JOINT INVERSION

OF COMPACT OPERATORS

by

James Ford

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Mathematics

Boise State University

August 2017

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

James Ford

Thesis Title: Joint Inversion of Compact Operators

Date of Final Oral Examination: 24 May 2017

The following individuals read and discussed the thesis submitted by student James Ford, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Jodi Mead, Ph.D.   Chair, Supervisory Committee

John Bradford, Ph.D.   Member, Supervisory Committee

Grady Wright, Ph.D.   Member, Supervisory Committee

The final reading approval of the thesis was granted by Jodi Mead, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

dedicated to Cassidy Carroll

ACKNOWLEDGMENTS

ABSTRACT

The first mention of joint inversion came in [22], where the authors used the singular value decomposition to determine the degree of ill-conditioning in inverse problems. The authors demonstrated in several examples that combining two models in a joint inversion, and effectively stacking discrete linear models, improved the conditioning of the problem. This thesis extends the notion of using the singular value decomposition to determine the conditioning of discrete joint inversion to using the singular value expansion to determine the well-posedness of joint linear operators. We focus on compact linear operators related to geophysical, electromagnetic subsurface imaging.

The operators are based on combining Green's function solutions to differential equations representing different types of data. Joint operators are formed by extending the concept of stacking matrices to one of combining operators. We propose that the effectiveness of joint inversion can be evaluated by comparing the decay rate of the singular values of the joint operator to those from the individual operators. The joint singular values are approximated by an extension of the Galerkin method given in [9, 18]. The approach is illustrated on a one-dimensional ordinary differential equation where slight improvement is observed when naively combining differential equations. Since this approach relies primarily on the differential equations representing data, it provides a mathematical framework for determining the effectiveness of joint inversion methods. It can be extended to more realistic differential equations in order to better inform the design of field experiments.

TABLE OF CONTENTS

LIST OF FIGURES

1

CHAPTER 1

INTRODUCTION

This thesis is motivated by electromagnetic imaging of the Earth's subsurface using two data collection methods – ground penetrating radar (GPR) and electrical resistivity (ER), both depicted in Figure 1.1. The ER survey on the left uses electrodes to introduce and measure current within the Earth's subsurface, from which the electrical conductivity can be estimated within a localized region. The GPR survey on the right involves a transmitter emitting radar pulses into the subsurface. The energy reflected back by dielectric interfaces within the subsurface is recorded by the receiver, and conductivity can also be inferred. Joint inversion of these data can give conductivity estimates on a vast range of frequency bands from 10^0 to 10^9 Hz. A comprehensive introduction to geophysical electromagnetic imaging can be found in [19, 23].
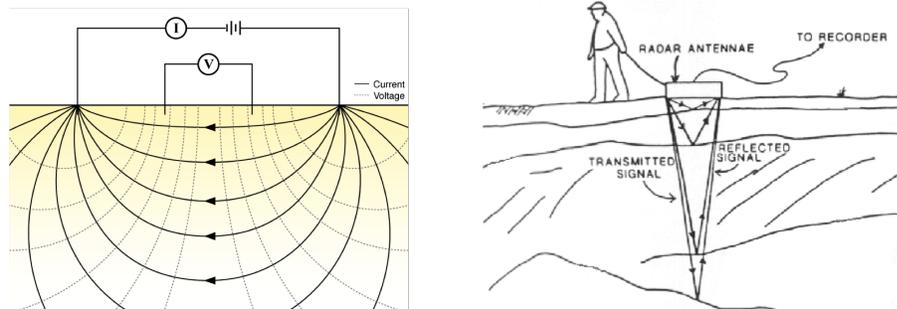
Figure 1.1: Electrical resistivity setup [4] (left) and ground penetrating radar [5] (right)

Maxwell's equations govern the behavior of both ER and GPR, relating input sources to observable responses [14]. If we assume the domain we are interested in imaging is linear and isotropic, Maxwell's equations can be decoupled into the wave equation and the diffusion equation. Under these assumptions, the damped wave equation governs GPR:

$$\varepsilon \frac{\partial^2 E}{\partial t^2} + \sigma \frac{\partial E}{\partial t} = \frac{1}{\mu} \nabla^2 E + f,$$

where E is the electric field component, f is the source; while the parameters μ, ε, and σ are scalar functions of space representing magnetic permeability, dielectric permittivity and conductivity, respectively. The forward problem takes the parameters as input and produces the electric field as a function of space and time. The inverse problem takes discrete measurements of the electric field and produces parameter estimates that can be used to image the Earth's subsurface.

Alternatively, the diffusion equation governs ER and assumes no time dependence of the electric field:

$$-\nabla \cdot \sigma \nabla \phi = \nabla \cdot J_s,$$

where J_s is the source current density and ϕ is the electric field potential. Similar to GPR, given discrete measurements of the electric field potential, the conductivity can be estimated through an inversion.

In the early 1900s, French mathematician Jacques Hadamard introduced the concept of a well-posed problem of mathematical physics [21]. He proposed that mathematical models of physical phenomena should have the properties that for every set of inputs, a solution exists, the solution is unique, and the problem is stable. i.e., the solutions behavior changes continuously with the inputs. A problem that does not satisfy these conditions is said to be ill-posed. Problems failing to meet

these conditions arise with great frequency in many areas of science and engineering, including electromagnetic subsurface imaging [11]. Since inverse problems are notoriously ill-posed, approaches to solving them will be the focus of this thesis.

The primary tools we will use for the analysis of inverse problems are the singular value decomposition (SVD) and its continuous analogue, the singular value expansion (SVE). We begin in Chapter 2 with a brief introduction to the solutions of discrete inverse problems using the SVD. Green's functions are then introduced since they continuously describe the solution of differential equations as a function of unknown parameters or sources. The corresponding operators are consequently decomposed using the SVE. A simple example is given with the focus on the behaviors of the corresponding singular values. Tikhonov regularization in the context of the SVE is also described.

In Chapter 3 we extend the concepts in Chapter 2 to joint inversion. Analysis of discrete joint inverse problems is somewhat straightforward because we simply view two models as one stacked matrix. However, in the continuous setting, the joint operator requires more careful consideration. Therefore we extend notions of Tikhonov regularization found in [6] to the more general direct sum of integral operators on Hilbert spaces, as in [8]. Approaches to calculating the continuous joint singular values are then given, and they include both analytical and numerical methods. Finally, in Chapter 4 we illustrate the approaches on a one-dimensional ODE. The singular values of the individual operators are found analytically, while those for the joint operator are found numerically. The singular values are compared and slight improvement with joint inversion is observed qualitatively through the decay rate of the singular values.

CHAPTER 2

INVERSION

2.1 Discrete inversion

Data collection is typically represented as a finite number of measurements taken at discrete points in space and time. It is therefore natural to first consider a discrete inverse problem, with observations denoted $b = (b_1, b_2, \ldots b_m)$ and parameters we seek to recover denoted $x = (x_1, x_2, \ldots x_n)$. Then, with g being the model relating x and b, we attempt to solve the system $g(x) = b$, or

$$
\begin{bmatrix}
g_1 (x_1, x_2, \ldots, x_n) \\
g_2 (x_1, x_2, \ldots, x_n) \\
\vdots \\
g_m (x_1, x_2, \ldots, x_n)
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_m
\end{bmatrix}
$$

for the parameter vector x. As previously mentioned, such problems may be ill-posed. A careful analysis of the model g can reveal if the problem has any solutions, a unique solution and if it is stable. Such insights can inform methods that allow us to proceed in any case.

Discrete inverse problems are well studied and general introductions to them can be found in [1, 10, 11, 12]. This section gives a brief overview of discrete inverse

problems as well as the methods used to analyze and solve them. This background will set the stage for the continuous analogues developed in this thesis.

2.1.1 Linear problems

The simplest case of discrete inverse problems occurs when the model and parameters are related linearly. That is,

$$
\begin{bmatrix}
g_1\left(x_1, x_2, \ldots, x_n\right) \\
g_2\left(x_1, x_2, \ldots, x_n\right) \\
\vdots \\
g_m\left(x_1, x_2, \ldots, x_n\right)
\end{bmatrix}
=
\begin{bmatrix}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{bmatrix}
$$

In this case, the model can be represented as matrix, and we solve $Ax = b$ to recover the parameters. We will devote our attention to linear problems. The assumption of linearity is not so restrictive, since most non-linear methods involve iteratively solving a linear problem. Therefore this thesis is devoted to analyzing and solving linear problems.

2.1.2 Singular value decomposition

The primary tool used for the analysis of linear inverse problems is the singular value decomposition (SVD). This algebraic decomposition represents a special case of the singular value expansion (SVE), which will be discussed in 2.2.2.

Theorem 2.1.1 (Singular Value Decomposition). Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank r. Then, there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that

$$A = U\Sigma V^T, \quad \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix},$$

where $\Sigma \in \mathbb{R}^{m \times n}$, $\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \ldots \sigma_r)$ and

$$\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0.$$

Letting v_j and u_j be the jth columns of V and U respectively, we have that

$$Av_j = \sigma_j u_j, \quad A^T u_j = \sigma_j v_j, \qquad \text{for } j = 1, \ldots, n$$

$$A^T u_j = 0, \qquad\qquad\qquad \text{for } j = n+1, \ldots, m$$

Proof. See [15]. □

This decomposition is enormously useful. For example, the norm of a vector is invariant under multiplication by an orthogonal matrix. So when considering $Ax = U\Sigma V^T x$, the scaling affect of the model can be isolated to Σ. This allows us to classify if solutions will exist and the conditioning of the model. Additionally, this can yield insights into the importance of each parameter, x_i. Parameters associated with large singular values will have a significantly greater influence on the outcome of the forward problem and can therefore be classified as more important [22].

