On the Use of Arnoldi and Golub-Kahan Bases to Solve Nonsymmetric Ill-Posed Inverse Problems

Matthew A. Brown

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Masters of Science
in
Mathematics

Julianne Chung, Chair
Eric de Sturler
Serkan Gugercin

February 13, 2015
Blacksburg, Virginia

Keywords: Ill-Posed, Inverse Problems, Arnoldi, Krylov subspace

On the Use of Arnoldi and Golub-Kahan Bases to Solve Nonsymmetric Ill-Posed Inverse Problems

Matthew A. Brown

(ABSTRACT)

Iterative Krylov subspace methods have proven to be efficient tools for solving linear systems of equations. In the context of ill-posed inverse problems, they tend to exhibit semiconvergence behavior making it difficult detect "inverted noise" and stop iterations before solutions become contaminated. Regularization methods such as spectral filtering methods use the singular value decomposition (SVD) and are effective at filtering inverted noise from solutions, but are computationally prohibitive on large problems. Hybrid methods apply regularization techniques to the smaller "projected problem" that is inherent to iterative Krylov methods at each iteration, thereby overcoming the semiconvergence behavior.

Commonly, the Golub-Kahan bidiagonalization is used to construct a set of orthonormal basis vectors that span the Krylov subspaces from which solutions will be chosen, but seeking a solution in the orthonormal basis generated by the Arnoldi process (which is fundamental to the popular iterative method GMRES) has been of renewed interest recently. We discuss some of the positive and negative aspects of each process and use example problems to examine some qualities of the bases they produce. Computing optimal solutions in a given basis gives some insight into the performance of the corresponding iterative methods and how hybrid methods can contribute.

Acknowledgments

My deepest gratitude goes to Prof. Julianne Chung for her guidance, enthusiasm, vision, and patience as my advisor, and also to Profs. Eric de Sturler and Serkan Gugercin for agreeing to serve on my committee and their contributions of ideas, hard questions, and that most valuable of resources, their time.

I would also like to thank Silvia Gazzola for a stimulating discussion and her contribution of perspectives that improved the presentation of ideas in this thesis.

Finally, thanks beyond words go to my wife, Lora, for her encouragement and support in this pursuit.

Contents

List of Figures

Chapter 1

Introduction

1.1 Inverse Problems

This thesis focuses on two Krylov subspace methods for finding solutions of linear, ill-posed, inverse problems. Inverse problems themselves compose a very large class of problems which occur when there is a desire to undo a process that has resulted in a known or observed dataset to recover the original information. For example, if f is some function and d is a known output of the function, one might be interested to know what input, x, (if any) yields such an output, ie.

$$\text{find } x \text{ such that } f(x) = d.$$

Inverse problems are commonly discretized either due to the practicalities of the problem (only discrete data points are available) or in pursuit of numerical solutions when a closed form solution is unavailable or difficult to use. In the case of systems of linear equations, the forward problem of computing $\mathbf{b}$ from a given matrix $\mathbf{A}$ and vector $\mathbf{x}$ leads to the general inverse problem in consideration here:

$$\text{given } \mathbf{A} \text{ and } \mathbf{b}, \text{ find } \mathbf{x} \text{ such that } \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{1.1}$$

We take $\mathbf{A}$ to be a general matrix in $\mathbb{R}^{n\times n}$, $\mathbf{x}$ a vector in $\mathbb{R}^n$, and $\mathbf{b}$ a vector in $\mathbb{R}^n$. While many methods for solving inverse problems may be applied to rectangular matrices, methods based on the Arnoldi process, which are of particular interest here, may not. So for the purposes of comparison, we restrict our attention here to square matrices.

The matrix $\mathbf{A}$ is not assumed to be invertible, so other methods are needed to solve the problem than matrix inversion. But even for an invertible $n \times n$ matrix $\mathbf{A}$, the inverse

solution given by

$$\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{b}$$

may be unavailable if $\mathbf{A}$ is too large to be efficiently inverted, or the inverse solution may be otherwise undesirable as explained in the next section on ill-posed problems. This motivates methods of solution other than matrix inversion.

Many problems are naturally represented by systems where the matrix $\mathbf{A}$ is symmetric. This structure often allows for significant computational enhancements and accordingly a host of iterative methods are designed specifically for this scenario. Notable among these are MINRES which seeks solutions that have minimal residuals and the method of conjugate gradients (CG).

Among the methods that can be applied to nonsymmetric systems are those based on the Golub-Kahan iterative bidiagonalization[1] process which implicitly computes relations based on the normal equations,

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b},$$

and those such as the iterative method GMRES which are based on the Arnoldi process. The latter typically are generalizations of MINRES to the nonsymmetric case. Both Arnoldi and Golub-Kahan processes produce sequences of basis vectors spanning their respective Krylov subspaces. These bases are used to construct approximations for $\mathbf{x}$, and are the specific interest of this thesis.

Compared to Golub-Kahan bidiagonalization, one advantage of the Arnoldi process is that it does not require $\mathbf{A}^T$ to compute the basis. This is useful in situations where the transpose is unavailable or otherwise unfeasible to compute or store in computer memory. Additionally, application of the normal equations results in effectively squaring the condition number of the problem. This is a concern when considering problems which are already ill-conditioned.

In some cases, the discretization matrix $\mathbf{A}$ may be nonsingular in exact arithmetic, but singular in finite precision arithmetic, for example if any of its singular values, σ_i are smaller than the machine precision, ϵ. These are effectively zero in finite precision and thus the computer representation of the discretization is singular. This scenario is studied with some

[1] These methods are sometimes referred to as "Lanczos" methods in the literature as attribution to [34], but here we follow the example in [29] by referring to them as Golub-Kahan [17] methods

detail and interesting results for the iterative method GMRES in [5].

The remainder of this chapter will introduce further complexities associated with 'ill-posed' inverse problems. Chapter 2 provides background material on Krylov subspaces and Arnoldi and Golub-Kahan based methods. Regularization for ill-posed problems is also discussed as well as hybrid methods which combine regularization and Krylov subspace methods. In Chapter 3, we investigate characteristics of the bases produced by these methods which span either $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ or $\mathcal{K}_k(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b})$, and how these characteristics affect the construction of approximate solutions. Specific example problems are presented as illustrations of the performance of various methods and lead to ideas for modification or improvement.

1.2 Ill-Posed Inverse Problems

An inverse problem may be said to be "ill-posed" if any of the three following conditions hold:

1. the problem has no exact solution
2. the problem has multiple solutions
3. solutions do not depend continuously on the data.

In some cases, the underlying model problem may be well-posed (ie. none of the three conditions apply) while the discrete numerical problem derived from it is ill-posed [24]. This can happen as a result of discretization errors, measurement errors, or numerical errors including the effects of round-off due to the use of finite-precision computations. On the other hand, some problems are inherently ill-posed. Here we will focus on problems of the form

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{exact} + \mathbf{n} \tag{1.2}$$

where the matrix $\mathbf{A}$ is extremely ill-conditioned and $\mathbf{n}$ is additive noise.

What we call 'noise' could be attributed to the contamination of measurements from outside sources, or it could be the result of errors like those listed above. In this work, we assume Gaussian white noise. That is, the elements of $\mathbf{n}$ come from a Gaussian distribution with zero mean and a fixed standard deviation. We will refer to the relative noise level, η, which is defined by

$$\eta = \frac{\|\mathbf{n}\|}{\|\mathbf{b}_{exact}\|}. \tag{1.3}$$

Here and in the rest of this thesis, the norm $\|\cdot\|$ being considered is the Euclidean norm. A precise measurement of η is not typically available in practice. On the other hand, η can sometimes be well approximated from the information that is available. For more details on noise in inverse problems, see for example Hansen's book [24]. We use Matlab's `randn` function to produce noise that follows a Gaussian distribution which may then be scaled to simulate different noise levels.

For some problems, the addition of noise to the right hand side causes

$$\mathbf{b} = \mathbf{b}_{exact} + \mathbf{n} \tag{1.4}$$

to lie outside the range of $\mathbf{A}$. This can occur for example when $\mathbf{A}$ has a strong smoothing effect and η is relatively large. Since there is no $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$, the problem exhibits the first type of ill-posed behavior. In this case, we typically relax our search for an exact solution and instead seek a least squares solution that minimizes the residual norm $\|\mathbf{b} - \mathbf{Ax}\|$. It is known that the least squares solutions is given by

$$\mathbf{x}_{LS} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

which is the inverse solution to the normal equations for the problem. However, for problems where $\mathbf{A}$ is severely ill-conditioned, computing $\mathbf{x}_{LS}$ or $\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{b}$ where $\mathbf{A}$ is invertible can still produce poor solutions. Consider the inverse solution

$$\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1}\mathbf{b}_{exact} + \mathbf{A}^{-1}\mathbf{n} = \mathbf{x}_{exact} + \mathbf{A}^{-1}\mathbf{n}.$$

If

$$\|\mathbf{x}_{exact}\| \ll \|\mathbf{A}^{-1}\mathbf{n}\|$$

the inverse solution is dominated by the inverted noise. That is, the inverse problem exhibits the third type of ill-posed behavior in that it is extremely sensitive to small perturbations in the data. This is reflected in the error bound of the exact solution

$$\frac{\|\mathbf{x}_{exact} - \hat{\mathbf{x}}\|}{\|\mathbf{x}_{exact}\|} \leq cond(\mathbf{A})\frac{\|\mathbf{n}\|}{\|\mathbf{b}_{exact}\|} \tag{1.5}$$

where $cond(\mathbf{A})$ is the condition number of the invertible matrix $\mathbf{A}$ which is determined as

$$cond(\mathbf{A}) = \sigma_{\max}(\mathbf{A})/\sigma_{\min}(\mathbf{A}) \tag{1.6}$$

with $\sigma_{\max}(\mathbf{A})$ and $\sigma_{\min}(\mathbf{A})$ respectively the largest and smallest singular values of $\mathbf{A}$. The situation for the least squares solution can be worse since

$$cond(\mathbf{A}^T\mathbf{A}) \approx cond(\mathbf{A})^2$$

If $\mathbf{A}$ is not invertible, then the condition number is unbounded and errors in approximate solutions may be arbitrarily large.

For ill-posed inverse problems, the least squares solution and the inverse solutions are contaminated by inverted noise. Regularization is a means to produce accurate reconstructions solving a nearby well-conditioned problem. Typically regularization is accomplished by imposing prior knowledge or constraints on the solution. For example, common regularization techniques seek solutions which are smoothest, have smallest total variation [1], have sparsity in some basis [10], or have specific upper and lower bounds.

Ill-posed inverse problem occur naturally in a variety of scientific and technical systems and have been well-studied in these contexts. General introductions to the area of ill-posed inverse problems can be found in [3], [19], [31], [38], and [46].

Image deblurring is a classic example of an ill-posed problem and the techniques we will explore here have been applied successfully to image deblurring problems in [4], [8], [15], [14], [16], and [27] among others. In this thesis, we consider smaller 1D examples of ill-posed inverse problems from Hansen's toolbox [23]. In the next section, we describe one such example that will be used throughout for illustration purposes.

1.3 Example

1-D Inverse Heat Problem

In the inverse heat problem, the task is to take an observed temperature reading and determine an initial temperature state that would have resulted in the observations after some period of time. The inverse heat problem is modeled using the Volterra integral equation of the first kind, which have the form

$$\int_0^s k(s-t)f(t)dt = g(s), \quad 0 \leq s \leq 1.$$

In the heat problem, the convolution kernel k is given by

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

The discretization of this problem by means of the midpoint rule leads to the linear system

$$\mathbf{b}_{exact} = \mathbf{A}\mathbf{x}_{exact} \tag{1.7}$$

where $\mathbf{A}$ is a nonsymmetric, Toeplitz, lower triangular matrix representing the convolution operation, $\mathbf{x}$ is an initial temperature state of the system, and $\mathbf{b}$ is the temperature state after some period of time.

