}}{\gamma^{(j)}} \right)^2 z_i^2 \right\|.
\]
\[
\begin{align*}
&= \frac{1}{m} \left[ \sum_{i=1}^{n} \left( \sum_{j=1}^{p} \frac{1}{\gamma(j)} \left( \frac{\gamma(j)}{p} - \frac{\hat{\gamma}(j) + 1}{\sigma_i^2 + \lambda_j} \right) \right)^2 z_i^2 \\
&\quad + \sum_{i=n+1}^{m} \left( \sum_{j=1}^{p} \frac{1}{\gamma(j)} \left( \frac{\gamma(j)}{p} - \frac{\hat{\gamma}(j) + 1}{\sigma_i^2 + \lambda_j} \right) \right)^2 z_i^2 \right],
\end{align*}
\]
giving the desired GCV function.

Using this result, the GCV function can be easily evaluated for overlapping windows for any values of \(\lambda(j), j = 1, \ldots, p\). For Shannon windows, we use the simpler formulation given in (3.13).

Notice that for the standard Tikhonov formulation, which corresponds to the \(p = 1\) window, \(\gamma = \hat{\gamma}\) and \(w_i^{(j)} = 1, i = 1, \ldots, n\), so both (3.13) and (3.12) reduce to the standard PRESS function from [3] and (2.6).

Acknowledgment. We are grateful to the referees for helpful comments that improved the presentation.

REFERENCES