In the case of a square matrix with full rank, $r = m = n$, we can very easily formulate the inverse of A using the SVD, and therefore the solution to $Ax = b$:

$$A^{-1} = V\Sigma^{-1}U^T, \quad x = V\Sigma^{-1}U^T b.$$

However, most applications result in non-square and or non-full rank matrices and

the inverse does not exist. In the case of full rank, overdetermined systems we can find the least squares solution to the problem; that is, the parameter vector that minimizes

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

The solution estimate in this case is

$$\hat{x} = \left(A^T A\right)^{-1} A^T b.$$

The generalized inverse allows us to define the solution estimate in terms of the SVD when A is not invertible. We construct the generalized inverse by first defining $\Sigma^\dagger \in \mathbb{R}^{n \times m}$ as

$$\Sigma^\dagger = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Then we can define the generalized inverse $A^\dagger$ as

$$A^\dagger = V \Sigma^\dagger U^T.$$

Further, we can define a generalized (or pseudo) inverse solution $x_\dagger$ by

$$x_\dagger = A^\dagger b$$
$$= V \Sigma^\dagger U^T b$$
$$= \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i.$$

It can be shown that $x_\dagger$ is the solution that minimizes $||b - Ax||_2^2$, [1].

2.1.3   Conditioning and Regularization

As stated previously, the third property that a well-posed problem should possess is stability, i.e., the solutions behavior changes continuously with the initial conditions. Typically, discrete problems are solved computationally. Therefore, the property we are interested in becomes numerical stability or conditioning. Given the SVD of a model, we are able to determine to what degree the inverse solution will be affected by perturbations in the inputs. The metric by which ill-conditioning is measured is called the condition number.

Definition 2.1.1. (Condition number) The condition number of A is defined to be $\kappa(A) = \sigma_1/\sigma_r$, where $r = rank(A)$.

As can be seen above, the generalized solution, $x_\dagger$, is dependent on the inverse of the singular values of A. If $\sigma_1^{-1} << \sigma_p^{-1}$, for some p where $1 < p \le r$, then the portion of the summation

$$\sum_{i=p}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

will dominate the solution. The ratio of the largest singular value to the smallest quantifies this effect. For $\kappa(A)$ large, the values in b associated with σ_r will be scaled by a much greater amount than the values associated with σ_1. Therefore, small perturbations in b can result in large changes to the solution. There are a vast number of regularization methods for addressing this issue of ill-conditioning [1, 11, 10]. We will focus on two in particular, the truncated singular value decomposition and Tikhonov regularization.

Ill-conditioning occurs when $\kappa(A) = \sigma_1/\sigma_r$ is large. This encourages us to truncate the smallest singular values of A up to σ_k for some $k < r$, so that $\sigma_1/\sigma_k << \sigma_1/\sigma_r$. More precisely, let

$$A_k = U\Sigma V^T, \quad \Sigma = \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix},$$

where $\Sigma_k = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$. Then the solution using this truncated matrix is

$$x_\dagger^k = A_k^\dagger b$$

$$= \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i.$$

The matrix A_k is the best rank k approximation to A [10]. If we cut off a small number singular values, we will have a very good approximation of the original matrix, and our inverse solution will be less susceptible to perturbations in b. The choice of truncation value is challenging and not within the scope of this thesis. Though, in general, it is advantageous to keep as many singular values as possible, while maintaining stability.

Tikhonov regularization is another method of combating ill-conditioning. Here, we solve a damped least squares problem

$$\min_x \|b - Ax\|_2^2 + \alpha^2 \|x\|_2^2,$$

or equivalently

$$\min_x \left\| \begin{bmatrix} b \\ 0 \end{bmatrix} - \begin{bmatrix} A \\ \alpha I \end{bmatrix} x \right\|_2^2.$$

For large enough α, this damped least squares problem always has a solution. The solution $x_\dagger^\alpha$, is dependent on α and can be expressed in terms of the SVD of A [1]. Namely,

$$x_\dagger^\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T b}{\sigma_i} v_i.$$

By writing the solution in this form, we see that the contribution of singular values smaller than alpha are damped out. However, we still have the challenge of choosing an α that improves the conditioning, but does not significantly change the problem. Methods for choosing the regularization parameter include the discrepancy principle, L-curve and generalized cross validation (GCV) [1], but these fall outside the scope of this thesis.

2.2 Inversion with compact operators

In this section, we extend the notions in 2.1 to continuous functions and compact linear operators. Since we are interested in physical phenomena modeled by differential equations, we wish to consider operators with some physical significance. Green's functions solutions of inhomogeneous differential equations determine the response of the physical system for given a source. The integral operators corresponding to Green's functions are therefore the compact operators we will investigate. They provide a physically motivated, yet abstract paradigm in which we may consider inverse problems. A thorough discussion of Green's functions and their applications can be found in [3].

2.2.1 Green's functions

Let $\mathcal{L}_A = \mathcal{L}_A(t)$ be a linear differential operator. Then the corresponding Green's function, $K_A(t, s)$ satisfies

$$\mathcal{L}_A K_A(t, s) = \delta(t - s), \tag{2.1}$$

where δ denotes the delta function. We solve for the Green's function as a function of s over the same domain Ω as t to determine the response of the system to an impulse

(delta function).

Given the Green's function, we can find the solution to the inhomogeneous equation $\mathcal{L}_A u(t) = f(t)$ for a given specific, f. This is accomplished by integrating both sides of equation (2.1) against $f(s)$ over s in the domain Ω,

$$\int_\Omega \mathcal{L}_A K_A(t,s) f(s) ds = \int_\Omega \delta(t-s) f(s) ds.$$

Since $\mathcal{L}_A$ is a an operator acting only on t, it can be moved outside of the integral. Therefore we have,

$$\mathcal{L}_A \left(\int_\Omega K_A(t,s) f(s) ds \right) = f(t).$$

Thus the solution to $\mathcal{L}_A u(t) = f(t)$ for a given source f is

$$u(t) = \int_\Omega K_A(t,s) f(s) ds.$$

In other words, integrating the Green's function against the source gives the response of the system to that source.

If we define the operator A to be

$$Ah(t) = \int_\Omega K_A(t,s) h(s) ds,$$

then A maps a source h to the response of the system. In a physically motivated scenario, this equation describes how a model A relates sources $h(t)$ to a response $u(t)$. The forward problem predicts the response for a given source, while the inverse problem recovers a source, given the response. To analyze Green's functions, we must

first work through some general results concerning compact operators.

2.2.2 Singular value expansion

When considering discrete inverse problems, the singular value decomposition is the tool of choice for rigorous analysis of the problem and its least squares solution. The continuous extension of this tool is the singular value expansion (SVE) [6, 9, 16, 17]. When originally developed by Smith, the singular value decomposition of a matrix was treated a special case of the more general singular value expansion. A history of the early development of the SVD/SVE can be found in [20]. Our study of general linear inverse problems will proceed with the singular value expansion as our primary tool. In particular, we will use the SVE for compact operators acting on Hilbert spaces. This is the analogous concept of a matrix acting on a finite dimensional vector space.

Definition 2.2.1. (Compact operator) Let H and H_A be Hilbert spaces and let $A : H \to H_A$ be a linear operator. We say A is compact if the image under A of a bounded subset of H is a precompact subset of H_A.

With the assumption of compactness, a significant amount of matrix theory can be extended to continuous operators. In particular, using the SVE we decompose an operator into orthogonal functions as we would decompose a matrix into orthogonal vectors in the SVD.

Note that we do not wish to consider the case where A has only finitely many singular values, nor the case where A is not compact.

Theorem 2.2.1 (Singular Value Expansion)*.* Let H, H_A be Hilbert spaces, and let $A : H \to H_A$ be a compact linear operator. Then there exists orthonormal sequences

$\{\phi_k\} \subset H$ and $\{\psi_k\} \subset H_A$ and positive numbers $\sigma_1 \geq \sigma_2 \geq \cdots$ converging to zero, such that

$$A = \sum_{k=1}^{\infty} \sigma_k \psi_k \otimes \phi_k, \quad \text{and} \quad A^* = \sum_{k=1}^{\infty} \sigma_k \phi_k \otimes \psi_k.$$

We define $\psi_k \otimes \phi_k$ as

$$(\psi_k \otimes \phi_k)h = \langle h, \phi_k \rangle_H \psi_k,$$

for all $h \in H$. Note that A^* is also a compact linear operator and denotes the adjoint of A. Furthermore,

$$A\phi_k = \sigma_k \psi_k \text{ for all } k$$

and

$$Ah = \sum_{k=1}^{\infty} \sigma_k \langle \phi_k, h \rangle_H \psi_k \text{ for all } h \in H.$$

Additionally, $\{\phi_k\}$ is a complete orthonormal set for $\mathcal{N}(A)^{\perp}$ and $\{\psi_k\}$ is a complete orthonormal set for $\overline{\mathcal{R}(A)}$.

Proof. See [6] or [7]. □

To derive the singular values analytically, we typically consider solving

$$A^* A \phi = \sigma^2 \phi.$$

This yields a family of singular function, singular value pairs $\{(\sigma_k, \phi_k)\}_{k=1}^{\infty}$. As we will see, this is not at all trivial. However, the singular values in the continuous setting give analogous insights as those in the discrete setting. In particular, the singular values of an operator can define a metric for ill-posedness. A goal of this thesis is to define and understand how this metric can quantify ill-posedness in the inversion of Green's function solutions of differential equations.

In section 2.1.2 we defined the discrete generalized inverse for a matrix A that operates on real, finite dimensional Hilbert spaces with $A : \mathbb{R}^n \to \mathbb{R}^m$. Written as a summation,

$$A^\dagger = \sum_{i=1}^{r} \sigma_i^{-1} v_i \otimes u_i = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}.$$

The solution that minimizes $\|Ax - b\|_2^2$ is given by

$$x_\dagger = A^\dagger b = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i.$$

Now suppose H and H_A are infinite dimensional Hilbert spaces, with $A : H \to H_A$ a compact linear operator having singular system $\{\sigma_i, \phi_i, \psi_i\}$ as defined in Theorem 2.2.1. The generalized inverse is expressed in an analogous manner,

$$A^\dagger = \sum_{k=1}^{\infty} \sigma_k^{-1} \phi_k \otimes \psi_k. \tag{2.2}$$

It is clear from this formulation that $A^\dagger$ is not a compact operator, since σ_k^{-1} increase in an unbounded manner [13, 16].