The "forward" problem is to compute $\mathbf{b}$ given a known initial state $\mathbf{x}$ and is well posed: for any $\mathbf{x}$ there is a unique $\mathbf{b}$ satisfying (1.7), and solutions depend continuously on the data. However, the inverse heat problem (find $\mathbf{x}$ satisfying (1.7) for an observed final state $\mathbf{b}$) is ill-posed when $\kappa < 4$.

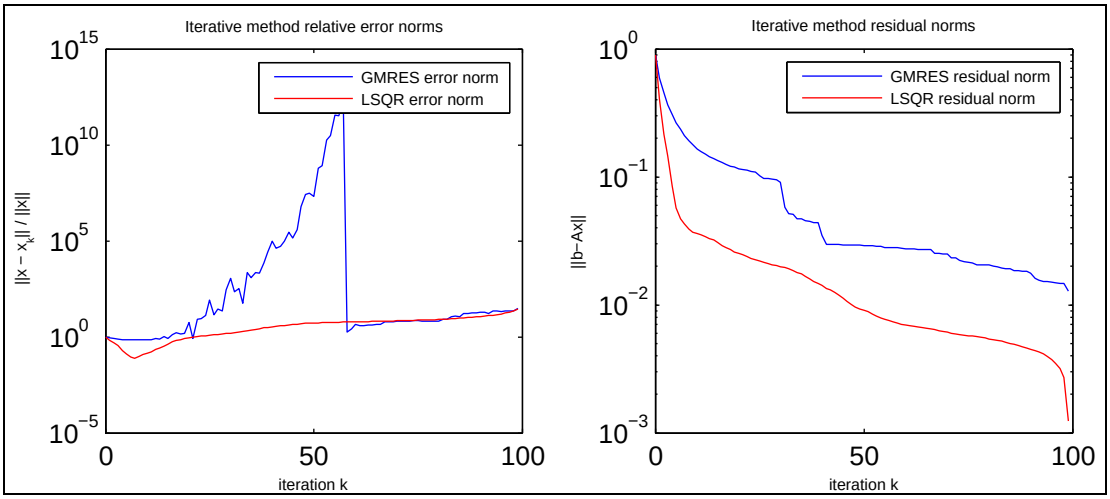
The source of the ill-posedness for this problem comes from the underlying physics. Since an imbalance in heat distribution will dissipate over time and tend towards an equilibrium, many initial heat distributions may result in very similar observed heat distributions especially after long periods of time. This indicates that a solution to the inverse heat problem may not be unique or that a small perturbation in the observation may correspond with a large change in the solution. This problem has been explored by many authors which use it as an introduction and motivation to ill-posed inverse problems [4].

Hansen has implemented a discretization of this system in the `heat` test problem provided in the Regularization Tools toolkit [23]. In this implementation, the parameter κ found in the kernel $k(\tau)$ can by any positive real number and affects the conditioning of the system. For example, $\kappa = 1$ is very ill-conditioned and $\kappa = 5$ is fairly well-conditioned. In the example here, the value $\kappa = 2$ is used. This produces a square, lower triangular matrix $\mathbf{A}$ whose diagonal elements a_{ii} are identical to each other and tend exponentially to zero with larger n. This leads to extreme ill-conditioning of $\mathbf{A}$ even for relatively small values of n. We use $n = 100$ here.

To simulate noisy data in the right hand side of the system, we compute

$$\mathbf{b} = \mathbf{b}_{exact} + \mathbf{n}$$

Figure 1.1: Iterative methods applied to ill-posed inverse problems often exhibit semiconvergence behavior. Here iterations of GMRES and LSQR are applied to the `heat` problem with a mirror image of the default $\mathbf{x}_{exact}$ and noise level $\eta = .05$. Note that residual norms decrease even while relative error norms increase.



where n is a Gaussian white noise vector $\mathbf{n}$ with relative noise level η. In the example that follows, a noise level of $\eta = 0.05$ is used.

To illustrate difficulties typical of solving ill-posed inverse problems, we apply the iterative methods GMRES and LSQR to this inverse heat problem with added noise. These two widely used methods are described in Chapter 2 and each produces in the kth step an approximate solution $\mathbf{x}_k$. Figure (1.1) shows plots of the norm of the relative error of each iterate, computed as

$$relative\ error = \frac{\|\mathbf{x}_{exact} - \mathbf{x}_k\|}{\|\mathbf{x}_{exact}\|}.$$

and also the residual norms

$$residual\ norm = \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|$$

versus the iteration index, k, for each of the methods. The exact solution $\mathbf{x}_{exact}$ is not available in practice, but is useful in test problems to analyze the behavior of solution techniques and in methods which use known solutions to develop a set of training data. Note the base-10 log scale on the vertical axis.

The behavior of the error norm and residual norm plots shown in Figure (1.1) is common when standard iterative methods are applied to ill-posed inverse problems. In the error

plot, we see what is called the semiconvergence of both methods. In the initial iterations the methods show improving approximations $\mathbf{x}_k$, but then they begin to worsen drastically. This is because we are initially constructing meaningful components of the true solution, but later are converging towards $\mathbf{A}^{-1}\mathbf{b}$ which is dominated by inverted noise.

The residual norm is a readily available tool that is often used to judge convergence of an iterative method towards the solution. GMRES and LSQR are designed to produce iterates $\mathbf{x}_k$ which minimize the residual. But as this example shows, solutions with smaller residuals are not necessarily accurate solutions for ill-posed inverse problems.

Also note that the relative error plots for GMRES and LSQR are very different from each other. These two methods share much of the same ideology in their derivation, but can have strikingly different performance characteristics depending on the problem. These differences are of particular interest in this thesis.

Chapter 2

Background

When finding approximate solutions to large scale discrete ill-posed inverse problems, we want to minimize computational effort while suppressing the influence of inverted noise. Iterative methods such as those based on Krylov subspace methods are often good at the former while regularization techniques such as spectral filtering provide the latter.

In this chapter we give background information on a class of iterative methods called Krylov methods. Then we discuss SVD based regularization methods including TSVD and Tikhonov regularization. Lastly, we address hybrid methods which seek to combine iterative methods with direct regularization in an effort to balance the two goals.

2.1 Iterative Krylov Methods

For a matrix $\mathbf{A}$, vector $\mathbf{b}$, and natural number k, the kth Krylov subspace, written $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ or when the context is clear, $\mathcal{K}_k$, is defined as

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = span\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\}.$$

Part of the appeal of Krylov subspaces is that, by their formation, they are adapted to the problem at hand [24]. Many iterative methods such as MINRES, GMRES, conjugate gradients (CG), and methods derived from these are categorized as Krylov subspace methods since their iterates are chosen from such spaces. In particular, for minimium residual methods, the goal is to compute a sequence of approximations to the linear least squares problem

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \tag{2.1}$$

by computing the best approximation in the Krylov subspace. That is, at each iteration k, we seek

$$\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A}, \mathbf{b}) \text{ such that } \mathbf{x}_k = \arg\min \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \tag{2.2}$$

2.1.1 Arnoldi Methods

Arnoldi methods are iterative methods that may be applied to the general inverse problem (1.1) where $\mathbf{A}$ is a general matrix in $\mathbb{R}^{n\times n}$. Arnoldi methods are typically used when $\mathbf{A}$ is nonsymmetric since symmetry can be exploited for a substantial reduction in computational effort. This class of methods uses the Arnoldi iteration (one step of the algorithm is given in Figure 2.1) to construct a set of orthonormal vectors that form a basis for $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$. The Arnoldi iteration [2] produces these vectors iteratively in a two-phase process of extending the basis by constructing a new vector and then orthonormalizing it against all the previous vectors. The initial basis vector $\mathbf{w}_1 \in \mathbb{R}^n$ is taken to be the normalized residual of the initial guess $\mathbf{x}_0 \in \mathbb{R}^n$,

$$\mathbf{w}_1 = \frac{\mathbf{b} - \mathbf{A}\mathbf{x}_0}{\|\mathbf{b} - \mathbf{A}\mathbf{x}_0\|}.$$

In the absence of prior information about the problem, it is usually advised to take $\mathbf{x}_0 = \mathbf{0}$ as the initial guess. The Arnoldi algorithm can be extended to the case where $\mathbf{x}_0$ is not the zero vector [45], but for simplicity we assume $\mathbf{x}_0 = \mathbf{0}$ from this point on. As a result

$$\mathbf{w}_1 = \frac{\mathbf{b}}{\|\mathbf{b}\|}. \tag{2.3}$$

At the first iteration, the Krylov subspace is spanned by only the single vector $\mathbf{w}_1$, which is by definition orthonormal. At each subsequent iteration, the set of vectors is extended by multiplying the previous orthonormalized vector with $\mathbf{A}$ to get a new vector $\widetilde{\mathbf{w}}_k = \mathbf{A}\mathbf{w}_{k-1}$. Then, this new vector $\widetilde{\mathbf{w}}_k$ is orthogonalized against each of the previous vectors one at a time.

This means that in the kth step, the new vector will be orthogonalized against $k-1$ vectors. This growth of work in each step is often cited as a disadvantage of the Arnoldi iteration since Golub-Kahan methods (details follow in the next section) exploit symmetry to require a smaller and constant amount of work for their orthogonalization in each step which makes them faster in situations where a moderately large number of iterations are performed. However, the extra work of the full Arnoldi orthogonalization can be beneficial since basis vectors tend to lose orthogonality in finite precision arithmetic.

Figure 2.1: One Step of the Arnoldi Process

Input : $\mathbf{A}, \mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_k}$
Result: $\mathbf{w}_{k+1}$
$\mathbf{w}_{k+1} = \mathbf{A}\mathbf{w}_k$;
for *i from* 1 *to* k **do**
  $h_{i,k} = \mathbf{w}_i^T \mathbf{w}_{k+1}$;
  $\mathbf{w}_{k+1} = \mathbf{w}_{k+1} - h_{i,k}\mathbf{w}_i$;
end
$h_{k+1,k} = \|\mathbf{w}_{k+1}\|_2$;
$\mathbf{w}_{k+1} = \mathbf{w}_{k+1}/h_{k+1,k}$;

Orthogonalization within the Arnoldi method may be accomplished in various ways, but often the modified Gram Schmidt (MGS) process is cited as the preference. The main concerns in choosing an orthogonalization method are computational efficiency and numerical stability. It is worth noting that the built-in Matlab implementation of GMRES uses Householder reflections for the orthogonalizations in its Arnoldi iterations as suggested in [48]. Our implementations of Arnoldi methods use MGS orthogonalization.

After the orthogonalizations are completed, the resulting vector is normalized for numerical stability of subsequent operations. After $k+1$ iterations of the Arnoldi process, we get the following relation:

$$\mathbf{A}\mathbf{W}_k = \mathbf{W}_{k+1}\mathbf{H}_{k+1,k} \tag{2.4}$$

where $\mathbf{W}_k = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_k] \in \mathbb{R}^{n\times k}$ and $\mathbf{H} \in \mathbb{R}^{(k+1)\times k}$ is an upper Hessenberg matrix with elements h_{ij} corresponding to the inner products computed during the orthonormalization process.

GMRES is an minimum residual iterative method which uses the Arnoldi process to generate the matrices $\mathbf{W}_k$ and $\mathbf{H}_{k+1,k}$ in the kth iteration to project (2.1) onto the Krylov subspace $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ to compute solutions $\mathbf{x}_k$ which satisfy (2.2).

Since the vectors $\{\mathbf{w}_i\}_{i=1}^k$ form an orthonormal basis for $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$, an approximate solution $\mathbf{x}_k$ that satisfies (2.2) will have the form

$$\mathbf{x}_k = \mathbf{W}_k\mathbf{y}_k \tag{2.5}$$

where $\mathbf{y}_k \in \mathbb{R}^k$. Further, from (2.3), we have

$$\mathbf{b} = \beta \mathbf{W}_{k+1} \mathbf{e}_1 \tag{2.6}$$

where $\beta = \|\mathbf{b}\|$ and $\mathbf{e}_1$ is the first column of the identity matrix of size $k+1$. Then using (2.6) and (2.5), then (2.4), we conclude that the residual vector corresponding to the solution at the kth iteration can be written as

$$\begin{aligned}
\mathbf{r}_k &= \mathbf{b} - \mathbf{A}\mathbf{x}_k \\
&= \beta \mathbf{W}_{k+1} \mathbf{e}_1 - \mathbf{A}\mathbf{W}_k \mathbf{y}_k \\
&= \beta \mathbf{W}_{k+1} \mathbf{e}_1 - \mathbf{W}_{k+1} \mathbf{H}_{k+1,k} \mathbf{y}_k \\
&= \mathbf{W}_{k+1} (\beta \mathbf{e}_1 - \mathbf{H}_{k+1,k} \mathbf{y}_k)
\end{aligned}$$

Since $\mathbf{W}_{k+1}$ has orthonormal columns, it follows that for $\mathbf{x}_k \in \mathcal{R}(\mathbf{W}_k)$

$$\|\mathbf{r}_k\| = \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\| = \|\beta \mathbf{e}_1 - \mathbf{H}_{k+1,k} \mathbf{y}_k\| \tag{2.7}$$

so we can approximate the least squares problem (2.1) with the projected least squares problem

$$\min_{\mathbf{x} \in R(\mathbf{W}_k)} \|\mathbf{r}_k\| = \min_{\mathbf{y} \in \mathbb{R}^k} \|\beta \mathbf{e}_1 - \mathbf{H}_{k+1,k} \mathbf{y}\| \tag{2.8}$$

and choose the approximate solution (2.5), where $\mathbf{y}_k$ is the solution to (2.8).

From a computational perspective, the projected least squares problem (2.8) is significantly easier and faster to solve due its smaller size and the favorable structure of $\mathbf{H}_{k+1,k}$.

In some extreme cases, cancellation errors may occur in finite precision arithmetic during the orthogonalization steps that result in a loss of orthogonality among the computed basis vectors, $\{\mathbf{w}_i\}_{i=1}^k$. This can be problematic if the stability or accuracy of the method depends on the orthogonality of the basis vectors.

If loss of orthogonality is expected or observed in an Arnoldi based method for a particular problem, then reorthogonalization of the basis vectors can help but effectively doubles the computational work in each step. An algorithm is given for complete reorthogonalization in Figure (2.2).

Figure 2.2: One Step of the Arnoldi Process with Complete Reorthogonalization

Input : $\mathbf{A}, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k$
Result: $\mathbf{w}_{k+1}, h_{1,k}, \cdots, h_{k+1,k}$
$\mathbf{w}_{k+1} = \mathbf{A}\mathbf{w}_k$;
for *i from* 1 *to* k **do**
  $h_{i,k} = \mathbf{w}_i^T \mathbf{w}_{k+1}$;
  $\mathbf{w}_{k+1} = \mathbf{w}_{k+1} - h_{i,k}\mathbf{w}_i$
end
for *i from* 1 *to* k **do**
  $\rho = \mathbf{w}_i^T \mathbf{w}_{k+1}$;
  $\mathbf{w}_{k+1} = \mathbf{w}_{k+1} - \rho \mathbf{w}_i$;
  $h_{i,k} = h_{i,k} + \rho$;
end
$h_{k+1,k} = \|\mathbf{w}_{k+1}\|$;
$\mathbf{w}_{k+1} = \mathbf{w}_{k+1}/h_{k+1,k}$

2.1.2 Golub-Kahan Iterative Bidiagonalization

Contrary to Arnoldi methods, Golub-Kahan based methods seek minimum norm residual solutions in the Krylov subspace defined by

$$\mathcal{K}_k(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b}).$$

so that at each iteration we seek a solution to

$$\min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b})} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2, k = 1, 2, ... \tag{2.9}$$

The Golub-Kahan bidiagonalization process is closely related to the Lanczos method [34] applied to real symmetric matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ or alternatively to $\begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{pmatrix}$, and follows the conceptual ideas of the Arnoldi process described above. However, the Golub-Kahan process takes advantage of the symmetry of $\mathbf{A}^T\mathbf{A}$ to reduce the work in each iteration and the storage requirements. This reduction is accomplished through a shorter orthogonalization step which only depends on two vectors in each iteration instead of k vectors as in Arnoldi.

Define $\mathbf{u}_1$ by $\beta\mathbf{u}_1 = \mathbf{b}$, i.e. $\mathbf{u}_1 = \frac{\mathbf{b}}{\|\mathbf{b}\|}$ and define $\mathbf{v}_1$ by $\alpha_1\mathbf{v}_1 = \mathbf{A}^T\mathbf{u}_1$ where $\alpha_1 = \|\mathbf{A}^T\mathbf{u}_1\|$. Then for $i = 1, 2, ...$

Figure 2.3: One Step of Golub-Kahan Iterative Bidiagonalization

Data: $\mathbf{A}, \mathbf{v}_{k-1}, \mathbf{u}_k, \alpha_{k-1}, \beta_k$
Result: $\mathbf{v}_k, \mathbf{u}_{k+1}, \alpha_k, \beta_{k+1}$
$\mathbf{v} = \mathbf{A}^T\mathbf{u}_k - \beta_k\mathbf{v}_{k-1}$;
$\alpha_k = \|\mathbf{v}\|$;
$\mathbf{v}_k = \mathbf{v}/\alpha_k$;
$\mathbf{u} = \mathbf{A}\mathbf{v}_k - \alpha_k\mathbf{u}_k$;
$\beta_{k+1} = \|\mathbf{u}\|$;
$\mathbf{u}_{k+1} = \mathbf{u}/\beta_{k+1}$;

$$\beta_{i+1}\mathbf{u}_{i+1} = \mathbf{A}\mathbf{v}_i - \alpha_i\mathbf{u}_i$$
$$\alpha_{i+1}\mathbf{v}_{i+1} = \mathbf{A}^T\mathbf{u}_{i+1} - \beta_{i+1}\mathbf{v}_i$$

where β_i and α_i are constants such that $\mathbf{u}_i$ and $\mathbf{v}_i$ have unit norm for $i = 1, 2, \ldots, n$. df An algorithm for computing these vectors and coefficients is given in Figure (2.3).

After k iterations, the matrices $\mathbf{U}_k$ and $\mathbf{V}_k$ in $\mathbb{R}^{n\times k}$ defined by

$$\mathbf{U}_k = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_k] \quad \text{and} \quad \mathbf{V}_k = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_k]$$

have orthonormal columns in exact arithmetic and together with the lower bidiagonal matrix

$$\mathbf{B}_{k+1,k} = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{bmatrix}$$

satisfy the following relations for $k = 1, 2, \ldots, ..., n$

$$\mathbf{A}^T\mathbf{U}_{k+1} = \mathbf{V}_k\mathbf{B}_{k+1,k}^T + \alpha_{k+1}\mathbf{v}_{k+1}e_{k+1}^T \tag{2.10}$$
$$\mathbf{A}\mathbf{V}_k = \mathbf{U}_{k+1}\mathbf{B}_{k+1,k}. \tag{2.11}$$

We can use these relations to iteratively find solutions to the projected problem (2.9). Since the vectors $\{\mathbf{v}_i\}_{i=1}^k$ form an orthonormal basis for $\mathcal{K}_k(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b})$ by construction, an approximate solution $\mathbf{x}_k$ that is in this Krylov subspace will have the form

$$\mathbf{x}_k = \mathbf{V}_k \mathbf{f}_k. \tag{2.12}$$

where $\mathbf{f}_k \in \mathbb{R}^k$. Then using (2.11) and the fact that $\mathbf{b} = \beta_1 \mathbf{U}_{k+1}\mathbf{e}_1$ we can write the residual corresponding to the solution at the kth iteration as

$$\begin{aligned}
\mathbf{r}_k &= \mathbf{b} - \mathbf{A}\mathbf{x}_k \\
&= \beta \mathbf{U}_{k+1}\mathbf{e}_1 - \mathbf{A}\mathbf{V}_k\mathbf{f}_k \\
&= \beta \mathbf{U}_{k+1}\mathbf{e}_1 - \mathbf{U}_{k+1}\mathbf{B}_{k+1,k}\mathbf{f}_k \\
&= \mathbf{U}_{k+1}\left(\beta \mathbf{e}_1 - \mathbf{B}_{k+1,k}\mathbf{f}_k\right)
\end{aligned}$$

Then since $\mathbf{U}_{k+1}$ has orthonormal columns

$$\|\mathbf{r}_k\| = \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\| = \|\beta \mathbf{e}_1 - \mathbf{B}_{k+1,k}\mathbf{f}_k\|,$$

and we can approximate the least squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$$

with the projected least squares problem

$$\min_{\mathbf{x}\in R(\mathbf{V}_k)} \|\mathbf{r}_k\| = \min_{\mathbf{f}\in\mathbb{R}^k} \|\beta \mathbf{e}_1 - \mathbf{B}_{k+1,k}\mathbf{f}\| \tag{2.13}$$

and choose the approximate solution as (2.12) where $\mathbf{f}_k$ solves (2.13).

As with the Arnoldi process, reorthogonalization has been proposed for the Golub-Kahan bidiagonalization process to mitigate loss of orthogonality among the computed basis vectors for the Krylov subspace which is caused by cancellation errors. An algorithm incorporating complete reorthogonalization is given in Figure (2.4).

When considering the use of this algorithm, it is worth noting first that the reorthogonalization is computationally expensive and grows with each iteration as in the Arnoldi iteration since the new vector is orthogonalized against all the previous vectors. Additionally, implementing reorthogonalization in this way for the Golub-Kahan process may result in the failure of relations (2.10) and (2.11) to hold since the matrix $\mathbf{B}_{k+1,k}$ is not updated to reflect changes made to the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$.

Figure 2.4: One Step of Golub-Kahan Iterative Bidiagonalization with Complete Reorthogonalization

Input : $\mathbf{A}, \mathbf{v}_1, \cdots, \mathbf{v}_{k-1}, \mathbf{u}_1, \cdots, \mathbf{u}_k, \alpha_{k-1}, \beta_k$
Result: $\mathbf{v}_k, \mathbf{u}_{k+1}, \alpha_k, \beta_{k+1}$
$\mathbf{v} = \mathbf{A}^T\mathbf{u}_k - \beta_k\mathbf{v}_{k-1}$;
for *j from* 1 *to* $k-1$ **do**
  $\mathbf{v}_k = \mathbf{v}_{k-1} - (\mathbf{v}_j^T\mathbf{v}_k)\mathbf{v}_j$; ;
end
$\alpha_k = \|\mathbf{v}\|$;
$\mathbf{v}_k = \mathbf{v}/\alpha_k$;
$\mathbf{u} = \mathbf{A}\mathbf{v}_k - \alpha_k\mathbf{u}_k$;
for *j from* 1 *to* k **do**
  $\mathbf{u} = \mathbf{u} - (\mathbf{u}_j^T\mathbf{u})\mathbf{u}_j$; ;
end
$\beta_{k+1} = \|\mathbf{u}\|$;
$\mathbf{u}_{k+1} = \mathbf{u}/\beta_{k+1}$;

The Golub-Kahan process is an attractive option because the computed matrix $\mathbf{B}_{k+1,k}$ has a sparse, bidiagonal structure making the minimization problem (2.13) easy to solve very quickly with small storage demand. Updating $\mathbf{B}_{k+1,k}$ to reflect the reorthogonalization of the columns of $\mathbf{V}_k$ and $\mathbf{U}_k$ would be easy to implement (as done for reorthogonalization of Arnoldi) but would result in $\mathbf{B}_{k+1,k}$ losing its bidiagonal structure and becoming Hessenberg instead [35]. This greatly increases the amount of computational effort and storage needed to solve the projected minimization problem.