As with the discrete case, it will most likely be the case that we cannot solve $Ah = f$ exactly. we again find the least squares solution that minimizes $\|Ah - f\|_{H_A}^2$. In terms of the SVE, this gives,

$$h = A^\dagger f = \sum_{k=1}^{\infty} \sigma_k^{-1} (\phi_k \otimes \psi_k) f = \sum_{k=1}^{\infty} \frac{\langle \psi_k, f \rangle_{H_A}}{\sigma_k} \phi_k \text{ for all } f \in D\left(A^\dagger\right).$$

2.2.3 Conditioning and ill-posedness

The condition number of compact linear operators is still defined as σ_1/σ_r, however in infinite dimensions the operator can have an infinite sequence of singular values

decaying towards zero. The condition number is therefore not a sufficient metric by which to measure ill-posedness. As in the discrete case, it is clear that small singular values (relative to σ_1) will disproportionately amplify the contribution from corresponding singular vectors or functions. If there is noise in the data, this too will be amplified, perhaps to an unacceptable level.

We characterize the ill-posedness of the problem in terms of the decay rate of its singular values. In particular, if $\sigma_k(A)$ decays like k^{-q}, we call q the degree of ill-posedness. Thus larger values of q indicate larger degrees of ill-posedness. In Chapter 3, we will use this metric to determine how joint inversion improves single inversions. Let us consider the following example to illustrate this measure of ill-posedness.

Example

Define the compact linear operator $A : H \to H_A$, where $H = H_A = L^2(0, 1)$, by

$$Ah(t) = \int_0^t h(s)ds,$$

Then, the adjoint operator $A^* : H_A \to H$ is

$$A^*f(t) = \int_t^1 f(s)ds$$

and the self-adjoint operator $A^*A : H \to H$ is

$$A^*Ah(t) = \int_t^1 \left(\int_0^s h(\tau)d\tau \right) ds.$$

Now we solve $A^*A\phi = \sigma^2\phi$ for σ the singular values and ϕ the right-singular functions of A. If we let $\lambda = \sigma^2$, we may write the problem as

$$A^*A\phi = \lambda\phi(t) = \int_t^1 \left(\int_0^s \phi(\tau)d\tau \right) ds.$$

Notice, that substituting in $t = 1$ to both sides of the integral equation gives us $\phi(1) = 0$ for $\lambda \neq 0$. Differentiating both sides of the equation with respect to t and applying a Leibniz integral rule yields

$$\begin{aligned} \lambda\phi'(t) &= \frac{d}{dt} \int_t^1 \left(\int_0^s \phi(\tau)d\tau \right) ds \\ &= \int_t^1 \left(\frac{\partial}{\partial t} \int_0^s \phi(\tau)d\tau \right) ds - \int_0^t \phi(\tau)d\tau \\ &= - \int_0^t \phi(\tau)d\tau. \end{aligned}$$

Substituting in $t = 0$ to both sides of the above equation gives us $\phi'(0) = 0$ for $\lambda \neq 0$. Differentiating both sides of the equation again with respect to t yields

$$\begin{aligned} \lambda\phi''(t) &= -\frac{d}{dt} \int_0^t \phi(\tau)d\tau \\ &= -\phi(t). \end{aligned}$$

We can now observe that the original problem is equivalent to

$$\lambda\phi'' + \phi = 0, \quad \text{with} \quad \phi(1) = \phi'(0) = 0.$$

To find the solution of this boundary value problem, consider the roots of the auxiliary equation $\lambda r^2 + 1 = 0$. They are

$$r_1 = \frac{1i}{\sqrt{\lambda}}, \quad r_2 = \frac{-1i}{\sqrt{\lambda}}.$$

This gives us the general solution to the ODE and its derivative as

$$\phi(t) = c_1 \cos\left(\frac{t}{\sqrt{\lambda}}\right) + c_2 \sin\left(\frac{t}{\sqrt{\lambda}}\right),$$

$$\phi'(t) = -\frac{c_1}{\sqrt{\lambda}} \sin\left(\frac{t}{\sqrt{\lambda}}\right) + \frac{c_2}{\sqrt{\lambda}} \cos\left(\frac{t}{\sqrt{\lambda}}\right).$$

Using the boundary conditions $\phi(1) = \phi'(0) = 0$, we deduce that

$$c_2 = 0, \text{ and}$$

$$\lambda_k = \frac{4}{(2k-1)^2 \pi^2}.$$

Thus the singular functions are of the form

$$\phi_k(t) = c_1 \cos \frac{2k-1}{2} \pi t,$$

where c_1 must be chosen to satisfy the orthonormality condition $\langle \phi_k, \phi_k \rangle_H = 1$, for $k = 1, 2, \ldots, \infty$. This yields a family of functions,

$$\phi_k(t) = \sqrt{2} \cos \frac{2k-1}{2} \pi t, \ k \in \mathbb{N}, \text{ and } \lambda_k = \frac{4}{(2k-1)^2 \pi^2}, \ k \in \mathbb{N}.$$

The singular values of the compact operator A are the square roots of λ, that is

$$\sigma_k = \frac{2}{(2k-1)\pi}, \ k \in \mathbb{N}.$$

We can see from Figure 2.1 that the singular values of the operator A decay like k^{-1}. This is classified as a mildly ill-posed problem [1]. Alternatively, the problem would be severely ill-posed if the singular values decayed at an exponential rate. The
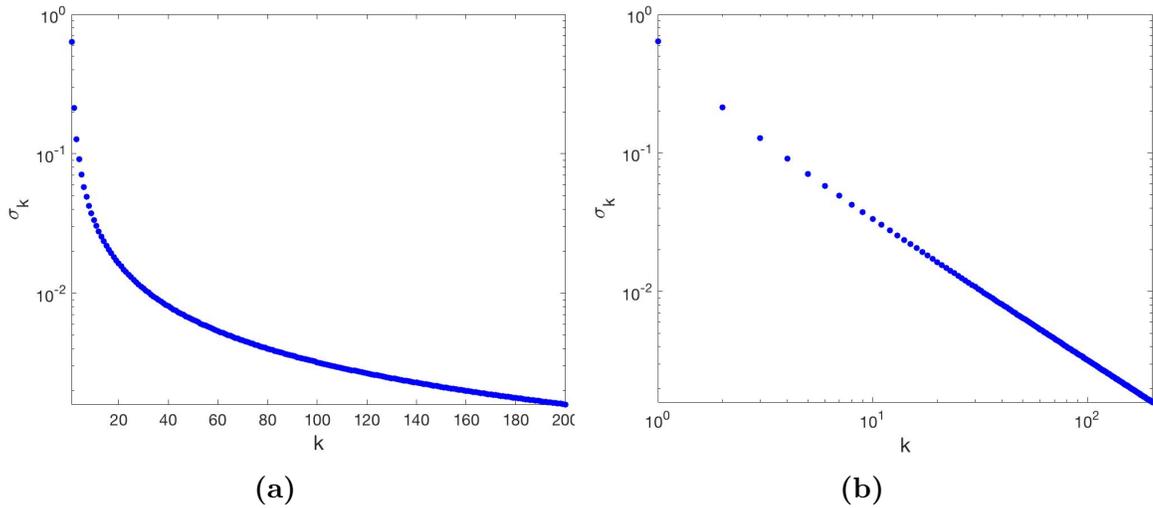
Figure 2.1: Singular values of A on a semi-log scale (a), and a log-log scale (b)

diffusion equation is an example of a severely ill-posed problem.

2.2.4 Regularization

The negative effect decaying singular values have on the parameter estimates in an ill-posed problem can be alleviated with regularization. In infinite dimensional Hilbert spaces, a truncated SVE approximation to the operator A can be formed analogously to the truncated SVD of a matrix. This requires the truncation of infinitely many singular values, and we will not investigate this finite sum approximation. Alternatively, we focus on Tikhonov regularization for compact operators.

Tikhonov regularization

The operators of interest can be expressed as an infinite summation and often do not have a well-defined inverse. Similar to the discrete case, Tikhonov regularization changes the problem to one which has an invertible operator, and therefore well-defined inverse solution. The definition of this new, invertible operator will require us

to consider the space $H_A \times H = \{(h_A, h) : h_A \in H_A, h \in H\}$ which is a Hilbert space under the inner product

$$\langle (h_{A,1}, h_1), (h_{A,2}, h_2) \rangle_{H_A \times H} = \langle h_{A,1}, h_{A,2} \rangle_{H_A} + \langle h_1, h_2 \rangle_H.$$

Now we may define a new operator $T_\lambda : H \to H_A \times H$ by

$$T_\lambda h = (Ah, \sqrt{\lambda}h).$$

In place of $\|Ah - f\|^2_{H_A}$, we instead consider minimizing

$$\|T_\lambda h - (f, 0)\|^2_{H_A \times H} = \|Ah - f\|^2_{H_A} + \lambda \|h\|^2_H.$$

Note that λ is being used as a regularization parameter, and not an eigenvalue.

Theorem 2.2.2. Suppose $\lambda > 0$. Then $\mathcal{R}(T_\lambda)$ is closed and $\mathcal{N}(T_\lambda)$ is trivial. Therefore $T_\lambda h = (f, 0)$ has a unique least squares solution for all $f \in H_A$, [6].

Proof. Consider the normal equation for this problem:

$$T^*Th = T^*(f, 0)$$

$$T^*(Ah, \sqrt{\lambda}h) = T^*(f, 0)$$

$$A^*Ah + \lambda h = A^*f + \sqrt{\lambda} \cdot 0$$

$$(A^*A + \lambda I)h = A^*f.$$

It is straightforward to show that $(A^*A + \lambda I)$ is invertible with a bounded inverse. Therefore a unique solution to the normal equations exists. □

Tikhonov regularization replaces the not necessarily invertible operator A^*A with the necessarily invertible $(A^*A+\lambda I)$ in the normal equations. A solution is guaranteed for appropriate λ, however we have restricted the space of acceptable solutions.