It is known that the Arnoldi basis is better for some problems, while the Golub-Kahan basis is better for others. Previous work on analyzing the comparative convergence properties of LSQR and GMRES can be found in [5], [39], [30].

2.2 Regularization for Ill-Posed Inverse Problems

Minimum residual methods based on the Arnoldi method and Golub-Kahan method have been applied successfully to numerous problems. But as we mentioned in the introduction, semiconvergence often makes them difficult to use for solving ill-posed inverse problems, particularly when the observed data are contaminated by additive noise.

In this section, we step out of the context of Krylov subspace methods and explore regularization techniques that can be used to handle noise contamination in a general setting. We return to iterative methods in the context of regularization in the following section and

discuss hybrid methods that combine both techniques.

Here our interest is still in solving problems which have the form (1.2) and are contaminated by additive Gaussian white noise $\mathbf{n}$ with relative noise level $\eta = \frac{\|\mathbf{n}\|}{\|\mathbf{b}_{exact}\|}$. We wish to find an approximate solution as near as possible to the unknown true solution $\mathbf{x}_{exact}$. Due to the added noise and ill-conditioning of $\mathbf{A}$, we generally do not want solutions $\mathbf{A}^{-1}\mathbf{b}$ or $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ and prefer regularized solutions which have been constructed to preserve useful components of the solution while suppressing the influence of the inverted noise.

Here we discuss two regularization methods which can be expressed in terms of the singular value decomposition (SVD) of the matrix $\mathbf{A}$. The SVD of $\mathbf{A} \in \mathbb{R}^{n\times n}$ is given as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{2.14}$$

where $\mathbf{U}$ and $\mathbf{V}^T$ are orthogonal matrices in $\mathbb{R}^{n\times n}$ and $\mathbf{\Sigma} = diag(\sigma_1, \sigma_2, \cdots, \sigma_n)$ is a diagonal matrix with the n singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ of $\mathbf{A}$ on the diagonal. The columns, $\mathbf{u}_i$ of $\mathbf{U}$ and $\mathbf{v}_i$ of $\mathbf{V}$ are called respectively the left and right singular vectors of $\mathbf{A}$. Since $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, we have the useful property that

$$\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I} \quad \text{and} \quad \mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}.$$

In the case that $\mathbf{A}$ is non-singular, this leads to

$$\mathbf{A}^{-1} = \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$$

where $\mathbf{\Sigma}^{-1}$ is a diagonal matrix with diagonal elements σ_i^{-1}.

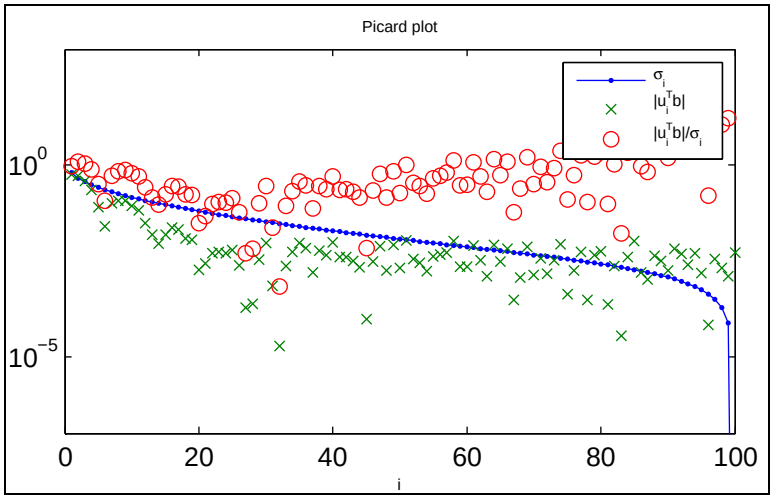
Using the SVD, the inverse solution can be written as

$$\hat{\mathbf{x}} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i. \tag{2.15}$$

Even if $\mathbf{A}$ is not full rank, the Moore-Penrose pseudoinverse gives a similar representation in terms of the SVD.

It is typical of ill-posed inverse problems that the singular values of $\mathbf{A}$ decay to and cluster near zero with no discernable gap between consecutive values. A discretized problem is said to satisfy the Discrete Picard Condition (DPC) if the SVD coefficients $|\mathbf{u}_i^T\mathbf{b}|$ on average

Figure 2.5: Picard Plot of `heat` test problem. $|\mathbf{u}_i^T\mathbf{b}|$ coefficients level off around $\eta = 0.05$ while singular values decay.



decay faster than the corresponding singular values σ_i [21].

The Picard plot is often used to illustrate these concepts. In Figure (2.5), we provide a Picard plot for the heat example from Section 1.3. The Picard plot shows the singular values σ_i, the SVD coefficients $|\mathbf{u}_i^T\mathbf{b}|$, and the solution coefficients $\frac{|\mathbf{u}_i^T\mathbf{b}|}{\sigma_i}$.

This phenomenon is usually visible from the Picard plot. The SVD coefficients will be decreasing with the index i until they reach the relative noise level η at some k, at which point they will level off.

This behavior can be explained by the presence of additive noise in the right hand side. If we used $\mathbf{b}_{exact}$ in the Picard plot instead, then we would observe that the coefficients $|\mathbf{u}_i^T\mathbf{b}|$ decay to zero faster that the singular values and thus the solution coefficients would tend towards zero in absolute value instead of increasing after some index k. This is because for all $i = 1, \ldots, n$ we have the expansion

$$\mathbf{u}_i^T\mathbf{b} = \mathbf{u}_i^T\left(\mathbf{b}_{exact} + \mathbf{n}\right) = \mathbf{u}_i^T\mathbf{b}_{exact} + \mathbf{u}_i^T\mathbf{n}$$

and for large i

$$\mathbf{u}_i^T\mathbf{b}_{exact} \approx 0 \quad \text{and} \quad \mathbf{u}_i^T\mathbf{n} \approx \eta.$$

Then, since the singular values continue to decay to zero, the solution coefficients for large i

$$\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i} \approx \frac{\eta}{\sigma_i}$$

become very large. This results in an amplification of the contribution of singular vectors $\mathbf{v}_i$ that correspond to small singular values and are highly oscillatory. This effect is often called "inverted noise".

Hansen [24] gives an informative development of SVD based regularization beginning with the SVE (continuous singular value expansion for which the SVD is the discrete analog).

Here we consider spectral filtering methods, where regularized solutions can be written as

$$\mathbf{x}_{reg} = \sum_{i=1}^{n} \phi_i \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i. \tag{2.16}$$

where ϕ_i are filter factors. The filter factors ϕ_i are so named because they usually take on values in $[0, 1]$ with the purpose of filtering any components of the solution that are dominated by inverted noise while allowing to pass any components which contain useful information about the solution. Choosing a good set of filter factors determines the success of the regularization method. Next we discuss two filtering methods: TSVD and Tikhonov.

2.2.1 TSVD

The first regularization method we will look at is the Truncated SVD regularization method or simply TSVD. As the name suggests, it attempts to prevent inverted noise from entering an approximate solution by filtering all components of the solution after an index k, which is intended to be the index at which the SVD components have reached the relative noise level η of the right hand side $\mathbf{b}$. That is, we take the filter factors $\phi_i = 1$ for $i \leq k$ and $\phi_i = 0$ for $i > k$ and construct the regularized solution as

$$\mathbf{x}_{TSVD} = \sum_{i=1}^{n} \phi_i \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i \tag{2.17}$$

Another interpretation of TSVD is obtained by constructing the reduced matrix

$$\mathbf{A}_k = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ | & & | \end{bmatrix}^T$$

and consider solving

$$\min \|\mathbf{b} - \mathbf{A}_k\mathbf{x}\|$$

which is better conditioned since typically

$$cond(\mathbf{A}_k) = \sigma_1/\sigma_k \ll \sigma_1/\sigma_n = cond(\mathbf{A})$$

but requires additional constraints because $\mathbf{A}_k$ is rank deficient. It can be shown that (2.17) is the exact solution to the problem

$$\min \|\mathbf{x}\| \quad \text{subject to} \quad \|\mathbf{b} - \mathbf{A}_k\mathbf{x}\| = \min$$

for fixed k. The truncation parameter k should be chosen so that $|\mathbf{u}_k^T\mathbf{b}| \approx \eta$ for $i > k$. There are many parameter selection techniques which may be paired with TSVD. We will address some of these in the next section.

2.2.2 Tikhonov Regularization

In Tikhonov regularization [41], [44], solutions have the form (2.16) where the filter factors ϕ_i take on real values between zero and one, not just the integer values 0 and 1 as in TSVD. This allows a weighted filtering which may be useful since for $i \approx k$ where k is the index where inverted noise begins to dominate solution coefficients, vectors $\mathbf{v}_i$ may still be able to contribute meaningfully to the solution if the noise is appropriately damped in the solution coefficients.

The standard-form Tikhonov regularization problem can be formulated as the linear least squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\| \begin{pmatrix} \mathbf{A} \\ \alpha\mathbf{I} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\| \tag{2.18}$$

which is often written as

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 + \alpha^2 \|\mathbf{x}\|^2 . \tag{2.19}$$

where $\alpha \in \mathbb{R}$ is a regularization parameter. By solving the corresponding normal equations, the Tikhonov solution can be written as

$$\mathbf{x}_\alpha = \left(\mathbf{A}^T\mathbf{A} + \alpha^2\mathbf{I}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{2.20}$$

Many researchers have considered using iterative methods to compute the Tikhonov solution. Here we use Tikhonov regularization as a spectral filtering approach. Using the SVD, equation (2.20) can be written as

$$\begin{aligned}
\mathbf{x}_\alpha &= \left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T + \alpha^2\mathbf{V}\mathbf{V}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{b} \\
&= \left(\mathbf{V}\left(\mathbf{\Sigma}^2 + \alpha^2\mathbf{I}\right)\mathbf{V}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{b} \\
&= \mathbf{V}\left(\mathbf{\Sigma}^2 + \alpha^2\mathbf{I}\right)^{-1}\mathbf{\Sigma}\mathbf{U}^T\mathbf{b} \\
&= \mathbf{V}\left[\left(\mathbf{\Sigma}^2 + \alpha^2\mathbf{I}\right)^{-1}\mathbf{\Sigma}^2\right]\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b} \\
&= \mathbf{V}\mathbf{\Phi}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}
\end{aligned}$$

where $\mathbf{\Phi} = \left(\mathbf{\Sigma}^2 + \alpha^2\mathbf{I}\right)^{-1}\mathbf{\Sigma}^2$ is a diagonal matrix whose diagonal elements are given by

$$\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \approx \begin{cases} 1, & \sigma_i \gg \alpha \\ \dfrac{\sigma_i^2}{\alpha^2}, & \sigma_i \ll \alpha \end{cases}. \tag{2.21}$$

This leads to the Tikhonov filtered solution which has the form (2.16) and uses the filter factors given in (2.21).

The choice of the parameter α is crucial. When α is small, solutions will tend to be more like the inverse or least squares solution and thus may be contaminated with inverted noise. When α is large, solutions will be weighted to favor regularity or smoothness in $\mathbf{x}$. Thus if α is too large, solutions will be very regular but not be a good fit for the original problem.

Much study has been invested in parameter selection methods to select truncation index k for TSVD or regularization parameter α for Tikhonov. For problems where the noise level η is provided or can be estimated from the data, parameter selection methods such as the Discrepancy Principle or UPRE can be used for estimating the desired parameter.

The discrepancy principle [27], [24], [37], [43] is intuitively easy to understand: if the norm of the noise in the right hand side is known in advance, then the parameter should be chosen so that the residual norm is approximately equal to the norm of the noise. One difficulty with this approach is that it can be very sensitive to estimate of the noise level, so it may be difficult to get good results. Furthermore, even with a good estimate, the discrepancy

principle has been known to produce over-regularized solutions[24].