We now examine the regularized solutions expressed in terms of the singular value expansion of A. Given the singular value expansion of A and its adjoint A^*, A^*A applied to h can be expressed as

$$A^*Ah = A^*\left(\sum_{k=1}^{\infty}\sigma_k\langle\phi_k,h\rangle_H\psi_k\right)$$
$$= \sum_{k=1}^{\infty}\sigma_k\langle\phi_k,h\rangle_H A^*\psi_k$$
$$= \sum_{k=1}^{\infty}\sigma_k^2\langle\phi_k,h\rangle_H\phi_k.$$

Equivalently,

$$A^*A = \sum_{k=1}^{\infty}\sigma_k^2\phi_k\otimes\phi_k.$$

Recall that $\{\phi_k\}$ is an orthonormal set for the $\mathcal{N}(A)^{\perp}$. So, $h\in H$ can be expressed as

$$h = \mathrm{proj}_{\mathcal{N}(A)^{\perp}}h + \mathrm{proj}_{\mathcal{N}(A)}h$$
$$= \sum_{k=1}^{\infty}\langle h,\phi_k\rangle_H\phi_k + \mathrm{proj}_{\mathcal{N}(A)}h$$

We may now consider writing the Tikhonov operator $(A^*A + \lambda I)$ applied to h in terms of the SVE of A. Namely,

$$(A^*A + \lambda I)h = A^*A\left(\sum_{k=1}^{\infty}\langle h, \phi_k\rangle_H\phi_k + \text{proj}_{\mathcal{N}(A)}h\right) + \lambda\left(\sum_{k=1}^{\infty}\langle h, \phi_k\rangle_H\phi_k + \text{proj}_{\mathcal{N}(A)}h\right)$$

$$= A^*A\left(\sum_{k=1}^{\infty}\langle h, \phi_k\rangle_H\phi_k\right) + \lambda\sum_{k=1}^{\infty}\langle h, \phi_k\rangle_H\phi_k + \lambda\text{proj}_{\mathcal{N}(A)}h$$

$$= \sum_{k=1}^{\infty}\langle h, \phi_k\rangle_H A^*A\phi_k + \sum_{k=1}^{\infty}\lambda\langle h, \phi_k\rangle_H\phi_k + \lambda\text{proj}_{\mathcal{N}(A)}h$$

$$= \sum_{k=1}^{\infty}\sigma_k^2\langle h, \phi_k\rangle_H\phi_k + \sum_{k=1}^{\infty}\lambda\langle h, \phi_k\rangle_H\phi_k + \lambda\text{proj}_{\mathcal{N}(A)}h$$

$$= \sum_{k=1}^{\infty}\left(\sigma_k^2 + \lambda\right)\langle h, \phi_k\rangle_H\phi_k + \lambda\text{proj}_{\mathcal{N}(A)}h.$$

Equivalently,

$$(A^*A + \lambda I) = \sum_{k=1}^{\infty}\left(\sigma_k^2 + \lambda\right)\phi_k \otimes \phi_k + \lambda\text{proj}_{\mathcal{N}(A)}.$$

This allows us to consider solutions $h^{(\lambda)} \in H$ satisfying $(A^*A + \lambda I)h^{(\lambda)} = A^*f$ in terms of the SVE of A. From our above expansions, we can see that $h^{(\lambda)}$ must equivalently satisfy

$$\sum_{k=1}^{\infty}\left(\sigma_k^2 + \lambda\right)\langle h^{(\lambda)}, \phi_k\rangle_H\phi_k + \lambda\text{proj}_{\mathcal{N}(A)}h^{(\lambda)} = \sum_{k=1}^{\infty}\sigma_k\langle \psi_k, f\rangle_{H_A}\phi_k.$$

Since the two summations are linear combinations of $\{\phi_k\}$, they represent elements in $\mathcal{N}(A)^\perp$. Therefore, it must be the case that $\lambda\text{proj}_{\mathcal{N}(A)}h^{(\lambda)} = 0$, and $h^{(\lambda)}$ lies in $\mathcal{N}(A)^\perp$. So we have that $h^{(\lambda)}$ must satisfy

$$\sum_{k=1}^{\infty}\left(\sigma_k^2 + \lambda\right)\langle h^{(\lambda)}, \phi_k\rangle_H\phi_k = \sum_{k=1}^{\infty}\sigma_k\langle \psi_k, f\rangle_{H_A}\phi_k$$

$$\implies \left(\sigma_k^2 + \lambda\right) \langle h^{(\lambda)}, \phi_k \rangle_H = \sigma_k \langle \psi_k, f \rangle_{H_A} \text{ for all } k \in \mathbb{N}$$

$$\implies \langle h^{(\lambda)}, \phi_k \rangle_H = \frac{\sigma_k}{\sigma_k^2 + \lambda} \langle \psi_k, f \rangle_{H_A} \text{ for all } k \in \mathbb{N}$$

$$\implies h^{(\lambda)} = \sum_{k=1}^{\infty} \frac{\sigma_k}{\sigma_k^2 + \lambda} \langle \psi_k, f \rangle_{H_A} \phi_k.$$

This implies that the generalized inverse operator for the modified least squares problem is

$$A_\lambda^\dagger = (A^*A + \lambda I)^{-1} A^* = \sum_{k=1}^{\infty} \frac{\sigma_k}{\sigma_k^2 + \lambda} \phi_k \otimes \psi_k.$$

Notice that

$$\frac{\sigma_k}{\sigma_k^2 + \lambda} \to 0, \text{ as } k \to \infty.$$

Therefore the operator $A_\lambda^\dagger$ is bounded, and inverse solutions depend continuously on f.

These solution estimates depend strongly on the regularization parameter λ, which restricts the space of acceptable solutions. In this thesis we study the impact of using physics and corresponding observations to restrict the solution space. This will allow more physically relevant solutions and is typically called simultaneous joint inversion. Joint inversion restricts the parameters to ones that satisfy two or more models. Joint inversion may not necessarily result in a well posed problem, but should require less regularization.

CHAPTER 3

JOINT INVERSION

3.1 Discrete joint inversion

3.1.1 Linear problems

The earliest mention of joint inversion comes from Vozoff and Jupp in their 1975 paper [22]. Here the authors considered combining several different kinds of geophysical measurements to avoid some of the ambiguity inherent in the individual methods. Mathematically speaking, they looked at a problem of two linear models A and B dependent on the same set of unknown parameters x, with corresponding data d_1 and d_2 respectively:

$$Ax = d_1 \text{ and } Bx = d_2.$$

When A and B are ill-conditioned, the individual problems require regularization. Alternatively, Vozoff and Jupp explored the potential of using less regularization by solving the joint inversion problem

$$\begin{bmatrix} A \\ B \end{bmatrix} x = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}.$$

To measure the effectiveness of this approach, the authors considered the decay rate of the singular values of A, B, and $\begin{bmatrix} A \\ B \end{bmatrix}$. In several examples, they demonstrated that the singular value decomposition of $\begin{bmatrix} A \\ B \end{bmatrix}$ yielded a significant increase in the number of usable values, as compared to A and B individually.

In this thesis we extend the discrete results of Vozoff and Jupp to a continuous setting. In particular, we measure the effectiveness of combining operators in an analogous way; the decay rate of singular values from individual and joint compact operators will be compared.

3.1.2 Green's functions example

First, we apply the ideas presented by Vozoff and Jupp to a pair of discretized Green's functions. As motivated by the imaging example in the introduction, we choose the Green's functions for the wave and diffusion equations in time and one spatial variable.

The Green's function for the wave and diffusion equations for homogeneous media in x and t are

$$
\begin{aligned}
K_w &= \frac{1}{2}\mathcal{H}((t-\tau) + (x-\xi))\mathcal{H}((t-\tau) - (x-\xi)), \\
K_d &= \frac{\mathcal{H}(t-\tau)}{\sqrt{4\pi^2(t-\tau)}} \exp\left(\frac{-(x-\xi)^2}{4(t-\tau)}\right)
\end{aligned}
$$

respectively, where $\mathcal{H}$ denotes the Heaviside function [3].

Given each kernel K, a solution u to the corresponding differential equation with source term f can be expressed as

$$u(x,t) = \int \int K(x,t,\xi,\tau) f(\xi,\tau) d\xi d\tau.$$

Using a two dimensional midpoint quadrature rule, we can approximate our solution on an $m \times n$ grid as

$$u(x_i, t_j) \approx \sum_{l=1}^{m} \sum_{k=1}^{n} K(x_i, t_j, \xi_k, \tau_l) f(\xi_k, \tau_l) \Delta\xi \Delta\tau.$$

With some simple reshaping, we approximate the operation of double integration against K as a matrix multiplication so that we solve $Af = u$ for the source function. The resulting singular values of the matrices from the discretized individual and joint operators are plotted in Figure 3.1. The wave equation in (a) had singular values decay slowly until the ninetieth. Then the singular values rapidly approached zero. The diffusion equation in (b) had singular values that decayed more rapidly than the wave equation, exhibiting noticeable drops every ten values past the forty-fifth, eventually dropping to machine precision. The joint operator in (c) had singular values that tapered off far less dramatically than either the wave or diffusion equation. It is significantly better conditioned than either individual model.

It must be noted that even though this simple discretization resulted in improved decay rates of singular values in joint inversion this will not always be the case. Having run many numerical experiments, it is still unclear what underlying features cause discrete models to interact well in a joint inversion. For this reason, we explore the more general notion of joint inversion on compact operators. In so doing, we more deeply understand what properties in physical models improve inverse estimates in ill-posed problems.
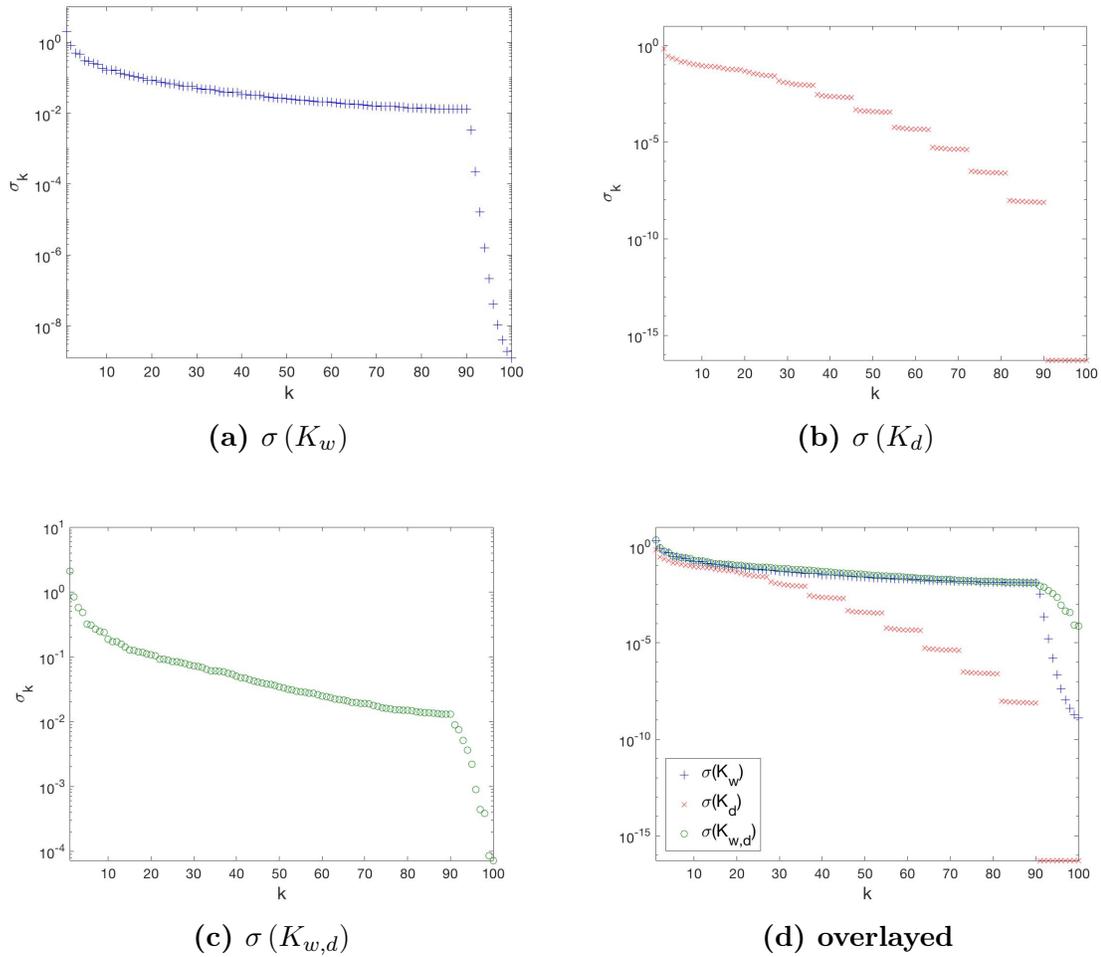
Figure 3.1: Singular values of individual operators and joint operator

3.2 Joint inversion with compact operators

Vozoff and Jupp vertically concatenated two discrete linear models into one stacked matrix. We now define a joint operator that combines two compact operators in an analogous way. The decay rate of the singular values of this joint operator compared to the decay rate of the singular values of the individual operators will give an indication of the effectiveness of joint inversion.

3.2.1   Joint operators

As with Tikhonov regularization, our joint operator will map into the Cartesian product of two Hilbert spaces. However, rather than consider the mathematically defined space H in Tikhonov regularization, we introduce the new physical space H_B defined by an additional data collection technique.

Definition 3.2.1. (Hilbert space direct sum) Let H_A and H_B be Hilbert spaces and

$$H_A \oplus H_B = \{(h_A, h_B) : h_A \in H_A, h_B \in H_B\}.$$

Define an inner product $\langle \cdot, \cdot \rangle$ on $H_A \oplus H_B$ by

$$\langle (h_{A,1}, h_{B,1}), (h_{A,2}, h_{B,2}) \rangle_{H_A \oplus H_B} = \langle h_{A,1}, h_{A,2} \rangle_{H_A} + \langle h_{B,1}, h_{B,2} \rangle_{H_B}.$$

With respect to this inner product, $H_A \oplus H_B$ is a Hilbert Space called the Hilbert space direct sum of H_A and H_B.

Remark. Let $A : H \to H_A$ and $B : H \to H_B$ be compact operators from the Hilbert space H to the Hilbert spaces H_A and H_B respectively. Define $C : H \to H_A \oplus H_B$ as $Ch = (Ah, Bh)$, for all $h \in H$. Then C is a compact operator between two Hilbert spaces, [2]. Therefore, C admits a singular value expansion.

As mentioned previously, the decay rate of the singular values provide a metric for the ill-posedness of an operator. With $A, B,$ and C defined as above, we will compare the singular value decay rates of the three operators to see if the joint operator yielded any improvement. For the purposes of visualization, it is helpful to think of C as defining a parametric curve in the space $H_A \oplus H_B$. Then it is clear that if A and B are compact, so is C.

Example

Define the Hilbert spaces $H = L^2(0, 2\pi)$, and $H_A = H_B = \mathbb{R}$. Define the compact operators $A : H \to H_A$ and $B : H \to H_B$ as

$$Ah = \int_0^{2\pi} h(y)\delta(y-5)dy, \quad Bh = \int_0^{2\pi} h(y)\delta(y-7)dy.$$

Define $C : H \to H_A \oplus H_B$ as

$$Ch = (Ah, Bh) = \left(\int_0^{2\pi} h(y)\delta(y-5)dy, \int_0^{2\pi} h(y)\delta(y-7)dy \right).$$

Now let us consider the image of a subset of H under C. Define $S = \{\cos kx : k \in \mathbb{R}, x \in [0, 2\pi]\}$, a continuum of cosine functions of each possible frequency. So for each frequency $k \in \mathbb{R}$ the mapping of $\cos kx$ under C is

$$C(\cos kx) = \left(\int_0^{2\pi} \cos(ky)\delta(y-5)dy, \int_0^{2\pi} \cos(ky)\delta(y-7)dy \right) \tag{3.1}$$

$$= (\cos(k \cdot 5), \cos(k \cdot 7)). \tag{3.2}$$

Since the codomain of C is $H_A \oplus H_B = \mathbb{R}^2$, we can represent the image of S under C graphically; see Figure 3.2.

3.2.2 Singular values

The singular values of the joint operator are found by solving $C^*C\phi = \sigma^2\phi$. With C defined as in the remark above, $h_A \in H_A$ and $h_B \in H_B$, the adjoint $C^* : H_A \oplus H_B \to H$ is

$$C^*(h_A, h_B) = A^*h_A + B^*h_B.$$
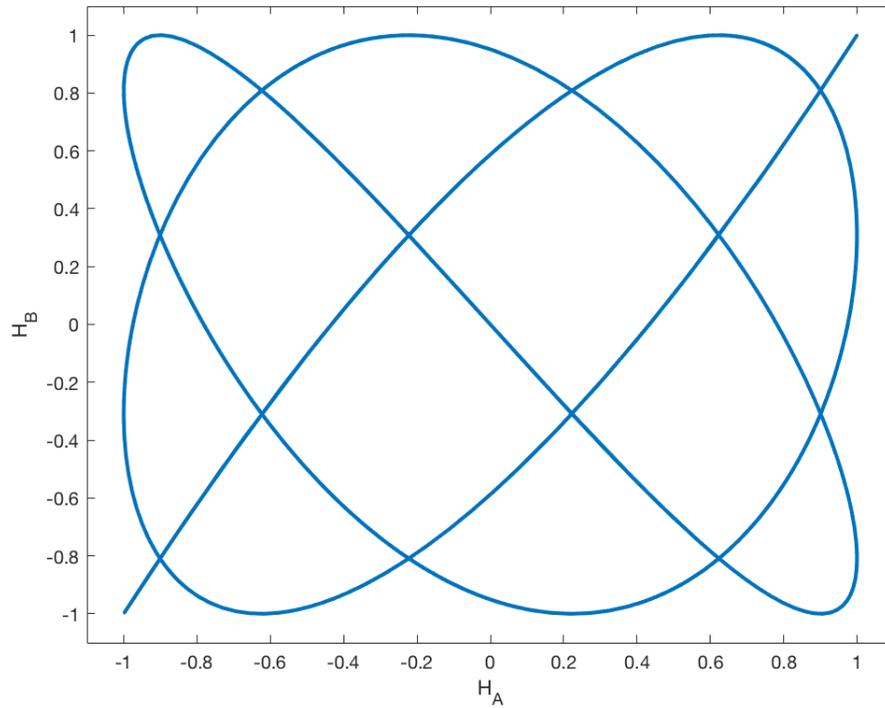
Figure 3.2: Parametric curve defined by $C(\cos kx)$ **for** $k \in [-5, 5]$

Expanding we get

$$\sigma^2 \phi = C^* C \phi$$

$$= C^* \left(A\phi, B\phi \right)$$

$$= A^* A\phi + B^* B\phi. \tag{3.3}$$

It is a challenge to solve this equation for the singular values and functions analytically. In the case where A and B are Green's function operators associated with the differential operations L_A and L_B respectively, we can see that this is an integral equation. For our motivating imaging problem,

$$Ah = \int_\Omega K_w h \quad \text{and,} \quad Bh = \int_\Omega K_d h,$$

where $h \in H$, and K_w, K_d are the Green's functions solutions to the wave and diffusion equations respectively. Since we are better equipped to solve differential equations, we take steps to transform equation (3.3) into an equivalent ODE. The following properties are what allows us to do so.