Another parameter selection method is the L-curve criterion [22], [26], [28]. The L-curve is a log-log plot of the solution norm $\|\mathbf{x}_\alpha\|$ versus the residual norm $\|\mathbf{b} - \mathbf{A}\mathbf{x}_\alpha\|$ for a range of values under consideration for the parameter α. This gives a visualization of balance between the goodness of fit and the enforced regularity of solutions.

We observed heuristically that when α is small compared to the singular values of $\mathbf{A}$, solutions will tend toward $\mathbf{A}^{-1}\mathbf{b}$ or the least squares solution and thus will be dominated by inverted noise while for large α, regularity is enforced over goodness-of-fit. As a result, the L-curve graph is expected to have distinct vertical and horizontal parts representing these phases. The idea then is to choose α in the transition area, specifically at the point of maximum curvature. This method can fail when the solution is very smooth.

Generalized Cross Validation (GCV) [18], [47] is another commonly used parameter selection method which uses the statistical method of cross validation to predict elements of the solution and then choose α so that the prediction error is minimized. Other parameter slection methods include the residual periodogram [42] and the χ^2 test [36].

2.2.3 Hybrid Methods

In Chapter 1, we demonstrated the semiconvergence behavior that occurs when standard iterative methods are applied to ill-posed inverse problems. In particular, early iterations are able to produce approximations that correspond to decreasing reconstruction errors, but additional iterations result in large errors due to contamination by inverted noise. Early termination of the iterative process is one approach to regularization, also known as iterative regularization, but it is known that selecting a good stopping iteration can be very difficult.

Hybrid methods attempt to blend the best aspects of regularization techniques with iterative methods for solving ill-posed inverse problems. SVD based techniques are very good at isolating and removing inverted noise from small scale problems but are too computationally expensive for large scale problems. Iterative Krylov methods allow the substitution of a large problem with a smaller projected problem and often lead to good approximations in a small number of iterations. But iterative methods exhibit semiconvergence when applied to ill-posed inverse problems and it can be difficult to identify the transition between decreasing and increasing error.

We consider hybrid methods, where a regularization technique is applied to the smaller pro-

jected problem at each iteration. The intent is to fully utilize the rapid convergence of the iterative method with relatively low computational effort and to use regularization to stabilize solutions.

Following Hansen's argument in [30], it is easy to see that Tikhonov regularization on the projected problem at the kth iteration is equivalent to applying k iterations of the Krylov method to the Tikhonov problem (2.19). More specifically let $\mathbf{W}_k$ and $\mathbf{H}_{k+1,k}$ be the matrices produced by the Arnoldi process (or equivalently the matrices $\mathbf{V}_k$ and $\mathbf{B}_{k+1,k}$ from the Golub-Kahan bidiagonalization process).

Then at the kth iteration of GMRES or LSQR applied to (2.19), we seek a solution to

$$\min_{\mathbf{x}\in\mathcal{K}_k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 + \alpha^2\|\mathbf{x}\|^2 \tag{2.22}$$

which is given by $\mathbf{x}_k = \mathbf{W}_k\mathbf{y}_k$ where $\mathbf{y}_k$ solves

$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\beta\mathbf{e}_1 - \mathbf{H}_{k+1,k}\mathbf{y}\|^2 + \alpha^2\|\mathbf{y}\|^2. \tag{2.23}$$

Notice that this is just Tikhonov regularization applied to the projected problem (2.8). In each iteration a new Tikhonov problem is solved with updated $\mathbf{H}_{k+1,k}$ and so the regularization parameter α must be estimated in each iteration. This process ensures that the amount of regularization is correlated with the noise level in the projected problem. By comparison, if Tikhonov regularization is applied before employing an iterative method, then the regularization parameter is fixed through the iterations.

Hybrid methods where Tikhonov regularization is applied inside Golub-Kahan based iterations have been explored in [6], [7], [9], [11], [20], [32], [33], [40], and others. They have shown to be particularly effective at stabilizing iterations near the minimum relative error level on many test problems, and thus mitigating the semiconvergence phenomenon.

Hybrid methods where Tikhonov is applied within Arnoldi methods have been a subject of renewed interest recently [16], [13], [12] in the context of large scale ill-posed problems where the matrix transpose $\mathbf{A}^T$ is unavailable or otherwise undesirable to compute or store in memory. The idea is that by avoiding computations with the matrix transpose, large scale problems may be solved more quickly, and that applying regularization to the projected problem will have a regularizing effect on the resulting solutions. Further investigation of hybrid methods in the context of seeking good solution coefficients for a given basis will be considered in Section 3.2.

Chapter 3

Analysis of Arnoldi and Golub-Kahan Subspaces and Solution Methods for Ill-Posed Inverse Problems

Since Krylov subspace methods seek approximate solutions in a sequence of subspaces, it is natural to ask how well the exact solution can be approximated in each of the subspaces, to what extent approximations might improve in subsequent subspaces, and how well an iterative method is able to utilize the subspace at its disposal. Further, by investigating the choice of coefficients for a given basis we can analyze the extent to which a hybrid method is an improvement on the underlying minimum residual method.

In this chapter we answer these questions in the context of problems for which we know the exact solutions and give some examples to illustrate how this analysis can be useful and what limitations it has. In Section 3.1 we investigate the quality of the basis for representing solutions for the problem at hand, and in Section 3.2 we consider the use of the bases for computing solutions.

3.1 On the Quality of the Basis

Suppose that an exact solution $\mathbf{x}_{exact} \in \mathbb{R}^n$ exists and is known for an ill-posed inverse problem (1.2) where $\mathbf{A}$ is of full rank and either the Arnoldi process or the Golub-Kahan process is applied to produce a set of orthonormal vectors $\{\mathbf{w}_i\}_{i=1}^n$ [1] and thus form a basis for $\mathbb{R}^n$. Consequently $\mathbf{x}_{exact} \in \mathcal{R}(\mathbf{W})$ where

[1] In Section 2.1, we used $\mathbf{w}_i$ for Arnoldi vectors and $\mathbf{v}_i$ for Golub-Kahan vectors, but the following analysis applies equally to any basis

$$\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_n].$$

Define the vector $\mathbf{p} \in \mathbb{R}^n$ as

$$\mathbf{p} = \mathbf{W}^T \mathbf{x}_{exact} \tag{3.1}$$

so that its components are given as

$$p_i = \mathbf{w}_i^T \mathbf{x}_{exact} \qquad \text{for } i = 1, \ldots, n. \tag{3.2}$$

We know the scalar product for any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, satisfies

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Then since the vectors $\mathbf{w}_i$ each have unit norm, we can define

$$\hat{\mathbf{p}} = \frac{\mathbf{p}}{\|\mathbf{x}_{exact}\|} = \begin{bmatrix} \cos\theta_1 \\ \cos\theta_2 \\ \vdots \\ \cos\theta_n \end{bmatrix}$$

where θ_k is the angle between $\mathbf{x}_{exact}$ and $\mathbf{w}_k$, and thus $|\hat{p}_k| \leq 1$ for all values of k. Note that $|\hat{p}_k|$ close to one indicates a small angle and $|\hat{p}_k| \approx 0$ indicates a basis vector that is nearly orthogonal to $\mathbf{x}_{exact}$, so the former contribute useful information to the reconstruction of the exact solution while the latter do not.

Computing the full set of basis vectors is undesirable in practice since the aim of the iterative method is to produce a high quality approximate solution in as few iterations (ie. with as little computational work) as possible. So a basis can be said to be well suited for a problem if the $|\hat{p}_i|$ corresponding to $\mathbf{w}_i$ are relatively large (eg. $|\hat{p}_i| > .3$) in the early iterations ($i \leq k << n$) and small in the later iterations ($i > k$). In this scenario, $\mathbf{x}_{exact}$ can be well represented in $\mathcal{R}(\mathbf{W}_k)$ and little is lost if iterations are terminated at k.

On the other hand, if $|\hat{p}_i|$ are relatively large for $i = 1, \cdots, n$, then all the basis vectors, including the last ones, are important for an accurate reconstruction of $\mathbf{x}_{exact}$ and early termination will yield poor reconstructions.

Besides helping to describe the quality of the basis for a given problem, components in (3.2) give the optimal solution for the minimizing the residual in a Krylov subspace.

Lemma 1. *Suppose the matrix* $\mathbf{W}_k \in \mathbb{R}^{n\times k}$ *with* $k < n$ *has orthonormal columns* $\mathbf{w}_1, \ldots, \mathbf{w}_k$ *and let* $\mathbf{x}_{exact} \in \mathbb{R}^n$ *be a given vector. Then the vector* $\hat{\mathbf{y}}$ *in* $\mathbb{R}^k$ *that satisfies*

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{x}_{exact} - \mathbf{W}_k\mathbf{y}\| \tag{3.3}$$

is given by $\mathbf{y} = \mathbf{W}_k^T\mathbf{x}_{exact}$.

Proof. Under these assumptions, the least squares solution is found by solving the normal equation

$$\mathbf{W}_k^T\mathbf{W}_k\mathbf{y} = \mathbf{W}_k^T\mathbf{x}_{exact}. \tag{3.4}$$

Since $\mathbf{W}_k$ contains orthonormal columns,

$$\mathbf{W}_k^T\mathbf{W}_k = \mathbf{I}_k \tag{3.5}$$

and the desired conclusion follows. □

It is clear then that any approximate solution taken from the $\mathcal{R}(\mathbf{W}_k)$ (eg. GMRES or Arnoldi-Tikhonov approximate solutions when $\mathbf{W}$ is determined by the Arnoldi process) will have relative error at least as large as the relative error in the optimal solution $\mathbf{W}_k\mathbf{p}_k$ where

$$\mathbf{p}_k = \mathbf{W}_k^T\mathbf{x}_{exact}.$$

By examining the relative errors corresponding to the optimal solutions for $k = 1, \ldots, n$ we can judge the qualities of the bases generated by the Arnoldi or Golub-Kahan processes for a particular problem and compare them.

To illustrate this, we return to the `heat` problem from the first chapter. The problem has been set up as before. In this instance, the right-hand side $\mathbf{b}$ is contaminated with a noise level of $\eta = 10^{-3}$. The Arnoldi and Golub-Kahan processes are each used with full reorthogonalization to compute their respective bases and the coefficients $\hat{\mathbf{p}}$ corresponding to the optimal solution in the Krylov subspaces are shown in the left plot of Figure 3.1.

The $|\hat{p}_i|$ for the vectors generated by the Golub-Kahan process exhibit the general features of a well-suited basis. The first few values of $|\hat{p}_i|$ are large and the remaining are relatively

Figure 3.1: `heat` problem, $\kappa = 1$, $\eta = 10^{-3}$: Left - values of $|\hat{p}_i|$ corresponding to optimal coefficients in Arnoldi and Golub-Kahan bases. Right - error in the partial reconstructions. Golub-Kahan provides a better basis

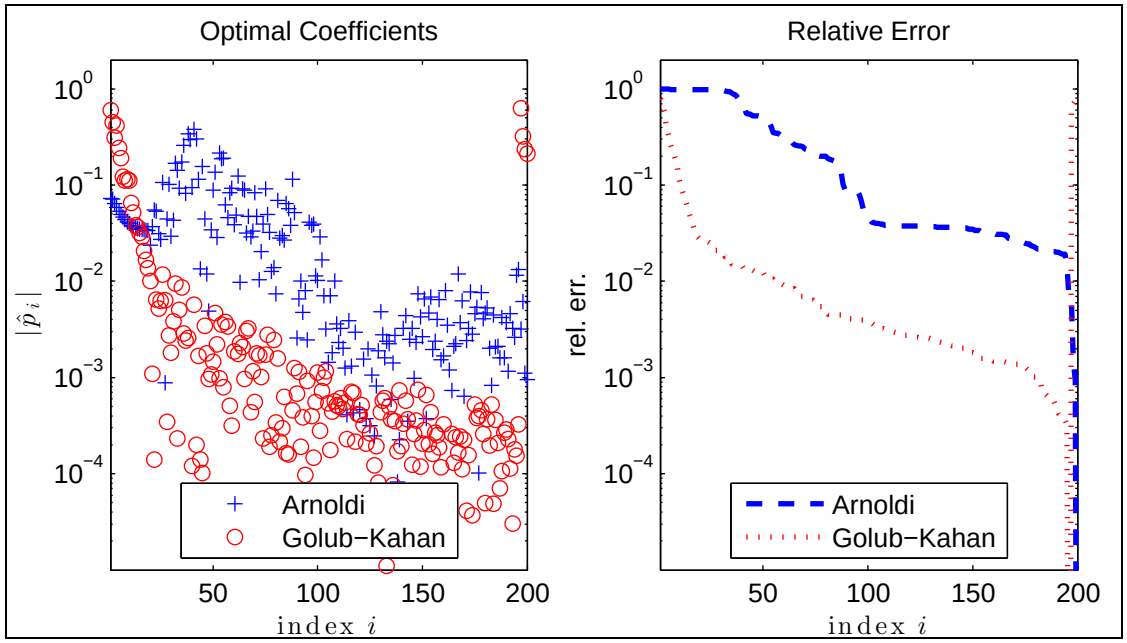


Figure 3.2: `heat` problem, $\kappa = 1$, $\eta = 10^{-3}$: Partial Reconstructions in Each Basis. Optimal Arnoldi solutions are poor in early iterations and never become as accurate as early Golub-Kahan optimal solutions

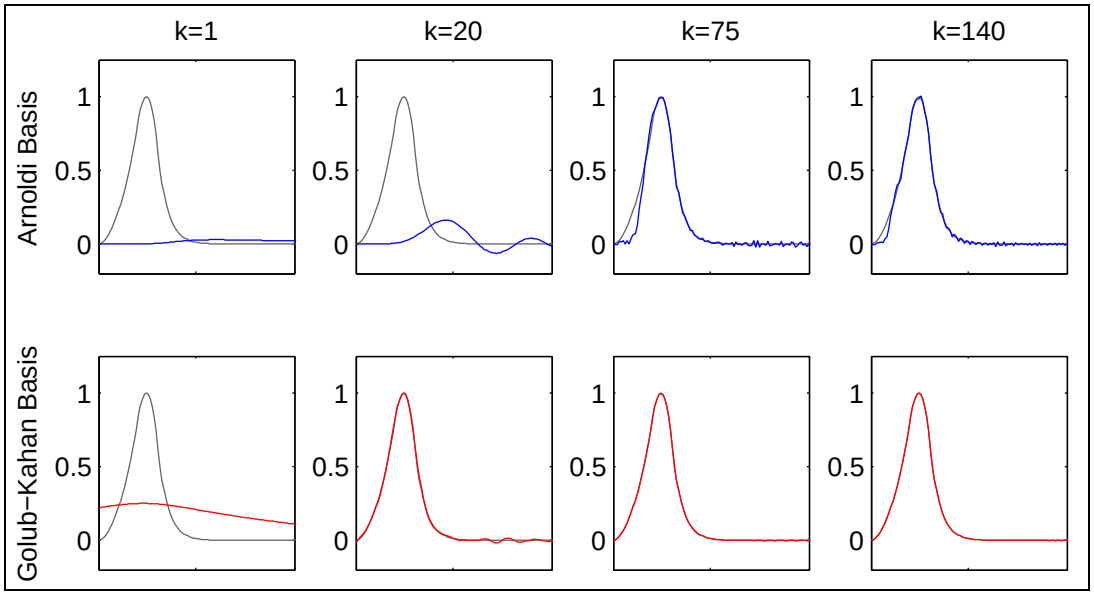
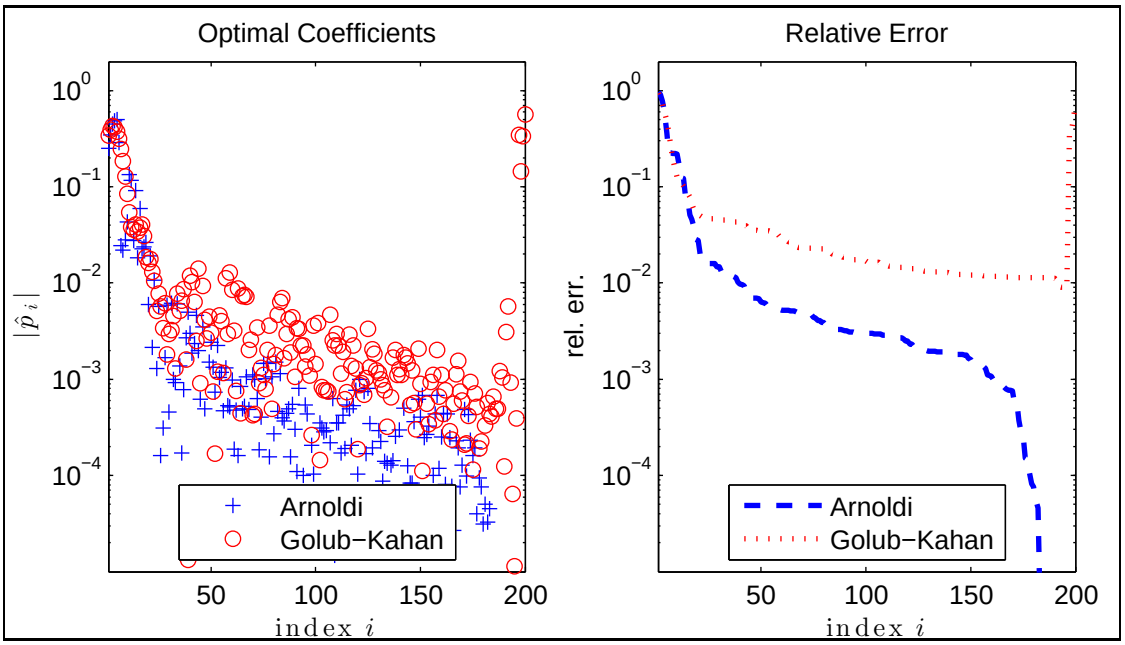


Figure 3.3: `heat` problem, $\kappa = 1$, $\eta = 10^{-3}$ with alternate $\mathbf{x}_{exact}$: the left plot shows values of $|\hat{p}_i|$ in Arnoldi and Golub-Kahan bases and the right plot show error in the optimal reconstructions for each k. Arnoldi provides a better basis in this example.



small.[2]

By comparison, the Arnoldi basis is not well-suited to this problem. The $|\hat{p}_i|$ associated with its initial vectors are small and so early partial reconstructions are poor. Also, the later $|\hat{p}_i|$ are consistently larger than those for the Golub-Kahan process by an order of magnitude. From this observation, we can infer that the overall quality of the reconstruction depends more on the later basis vectors from Arnoldi.

The error plot in Figure 3.1 makes it clear that for this problem, the Golub-Kahan basis is superior to the Arnoldi basis since the optimal approximations for $\mathbf{x}_{exact}$ in the Golub-Kahan bases have smaller relative error than those in the Arnoldi bases at any iteration k.

Figure 3.2 shows reconstructions at various iterations, along with the true solution. These plots confirm the quality of the basis.

In our observations, this general behavior is consistent for all noise levels in this problem for

[2]the exceptions are the $|\hat{p}_{197}|, \ldots, |\hat{p}_{200}|$ which are large and correspond to a loss of orthogonality among the basis vectors despite the reorthogonalization.

a given right-hand side. Hansen and Jensen in [25] studied how noise propogates through the Arnoldi and Golub-Kahan Krylov basis vectors in the context of image deblurring.

On the other hand, the same problem with a different $\mathbf{x}_{exact}$ (and thus a different right-hand side $\mathbf{b}$) can exhibit very different results. For instance, when $\mathbf{x}_{exact}$ is taken to be the reflection through a vertical line in the center of the default $\mathbf{x}$ from the `heat` problem, the Arnoldi process yields a basis that is superior to that of Golub-Kahan. This is shown in the plots of the $|\hat{p}_i|$ and relative errors in Figure 3.3.

This illustrates that the quality of a Krylov subspace basis depends on both the exact solution $\mathbf{x}_{exact}$ and on the matrix $\mathbf{A}$. Here we have shown two examples, one where the Arnoldi process is better and another where the Golub-Kahan basis is better. Only a small number of iterations are typically used in practice, and in the latter example the Arnoldi and Golub-Kahan bases are of comparable quality for the initial iterations. In the next section we look at how well these bases are utilized by some iterative methods.

3.2 On the Utilization of a Given Basis for Computing Solutions

In this section, we fix a basis and investigate how well iterative minimum residual and hybrid methods utilize the given basis. In Lemma 1 we established the optimal solutions in the kth Krylov subspace of the approximation as

$$\mathbf{x}_k^{(opt)} = \mathbf{W}_k \mathbf{W}_k^T \mathbf{x}_{exact} = \arg\min_{\mathbf{x} \in \mathcal{K}_k} \|\mathbf{x}_{exact} - \mathbf{x}\|.$$

For a given basis we can immediately conclude that the relative error of an approximate solution produced by a minimum residual method or a Tikhonov regularized hybrid approximation in the same basis must be a least as large.

By comparing the relative errors of the optimal solutions, minimum residual method solutions, and hybrid method solutions, we can evaluate how well these methods are using the given basis and in what manner hybrid methods might yield improved solutions. Note for comparison that the approximate solutions of our iterative methods, $\mathbf{x}_k^{(MR)}$, and hybrid methods, $\mathbf{x}_k^{(Tik)}$ satisfy

$$\mathbf{x}_k^{(MR)} = \arg\min_{\mathbf{x}\in\mathcal{K}_k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$$
$$\mathbf{x}_k^{(Tik)} = \arg\min_{\mathbf{x}\in\mathcal{K}_k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 + \alpha^2\|\mathbf{x}\|^2.$$

Since each method uses a different criteria to determine solutions, we can expect to have varied results depending on the problem. Since Arnoldi and Golub-Kahan based methods find solutions in different bases, we can expect to get different results when each is applied to a particular problem. These variances are readily observed and we give some examples and include information about optimal solutions in the basis for comparison.

As the primary interest of this thesis, all examples are nonsymmetric and severely ill-conditioned. Full reorthogonalization of the basis vectors is performed during both the Arnoldi and Golub-Kahan processes in each of the examples so that the results reflect as much as possible their implementation in exact arithmetic. Tikhonov regularization is used in the hybrid methods where the regularization parameter α is chosen to minimize the error between the Tikhonov solution and the true solution. A parameter selection method with less optimal results would normally be used in practice, but the purpose here is to illustrate the limitations in the optimal scenario.