Remark. Let $\mathcal{L}$ be a linear differential operator, and K be the corresponding Green's function which satisfies $\mathcal{L}K(t,s) = \delta(t-s)$. Let A be defined as $Ah = \int_\Omega K(t,s)h(s)ds$ for all h in the relevant Hilbert space H. Then we have

$$\int_\Omega \mathcal{L}K(t,s)h(s)ds = \int_\Omega \delta(t-s)h(s)ds$$
$$\mathcal{L}\int_\Omega K(t,s)h(s)ds = h(t)$$
$$\mathcal{L}Ah(t) = h(t), \tag{3.4}$$

for all $h \in H$. Similarly we have,

$$\mathcal{L}^* A^* h_A(t) = h_A(t). \tag{3.5}$$

for all h_A in the relevant codomain H_A.

We begin the process of transforming equation (3.3) by applying the adjoint differential operator, L_A^* to both sides and simplifying using the property in equation (3.5):

$$L_A^* \left(\sigma^2 \phi \right) = L_A^* \left(A^* A \phi + B^* B \phi \right),$$

$$\sigma^2 L_A^* \phi = A \phi + L_A^* B^* B \phi.$$

Now we apply the differential operator L_A to both sides and simplify using the property in equation (3.4):

$$L_A \left(\sigma^2 L_A^* \phi \right) = L_A \left(A \phi + L_A^* B^* B \phi \right),$$

$$\sigma^2 L_A L_A^* \phi = \phi + L_A L_A^* B^* B \phi.$$

This eliminated the integrals associated with the operator A^* and A. We must also apply L_B^* and L_B in the same way to eliminate the integrals associated with B^* and B.

Certain linear differential operators commute, for instance, linear differential operators in one dimension, with constant coefficients. Below we use the assumption of commutativity so that we can define an ODE from which we can find the singular values of the joint operator. It is not a realistic assumption for the application that motivates this work. However, we view this as preliminary work that can be extended to more realistic settings.

Assuming $L_A L_A^*$ and $L_B L_B^*$ commute and applying $L_B L_B^*$ to both sides of the equation, we find

$$L_B L_B^* \left(\sigma^2 L_A L_A^* \phi \right) = L_B L_B^* \left(\phi + L_A L_A^* B^* B \phi \right),$$

$$\sigma^2 \left(L_B L_B^* L_A L_A^* \right) \phi = L_B L_B^* \phi + L_A L_A^* L_B L_B^* B^* B \phi,$$

$$\sigma^2 \left(L_B L_B^* L_A L_A^* \right) \phi = L_B L_B^* \phi + L_A L_A^* \phi,$$

$$\sigma^2 \left(L_B L_B^* L_A L_A^* \right) \phi = \left(L_B L_B^* + L_A L_A^* \right) \phi.$$

Though not necessarily aesthetically pleasing, this is clearly an ODE in ϕ. Namely,

$$\sigma^2 \left(L_B L_B^* L_A L_A^* \right) \phi = \left(L_B L_B^* + L_A L_A^* \right) \phi. \tag{3.6}$$

This produced an ODE with much higher order than that of the given differential operators L_A or L_B, and introduced many more boundary conditions. In cases where equation (3.6) is too difficult to solve, we turn to a numerical approximation as described in Section 3.2.3.

3.2.3 Approximation of SVE with SVD

The singular value expansion of an individual integral kernel, such as a Green's function, can be approximated using the Galerkin method. It has been shown that the singular values derived using the Galerkin method converge to the true singular values. For further details concerning the results pertaining to individual operators, see [9, 18]. Here, we describe the algorithm for computing the SVE approximation and its convergence properties. We then extend the methods to joint operators, and prove the corresponding convergence properties.

Given a Green's function or other integral kernel, $K_A(s,t)$ defined over $\Omega_s \times \Omega_t$ and associated integral operator A, we choose orthonormal bases $\{p_j(t)\}_{j=1}^n$ and $\{q_i(s)\}_{i=1}^n$ for $L^2(\Omega_t)$ and $L^2(\Omega_s)$ respectively. The matrix $A^{(n)}$ with entries $a_{ij}^{(n)}$ approximates

the operator A, and is defined by

$$
\begin{aligned}
a_{ij}^{(n)} &= \langle q_i, A p_j \rangle \\
&= \langle q_i, \langle K_A, p_j \rangle \rangle \\
&= \int_{\Omega_s} \int_{\Omega_t} q_i(s) K_A(s,t) p_j(t) dt ds.
\end{aligned}
\tag{3.7}
$$

The SVD $A^{(n)}$ is denoted $U^{(n)} \Sigma^{(n)} \left(V^{(n)} \right)^T$ with $\Sigma^{(n)} = \operatorname{diag}\left(\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots \sigma_n^{(n)} \right)$ containing the discrete singular values $\sigma_i^{(n)}$ which approximate the continuous singular values σ_i.

Theorem 3.2.1. (SVE-SVD, [9, 18]) Define

$$
\begin{aligned}
\left(\Delta_A^{(n)} \right)^2 &= \| K_A \|^2 - \| A^{(n)} \|_F^2 \\
&= \sum_{i=1}^{\infty} \sigma_i^2 - \sum_{i=1}^{n} (\sigma_i^{(n)})^2.
\end{aligned}
$$

Then the following hold for all i and n, independent of the convergence of $\Delta_A^{(n)}$ to 0:

1. $\sigma_i^{(n)} \leq \sigma_i^{(n+1)} \leq \sigma_i$

2. $0 \leq \sigma_i - \sigma_i^{(n)} \leq \Delta_A^{(n)}$

Thus if $\lim_{n \to \infty} (\Delta_A^{(n)})^2 = 0$, the singular values σ_i of K_A are accurately approximated.

We extend this method to joint operators in the following fashion. Let $K_A(s,t)$ and $K_B(s,t)$ be Green's functions or other integral kernels, over $\Omega_s \times \Omega_t$ with associated integral operator A and B respectively, and $K_A \oplus K_B$ is the kernel of the direct sum integral operator $C = A \oplus B$, [8]. Compute $A^{(n)}$ and $B^{(n)}$ separately as described in equation (3.7). Now that we are in a discrete setting we can stack these matrices

found to form $C^{(n)} = \begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix}$. We now show that the singular values of this matrix $\sigma_i(C^{(n)})$ converge to the singular values $\sigma_i(C)$ of the operator C. We will begin by using an equivalent definition for the singular values and a Lemma which can be found in [9].

Definition 3.2.2. The singular values of an integral operator A with a real, square integrable kernel K_A are the stationary values of the functional

$$F[\phi, \psi] = \frac{\langle \psi, A\phi \rangle}{\|\phi\| \|\psi\|},$$

with the corresponding left and right singular functions given by $\phi/\|\phi\|$ and $\psi/\|\psi\|$ respectively.

The idea is then to approximate A with an integral operator whose kernel is degenerate. We accomplish this by restricting ϕ and ψ to a the span of finitely many, n, orthonormal basis functions.

Theorem 3.2.2 ([9]). The approximate singular values $\sigma_i(C^{(n)})$, where n is the number of basis functions, are increasingly better approximations to the true singular values $\sigma_i(C)$,

$$\sigma_i(C^{(n)}) \le \sigma_i(C^{(n+1)}) \le \sigma_i(C), \quad i = 1, 2, \ldots n.$$

Proof. The theorem follows from the facts that the basis functions $\{\phi_i\}_{i=1}^{n}$ and $\{\psi_i\}_{i=1}^{n}$ are orthonormal and that the singular values $\sigma_i(C^{(n)})$ and $\sigma_i(C^{(n+1)})$ are the stationary values of $F[\phi, \psi]$ with ϕ and ψ restricted to n-dimensional and $n+1$-dimensional function subspaces respectively. $\square$

Now we consider the error of joint approximation.

Theorem 3.2.3. Define

$$\left(\Delta_C^{(n)}\right)^2 = \|K_A \oplus K_B\|^2 - \left\|\left[\begin{matrix} A^{(n)} \\ B^{(n)} \end{matrix}\right]\right\|_F^2 \tag{3.8}$$

$$= \sum_{i=1}^{\infty} \sigma_i(C)^2 - \sum_{i=1}^{n} \sigma_i(C^{(n)})^2.$$

$\left(\Delta_C^{(n)}\right)^2$ is equivalent to $\left(\Delta_A^{(n)}\right)^2 + \left(\Delta_B^{(n)}\right)^2$, the sum of the errors from the discretization of A and B.

Proof. Expanding equation (3.8),

$$\left(\Delta_C^{(n)}\right)^2 = \langle K_A \oplus K_B, K_A \oplus K_B \rangle - \left[\|A^{(n)}\|_F^2 + \|B^{(n)}\|_F^2\right]$$

$$= \langle K_A, K_A \rangle + \langle K_B, K_B \rangle - \|A^{(n)}\|_F^2 - \|B^{(n)}\|_F^2$$

$$= \|K_A\|^2 - \|A^{(n)}\|_F^2 + \|K_B\|^2 - \|B^{(n)}\|_F^2$$

$$= \left(\Delta_A^{(n)}\right)^2 + \left(\Delta_B^{(n)}\right)^2. \tag{3.9}$$

Thus the joint error is the sum of the individual errors. □

This quantity is important in bounding the error.