Example 3.2.1: Inverse Laplace Transformation

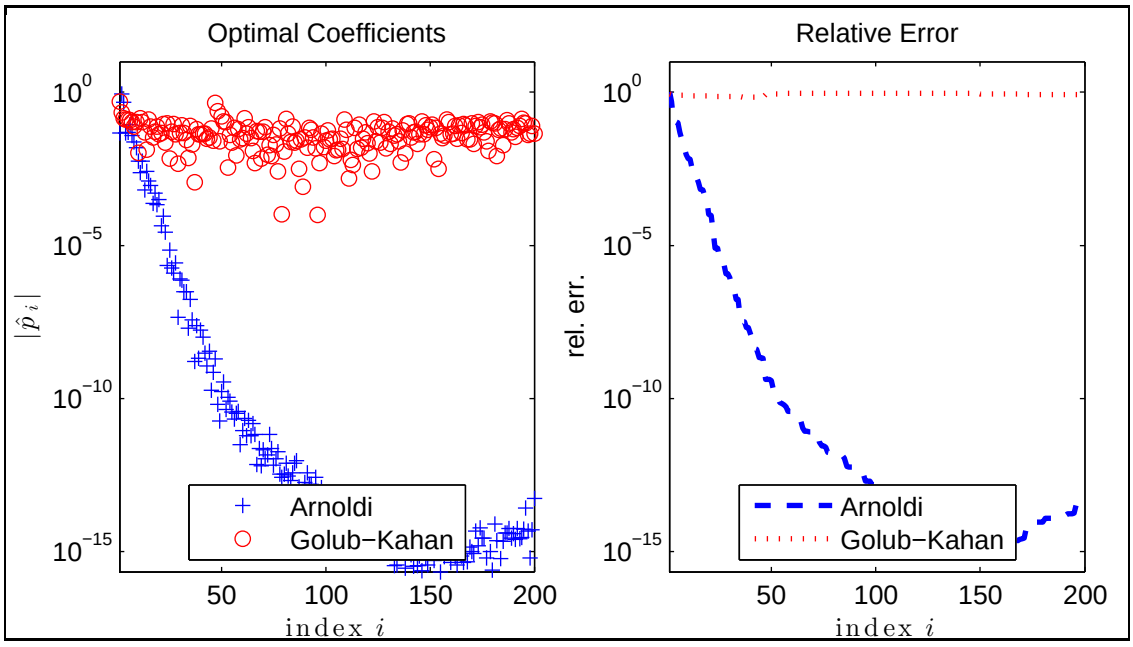
The `i_laplace` test problem from [23] gives a discretization of the inverse Laplace transformation by means of Gauss-Laguerre quadrature. Four corresponding solutions and right-hand sides are provided in the software. We use the second with size $n = 200$ and add 0.1% Gaussian white noise to the right-hand side.

Figure (3.4) shows the magnitudes of the optimal coefficients for the construction of $\mathbf{x}_{exact}$ in the Arnoldi and Golub-Kahan bases. Observe that in the Arnoldi basis, early magnitudes are relatively large but then decay quickly and remain small. As explained in the previous section, this indicates that the Arnoldi basis is good for this problem. This assessment is reflected in the plot of relative errors in the same figure.

In contrast, the magnitudes of the Golub-Kahan coefficients remain relatively large (> 0.05) for all values of k.

Figure (3.5) shows plots of the relative errors and residuals for both Arnoldi and Golub-Kahan bases. We compare results for the kth approximate solutions generated as optimal

Figure 3.4: Example 3.2.1 - `i_laplace` test problem #2, with noise level $\eta = 10^{-3}$: Arnoldi basis allows good reconstruction of $\mathbf{x}_{exact}$ but Golub-Kahan basis does not.



in each Krylov basis, the corresponding iterative minimum residual method (ie. GMRES or LSQR), and the solutions obtained when Tikhonov regularization applied to the projected problem (ie. Arnoldi Tikhonov (ATik) and Golub-Kahan Tikhonov (GKTik)).

Regardless of the basis, the relative errors for the iterative minimum residual and hybrid methods are bounded below by the relative error curve for the optimal solutions. For the Arnoldi basis, both the GMRES error curve and the Arnoldi-Tikhonov curve generally follow the trend of the basis in the initial iterations. Thus these approximations make good use of the basis. Applying Tikhonov regularization to the projected problem in the Arnoldi basis (ATik) improves the GMRES solution briefly, then increases to a higher relative error and unfortunately stabilizes at this level for the remaining iterations. This illustrates the importance of well-considered stopping criteria for ATik methods.

While the Golub-Kahan basis is not very good, the error plots for LSQR and Golub-Kahan-Tikhonov show that these methods perform was as good they can for the first ten iterations. This doesn't carry much weight since the optimal error remains high in early iterations, but it is more than can be said of GMRES diverges sharply after iteration seven. Additionally, the Tikhonov regularized Golub-Kahan solutions stabilize close to the optimal error level for the basis even when LSQR solutions diverge. This manner of stabilization is not guaranteed, but seems to be common in Tikhonov regularized Golub-Kahan solutions and can be a useful

Figure 3.5: Example 3.2.1 - `i_laplace` test problem #2, with noise level $\eta = 0.1\%$. Left, top & (bottom): relative error norms & (residual norms) for basis optimal, GMRES, and AT solutions. Arnoldi methods fail to make good use of the basis but regularization helps. Right, top & (bottom): relative error norms & (residual norms) for Golub-Kahan basis optimal, LSQR, and GKT solutions. The Golub-Kahan basis is unsuited for this problem and restricts LSQR and GKT performance.

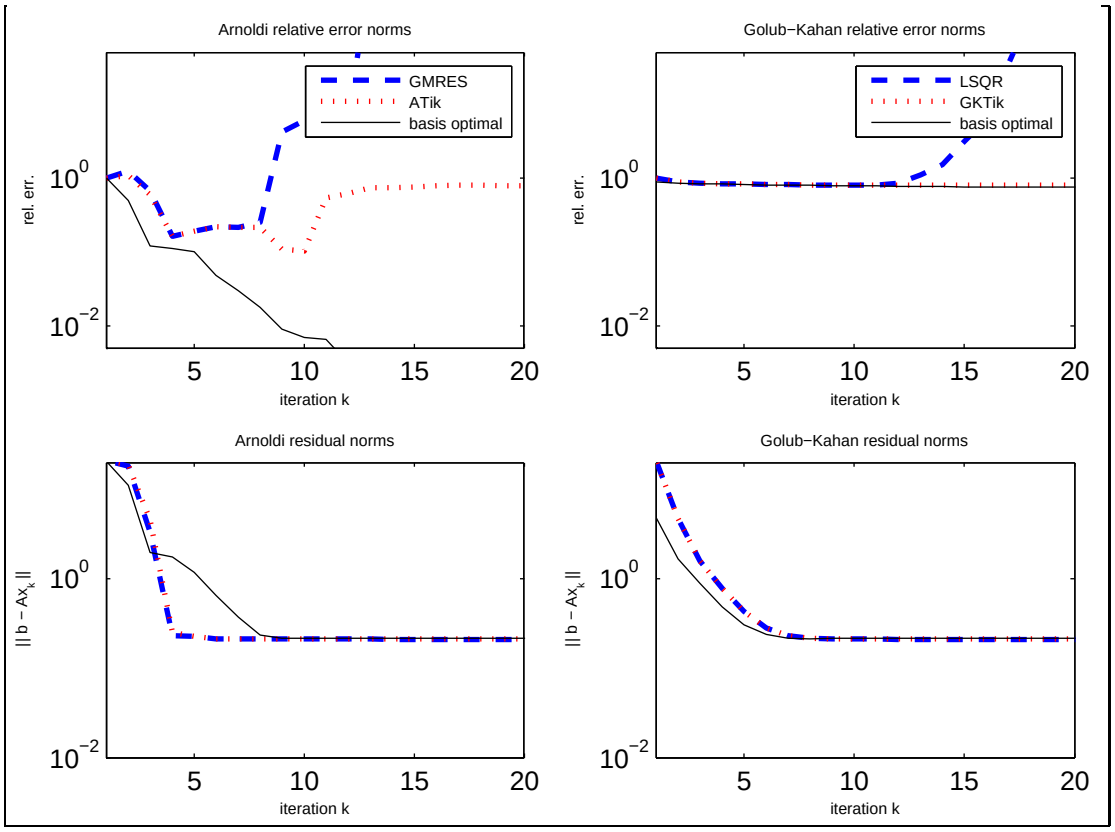
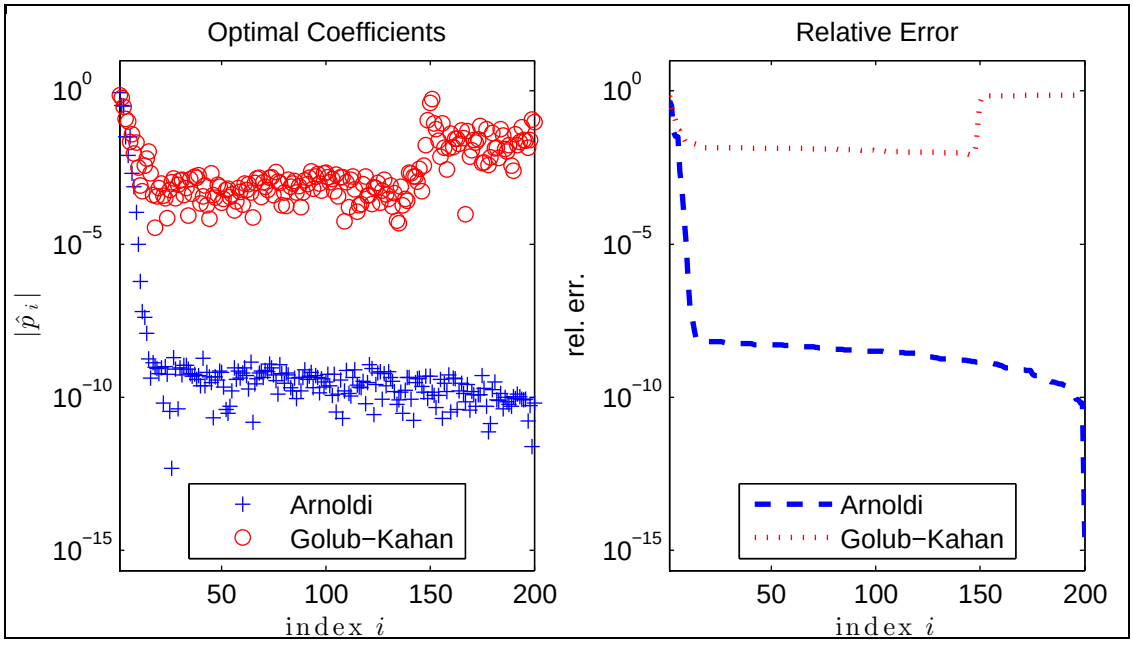


Figure 3.6: Example 3.2.2 - `baart` test problem, with noise level $\eta = 10^{-3}$: Arnoldi basis allows good reconstruction of $\mathbf{x}_{exact}$ but Golub-Kahan basis does not.



advantage [33].

Plots of residual norms are included in Figure (3.5) to reiterate the marked difference in the trends of the residual norms and relative error norm for ill-posed inverse problems.

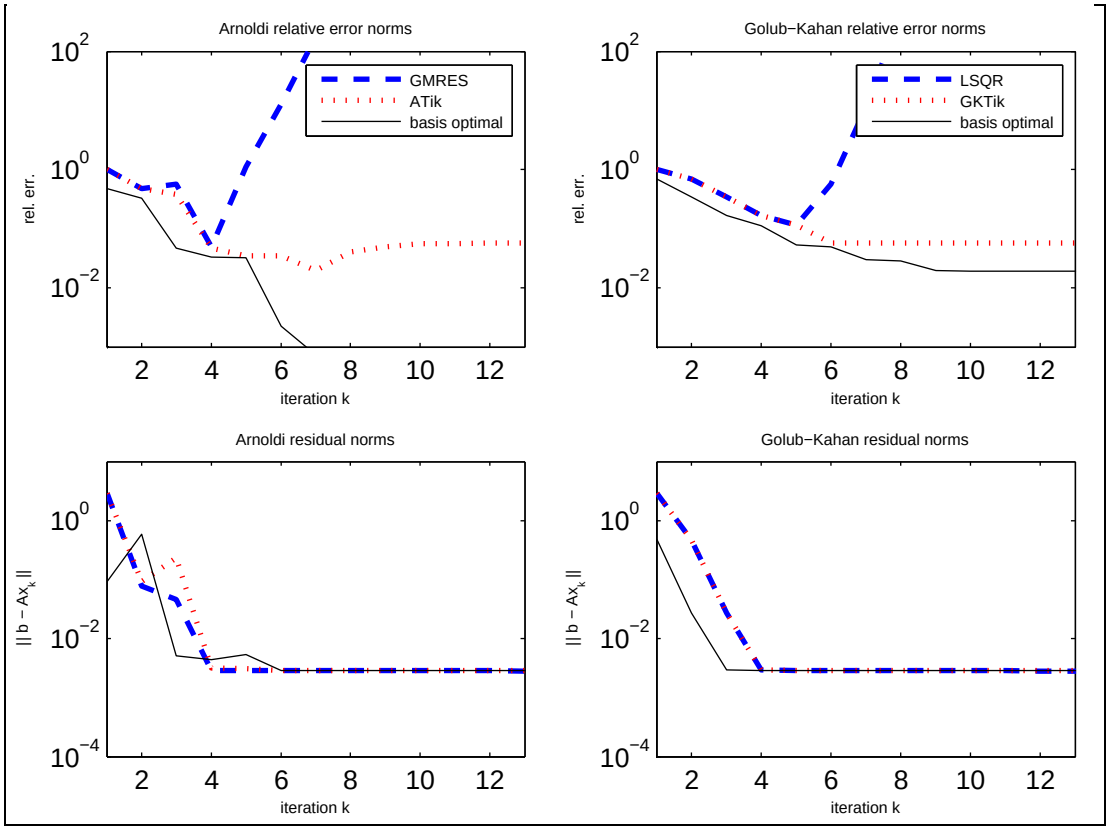
Example 3.2.2: `baart` Problem

The `baart` problem is another nonsymmetric problem included in [23] which has been used as an example of a problem for which methods based on the Arnoldi process result in better solutions than Golub-Kahan. We use it with relative noise level $\eta = 10^{-3}$ and problem size 200.

Figure 3.6 shows the magnitudes of the coefficients of the exact solution in both the Arnoldi and Golub-Kahan bases and the relative errors of these optimal solutions in the bases. The Arnoldi basis very rapidly allows accurate reconstructions. Apparently (as has been noted in [30]) the initial basis vectors produced by the Arnoldi process are very well suited for the problem. The Golub-Kahan basis does not do so well and we can see that the optimal solutions in any Golub-Kahan basis will not have relative error less than 0.01.

Figure (3.7) shows the relative error norm and residual norm curves for the optimal solu-

Figure 3.7: Example 3.2.2 - `baart` test problem with noise level $\eta = 10^{-3}$. Left, top & (bottom): relative error norms and (residual norms) for basis optimal solutions, GMRES solutions, and Arnoldi-Tikhonov solutions. Right, top & (bottom): relative error norms and (residual norms) for basis optimal solutions, LSQR solutions, and Golub-Kahan-Tikhonov solutions.



tions in the bases, the iterative minimum residual method solutions, and the hybrid method solutions. GMRES arrives at its best solution in the fourth iteration and the error sharply increases afterward. Tikhonov regularization stabilizes this behavior. By comparison with the optimal solution errors, we can see that GMRES and AT utilize the Arnoldi basis well in those initial iterations, but never achieves the extraordinarily low error that is possible in the Arnoldi basis.

In the Golub-Kahan basis, we see that LSQR and the corresponding regularized solutions are near optimal in the basis until the point at which LSQR begins to semiconverge. The regularized solutions remain stable at the point of lowest error on the LSQR curve. This is exactly the behavior we desire when employing hybrid methods to solve ill-posed inverse problems.

3.3 Discussion and Summary of Results

Our examples show how the optimal coefficients of a basis and the relative error curve of optimal solutions in the basis can be used to help interpret the results of applying iterative and hybrid methods to some test problems. Plots such as those presented in section 3.1 can provide at-a-glance insight into the suitability of a basis for reproducing a solution and also allows comparisons between bases.

The examples show that an Arnoldi basis is better for some problems while a Golub-Kahan basis is better for others, and they also showed how a different exact solution can have a strong influence on the basis quality. But we also found that basis quality does not fully answer the question of how well iterative methods using the bases will perform.

When applied to nonsymmetric ill-posed problems, methods based on the Golub-Kahan process (LSQR and GKTik) seem to produce solutions which are very close to optimal until the semiconvergence. When this occurs, hybrid regularization stabilizes solutions. On the other hand, methods based on the Arnoldi process (GMRES and ATik) often seem to have difficulty fully utilizing the basis for more than the first few iterations. The `baart` problem shows that there are scenarios where hybrid regularization stabilizes solutions, while the `i_laplace` problem shows that ATik solutions may not stabilize.

It should also be mentioned that using the Arnoldi basis provides an advantage by not requiring the matrix transpose and can require less computational work in early iterations by avoiding the matrix-vector product with $\mathbf{A}^T$, particularly when compared to a Golub-Kahan implementation with reorthogonalization. So neither Arnoldi nor Golub-Kahan processes can be completely discarded in favor of the other.

Chapter 4

Conclusion

Ill-posed inverse problems are difficult to solve, and regularization is needed to compute accurate solutions. Iterative methods such as LSQR and GMRES seek solutions in a particular Krylov subspace which is spanned by a set of orthonormal basis vectors. In this thesis, we investigate the use of the Arnoldi and Golub-Kahan process to construct these bases when seeking solutions to nonsymmetric ill-posed inverse problems.

For a given problem, we investigate the quality of the basis by looking at scaled optimal solution coefficients in the given basis and relative errors of reconstructions. Although such an analysis is not feasible in practice, it provides insight into which basis is better for a particular problem. Then, for a given basis, we investigate its use for computing solutions. We show that hybrid methods can better utilize the basis by computing nearly optimal solution coefficients and can thus overcome the semi-convergence behavior that is typical of minimum residual methods such as LSQR and to a lesser extent GMRES when applied to ill-posed inverse problems.

This work provides a way to separately judge the quality of a basis and the use of that basis by an iterative method. This may be useful to future work in several areas. In particular, it may be used to help analyze methods which propose to augment an Arnoldi or Golub-Kahan basis with vectors intended to aid in the reconstructing problematic features of exact solutions such as discontinuities. When additive noise is present in observations, we would like to know the manner in which noise is propagated among the basis vectors, when it might be damped, and whether certain bases provide better damping [29]. 'Hybrid' Tikhonov regularization works particularly well with Golub-Kahan based methods and a similarly strong pairing with Arnoldi methods would help make them more appealing to use in general.

Bibliography

[1] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), p. 1217.

[2] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math, 9 (1951), pp. 17–29.

[3] A. Bakushinsky and A. Goncharsky, *Ill-posed Problems: Theory and Applications*, Springer Netherlands, 1994.

[4] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, CRC press, 2010.

[5] P. N. Brown and H. F. Walker, *GMRES on (nearly) singular systems*, SIAM Journal on Matrix Analysis and Applications, 18 (1997), pp. 37–51.

[6] D. Calvetti, G. H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT Numerical Mathematics, 39 (1999), pp. 603–619.

[7] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, *Tikhonov regularization and the L-curve for large discrete ill-posed problems*, Journal of Computational and Applied Mathematics, 123 (2000), pp. 423–446.

[8] J. Chung, S. Knepper, and J. G. Nagy, *Imaging Handbook of Mathematical Methods in Imaging*, Springer, New York, 2011, ch. Large-Scale Inverse Problems in Imaging, pp. 43–86.

[9] J. Chung, J. G. Nagy, and D. P. O'Leary, *A weighted GCV method for Lanczos hybrid regularization*, Electronic Transactions on Numerical Analysis, 28 (2008), pp. 149–167.

[10] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.

[11] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT Numerical Mathematics, 17 (1977), pp. 134–145.

[12] S. Gazzola and J. G. Nagy, *Generalized Arnoldi-Tikhonov method for sparse reconstruction*, SIAM Journal on Scientific Computing, 36 (2014), pp. B225–B247.

[13] S. Gazzola and P. Novati, *Multi-parameter Arnoldi-Tikhonov methods*, Electronic Transactions on Numerical Analysis, 40 (2013), pp. 452–475.

[14] ——, *Automatic parameter setting for Arnoldi-Tikhonov methods*, Journal of Computational and Applied Mathematics, 256 (2014), pp. 180–195.

[15] S. Gazzola, P. Novati, and M. Russo, *On Krylov projection methods and Tikhonov regularization*, (2014).

[16] S. Gazzola, P. Novati, and M. R. Russo, *Embedded techniques for choosing the parameter in Tikhonov regularization*, arXiv preprint arXiv:1307.0334, (2013).

[17] G. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial & Applied Mathematics, Series B: Numerical Analysis, 2 (1965), pp. 205–224.

[18] G. H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.

[19] C. Groetsch, *Integral equations of the first kind, inverse problems and regularization: a crash course*, in Journal of Physics: Conference series, vol. 73, IOP Publishing, 2007, p. 012001.

[20] M. Hanke, *On Lanczos based methods for the regularization of discrete ill-posed problems*, BIT Numerical Mathematics, 41 (2001), pp. 1008–1018.

[21] P. C. Hansen, *The discrete Picard condition for discrete ill-posed problems*, BIT Numerical Mathematics, 30 (1990), pp. 658–672.

[22] ——, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM review, 34 (1992), pp. 561–580.

[23] ——, *Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems*, Numerical algorithms, 6 (1994), pp. 1–35.

[24] ——, *Discrete Inverse Problems: Insight and Algorithms*, vol. 7, SIAM, 2010.

[25] P. C. Hansen and T. K. Jensen, *Noise propagation in regularizing iterations for image deblurring*, Electronic Transactions on Numerical Analysis, 31 (2008), pp. 204–220.

[26] P. C. Hansen, T. K. Jensen, and G. Rodriguez, *An adaptive pruning algorithm for the discrete L-curve criterion*, Journal of computational and applied mathematics, 198 (2007), pp. 483–492.

[27] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, vol. 3, SIAM, 2006.

[28] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM Journal on Scientific Computing, 14 (1993), pp. 1487–1503.

[29] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT Numerical Mathematics, 49 (2009), pp. 669–696.

[30] T. K. Jensen and P. C. Hansen, *Iterative regularization with minimum-residual methods*, BIT Numerical Mathematics, 47 (2007), pp. 103–120.

[31] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160, Springer, 2005.

[32] M. E. Kilmer, P. C. Hansen, and M. I. Espanol, *A projection-based approach to general-form Tikhonov regularization*, SIAM Journal on Scientific Computing, 29 (2007), pp. 315–330.

[33] M. E. Kilmer and D. P. O'Leary, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM Journal on matrix analysis and applications, 22 (2001), pp. 1204–1221.

[34] C. Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 33–53.

[35] R. M. Larsen, *Lanczos bidiagonalization with partial reorthogonalization*, DAIMI Report Series, 27 (1998).

[36] J. L. Mead and R. A. Renaut, *A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems*, Inverse Problems, 25 (2009), p. 025002.

[37] V. A. Morozov, *On the solution of functional equations by the method of regularization*, in Soviet Math. Dokl, vol. 7, 1966, pp. 414–417.

[38] J. L. Mueller and S. Siltanen, *Linear and nonlinear inverse problems with practical applications*, vol. 10, SIAM, 2012.

[39] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM Journal on Matrix Analysis and Applications, 13 (1992), pp. 778–795.

[40] D. P. O'Leary and J. A. Simmons, *A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems*, SIAM Journal on Scientific and Statistical Computing, 2 (1981), pp. 474–489.

[41] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, Journal of the ACM (JACM), 9 (1962), pp. 84–97.

[42] B. W. Rust and D. P. O'Leary, *Residual periodograms for choosing regularization parameters for ill-posed problems*, Inverse Problems, 24 (2008), p. 034005.

[43] O. Scherzer, *The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems*, Computing, 51 (1993), pp. 45–60.

[44] A. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, in Soviet Math. Dokl., vol. 5, 1963, pp. 1035–1038.

[45] H. A. Van der Vorst, *Iterative Krylov Methods for Large Linear Systems*, vol. 13, Cambridge University Press, 2003.

[46] C. R. Vogel, *Computational Methods for Inverse Problems*, vol. 23, SIAM, 2002.

[47] G. Wahba, *Spline Models for Observational Data*, vol. 59, SIAM, 1990.

[48] H. F. Walker, *Implementation of the GMRES method using Householder transformations*, SIAM Journal on Scientific and Statistical Computing, 9 (1988), pp. 152–163.