Theorem 3.2.4. The sum of squares error of the approximate joint singular values is bounded by

$$\sum_{i=1}^{n} \left[\sigma_i(C) - \sigma_i(C^{(n)})\right]^2 \leq \left(\Delta_C^{(n)}\right)^2.$$

Proof.

$$\sum_{i=1}^{n} \left[\sigma_i(C) - \sigma_i(C^{(n)})\right]^2 = \sum_{i=1}^{n} [\sigma_i(C)]^2 + \sum_{i=1}^{n} \left[\sigma_i(C^{(n)})\right]^2 - 2\sum_{i=1}^{n} \sigma_i(C)\sigma_i(C^{(n)})$$

$$\leq \sum_{i=1}^{n} [\sigma_i(C)]^2 + \sum_{i=1}^{n} \left[\sigma_i(C^{(n)})\right]^2 - 2\sum_{i=1}^{n} \left[\sigma_i(C^{(n)})\right]^2$$

$$\leq \sum_{i=1}^{n} [\sigma_i(C)]^2 + \sum_{i=1}^{n} \left[\sigma_i(C^{(n)})\right]^2$$

$$= \left(\Delta_C^{(n)}\right)^2$$

$$= \left(\Delta_A^{(n)}\right)^2 + \left(\Delta_B^{(n)}\right)^2$$

$\square$

Thus, if $\lim_{n\to\infty}(\Delta_A^{(n)})^2 = 0$ and $\lim_{n\to\infty}(\Delta_B^{(n)})^2 = 0$, the singular values $\sigma(C)$ are accurately approximated by $\sigma(C^{(n)})$.

CHAPTER 4

APPLICATION WITH COMPACT OPERATORS

In this chapter, we give an example of our approach to determining the effectiveness of joint inversion by calculating the decay rate of singular values. Models based on the wave and the diffusion equations give the motivation for this work, but determining the corresponding singular values is outside the scope of this thesis. Therefore we apply our approach to the one dimensional steady state versions of the models.

4.1 Individual operators

The Green's functions for the two differential operators

$$L_A u = -u'', \quad u(0) = u(\pi) = 0,$$

$$L_B u = u'' + b^2 u, \quad u(0) = u(\pi) = 0, \quad \text{and } b \notin \mathbb{Z}$$

are well known, [3]. Let $A : L^2[0, \pi] \to L^2[0, \pi]$ be defined by

$$Ah(x) = \int_0^\pi K_A(x, y) h(y) dy,$$

where the kernel

$$K_A = \begin{cases} \frac{1}{\pi}(\pi - x)\,y, & 0 \le y \le x \le \pi, \\ \\ \frac{1}{\pi}(\pi - y)\,x, & 0 \le x \le y \le \pi. \end{cases}$$

Then $Af = u$ where $-u'' = f$ with zero Dirichlet boundary conditions.

Similarly, let $B : L^2[0, \pi] \to L^2[0, \pi]$ be defined by

$$Bh(x) = \int_0^\pi K_B(x, y)\, h(y) dy,$$

where the kernel

$$K_B = \begin{cases} -\frac{\sin(by)\sin[b(\pi - x)]}{b\sin(b\pi)}, & 0 \le y \le x \le \pi, \\ \\ -\frac{\sin(bx)\sin[b(\pi - y)]}{b\sin(b\pi)}, & 0 \le x \le y \le \pi. \end{cases}$$

Then $Bf = u$, where $u'' + b^2 u = f$ with zero Dirichlet boundary conditions. Thus A and B are the Green's function operators associated with L_A and L_B, respectively.

A is a self-adjoint compact operator, and therefore admits an eigenvalue expansion. In this case, the singular values of A are simply the absolute values of the eigenvalues. The equation $A\phi = \lambda\phi$ is equivalent to

$$\lambda\phi(x) = \int_0^\pi K_A(x, y)\, \phi(y) dy.$$

First notice that when evaluated at $x = 0$ or $x = \pi$, the integral of the kernel K_A is zero. This implies the conditions $\phi(0) = 0$ and $\phi(\pi) = 0$. Now, differentiating both sides of the integral equation with respect to x and applying the Leibniz integral rule yields

$$
\begin{aligned}
\lambda \phi'(x) &= \frac{d}{dx} \int_0^\pi K_x(x,y)\phi(y)dy \\
&= \int_0^\pi \frac{\partial}{\partial x} \left[K(x,y)\phi(y) \right] dy \\
&= \int_0^x \frac{\partial}{\partial x} \left[K(x,y) \right] \phi(y)dy + \int_x^\pi \frac{\partial}{\partial x} \left[K(x,y) \right] \phi(y)dy \\
&= \int_0^x -\frac{1}{\pi} y\phi(y)dy + \int_x^\pi \frac{1}{\pi}(\pi - y)\phi(y)dy.
\end{aligned}
$$

Differentiating both sides with respect to x and applying the Leibniz integral rule again gives

$$
\begin{aligned}
\lambda \phi''(x) &= \frac{d}{dx} \left(\int_0^x -\frac{1}{\pi}y\phi(y)dy + \int_x^\pi \frac{1}{\pi}(\pi - y)\phi(y)dy \right) \\
&= -\frac{1}{\pi}x\phi(x) - \frac{1}{\pi}(\pi - x)\phi(x) \\
&= -\phi(x).
\end{aligned}
$$

We may now observe that the original eigenvalue problem is equivalent to the ODE

$$
\lambda \phi'' = -\phi \text{ with } \phi(0) = \phi(\pi) = 0.
$$

Solutions to this boundary value problem are of the form

$$
\phi(x) = c_1 \cos\left(\frac{1}{\sqrt{\lambda}}x \right) + c_2 \sin\left(\frac{1}{\sqrt{\lambda}}x \right).
$$

Using the boundary conditions, we deduce that

$$c_1 = 0, \quad \text{and}$$

$$\lambda_k = \frac{1}{k^2}, \quad \text{for } k = 1, 2, \ldots, \infty.$$

Thus solutions are of the form

$$\phi_k(x) = c_2 \sin(kx),$$

where c_2 must be chosen to satisfy the condition $\langle \phi_k, \phi_k \rangle_{L^2[0,\pi]} = 1$, for $k = 1, 2, \ldots, \infty$. This yields a family of solutions,

$$\phi_k(x) = \sqrt{\frac{2}{\pi}} \sin(kx), \quad \text{and} \quad \lambda_k = \frac{1}{k^2}, \quad \text{for } k = 1, 2, \ldots, \infty.$$

Since $\lambda_k > 0$, the singular values for A are also $\sigma_k = \frac{1}{k^2}$ for $k = 1, 2, \ldots, \infty$.

B is also a self-adjoint operator, so we need only compute its eigenvalues. Note that we will reuse the ϕ, λ notation, but the solutions obtained here are not related to A. We follow an approach similar to that for the computation of the singular values of A, and we exploit the relationship between B and L_B.

Applying L_B to both sides, we find

$$L_B B \phi = L_B(\lambda \phi)$$

$$\phi = \lambda L_B \phi$$

$$\implies \phi = \lambda \left(\phi'' + b^2 \phi \right), \quad \phi(0) = \phi(\pi) = 0.$$

So we are left with the following equations

$$\left(\frac{1}{\lambda} - b^2\right)\phi = \phi'', \quad \phi(0) = \phi(\pi) = 0.$$

Solutions to this boundary value problem are of the form

$$\phi(x) = c_1 \cos\left(\sqrt{\frac{1}{\lambda} - b^2}x\right) + c_2 \sin\left(\sqrt{\frac{1}{\lambda} - b^2}x\right).$$

Using the boundary conditions, we deduce that

$$c_1 = 0, \quad \text{and}$$

$$\lambda_k = \frac{1}{k^2 + b^2}, \quad \text{for } k = 0, 1, \ldots, \infty$$

Again, since $\lambda_k > 0$, the singular values are $\sigma_k = \frac{1}{k^2+b^2}$. We omit finding the eigenfunctions since the decay rates of the singular values is the focus of this work.

The first 200 singular values for A and B are plotted in Figure 4.1 on a semi-log scale. There we see how the early values decay quadratically. This decay continues on for the infinitely many singular values approaching zero.

4.2   Joint operator

The joint operator C will not be self adjoint since it does not have the same domain and codomain. i.e. $C : H \to H_A \oplus H_B$. Therefore we must compute the singular values, as opposed to the eigenvalues as done in the previous two examples. Namely, we solve
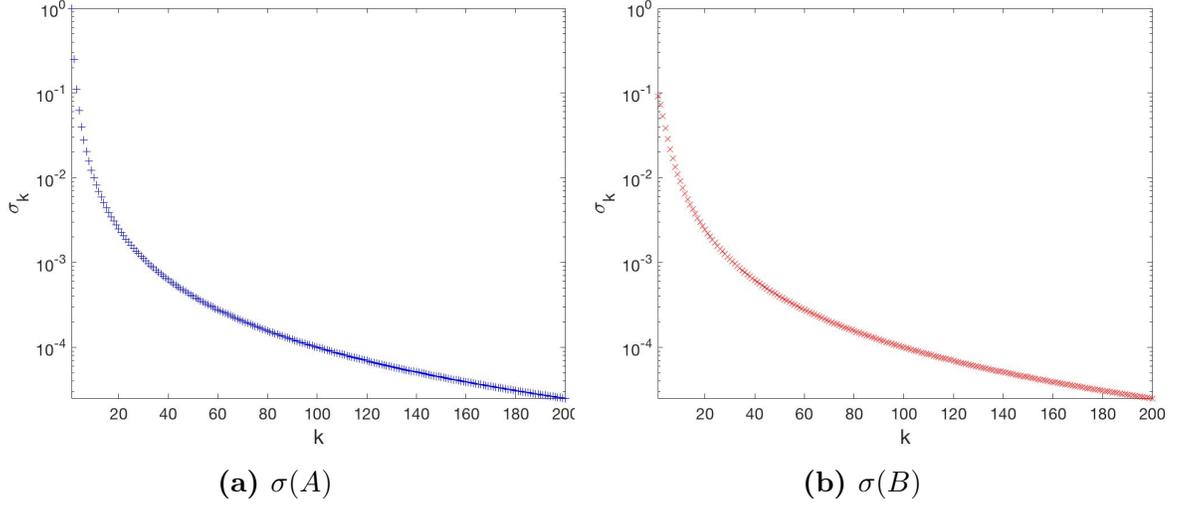
(a) $\sigma(A)$ **(b)** $\sigma(B)$

Figure 4.1: Singular values of individual operators

$$\sigma^2 \phi = C^* C \phi$$

$$= C^* \left(A\phi, B\phi \right)$$

$$= A^* A\phi + B^* B\phi.$$

Using the analytical approach to calculating the singular values in Section 3.2.2, we apply the differential operators and their adjoints to arrive at an ODE,

$$\sigma^2 \left(L_B L_B^* L_A L_A^* \right) \phi = \left(L_B L_B^* + L_A L_A^* \right) \phi \quad \text{or,}$$

$$\sigma^2 \left(\phi^{(8)} + 2b^2 \phi^{(6)} + b^4 \phi^{(4)} \right) = \phi^{(4)} + 2b^2 \phi^{(2)} + b^4 \phi + \phi^{(4)}. \tag{4.1}$$

As mentioned previously, some of our simplifications relied on the fact that $L_A L_A^*$ and $L_B L_B^*$ commute. We show that fact explicitly for this particular example.

Since the operators are self adjoint, L_A^* is equivalent to L_A and L_B^* is equivalent to L_B. We will show that

$$L_A L_A^* L_B L_B^* h = L_B L_B^* L_A^* L_A h,$$

for all $h \in \mathcal{R}\left(A^* A\right)$. Let us first expand the left hand side of the equation.

$$\begin{aligned} L_A L_A^* L_B L_B^* h &= L_A L_A^* L_B \left(h^{(2)} + b^2 h\right) \\ &= L_A L_A^* \left(h^{(4)} + 2b^2 h^{(2)} + b^4 h\right) \\ &= L_A \left(h^{(6)} + 2b^2 h^{(4)} + b^4 h^{(2)}\right) \\ &= h^{(8)} + 2b^2 h^{(6)} + b^4 h^{(4)}. \end{aligned}$$

Now, we expand the right hand side of the equation.

$$\begin{aligned} L_B L_B^* L_A L_A^* h &= L_B L_B^* L_A \left(h^{(2)}\right) \\ &= L_B L_B^* \left(h^{(4)}\right) \\ &= L_B \left(h^{(6)} + b^2 h^{(4)}\right) \\ &= h^{(8)} + 2b^2 h^{(6)} + b^4 h^{(4)}. \end{aligned}$$

Therefore, we can see that

$$L_A L_A^* L_B L_B^* h = L_B L_B^* L_A^* L_A h,$$

and so $L_A L_A^*$ and $L_B L_B^*$ commute.

Notice that the application of the differential operators at the previous steps imply the following boundary conditions:

$$\begin{aligned}
\phi(0) &= \phi(\pi) &= 0, \\
\phi^{(2)}(0) &= \phi^{(2)}(\pi) &= 0, \\
\phi^{(4)}(0) &= \phi^{(4)}(\pi) &= 0, \\
\phi^{(6)}(0) &= \phi^{(6)}(\pi) &= 0.
\end{aligned}$$

The singular function, singular value pairs $\{(\phi_k, \sigma_k)\}_{k=1}^{\infty}$ satisfy equation (4.1) and the corresponding boundary conditions. Even though this is a linear constant coefficient ODE, the eight roots of the characteristic polynomial make solving the boundary value problem analytically outside of the scope of this thesis. We therefore turn to the extension of the Galerkin method presented in section 3.2.3.

4.3   Joint singular values

The discretizations $A^{(n)}$ and $B^{(n)}$, described in Section 3.2.3, approximate the operators A and B with orthonormal bases. Following the approach in [9, 11], we use orthonormal box functions as our bases. The interval domains $\Omega_s = [0, \pi]$ and $\Omega_t = [0, \pi]$ are divided into n subintervals $\{\Omega_s^{(i)}\}$ and $\{\Omega_t^{(i)}\}$ with equal lengths h_s and h_t respectively. The orthonormal box functions are then

$$q_i(s) = \begin{cases} h_s^{-1/2}, & s \in \Omega_s^{(i)} \\ 0, & \text{else} \end{cases}$$

$$p_j(t) = \begin{cases} h_t^{-1/2}, & t \in \Omega_t^{(j)} \\ 0, & \text{else} \end{cases}$$

for $i, j = 1, \ldots n$. Using these in equation (3.7) we have

$$a_{ij}^{(n)} = h_s^{-1/2} h_t^{-1/2} \int_{\Omega_s^{(i)}} \int_{\Omega_t^{(j)}} K_A(s,t) dt ds,$$

and similarly

$$b_{ij}^{(n)} = h_s^{-1/2} h_t^{-1/2} \int_{\Omega_s^{(i)}} \int_{\Omega_t^{(j)}} K_B(s,t) dt ds.$$

For $b = \pi$ the analytically derived singular values are plotted with the numerical approximations on a semi-log scale in Figure 4.2. It should be noted that the use of orthonormal box functions in this Galerkin process is most effective for smooth kernels, as sharp features cannot be fully resolved with finitely many boxes.
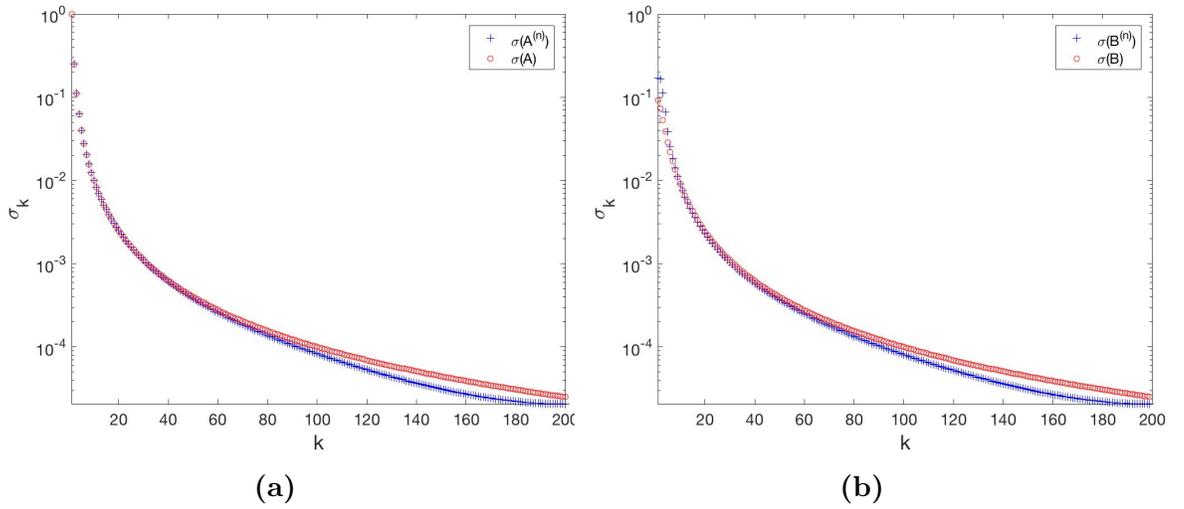
Figure 4.2: Analytical singular values and Galerkin method approximations using orthonormal box functions

The approximation $C^{(n)}$ to the joint operator $C = A \oplus B$ is formed by stacking the matrices $A^{(n)}$ and $B^{(n)}$. i.e.

$$C^{(n)} = \begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix}.$$

We approximate the continuous singular values of C with the discrete singular values of $C^{(n)}$. The values $\sigma(C^{(n)})$ are shown in Figure 4.3 on both a semi-log and log-log scale. Note that the first several singular values have a decay rate that is less than quadratic. Eventually the later ones decay quadratically to zero, but if we wished to consider a truncated SVD solution, we could include more singular values than if we used either model A or B individually.

This modest improvement can clearly be seen in Figure 4.4, where we plot the singular values of the joint operator and singular values of the individual operators. We hypothesis that when we increase the complexity of the models under consideration there will be more substantial gains from joint inversion. We expect this to be especially true when considering models with complementary physics, such as ER and GPR.
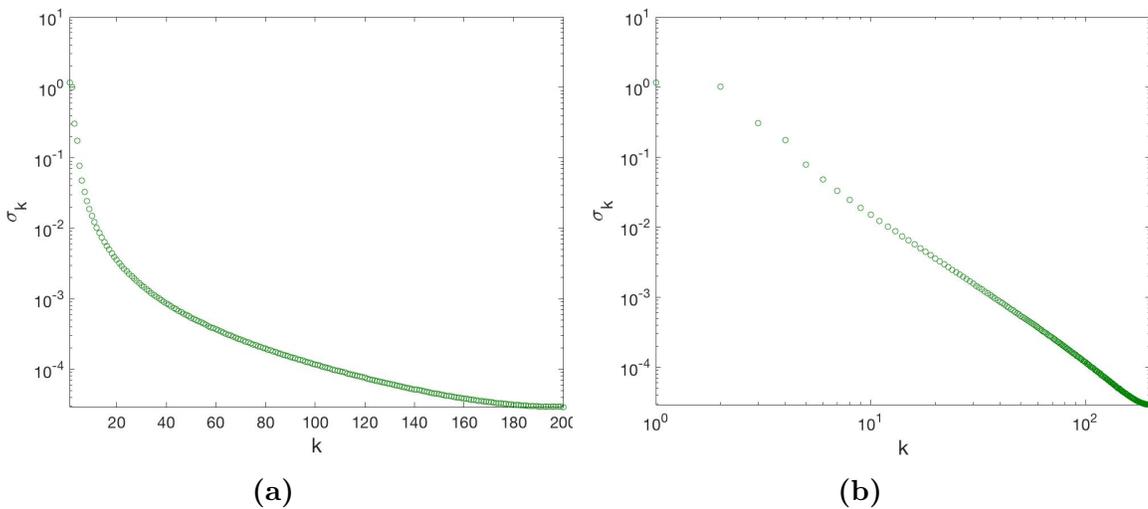
Figure 4.3: Approximate singular values of the joint operator C on a semi-log scale (a), and a log-log scale (b)

Chapter 3 and 4 define and demonstrate the extension of singular value analysis of joint inversion problems to a continuous domain. This differs from the analysis
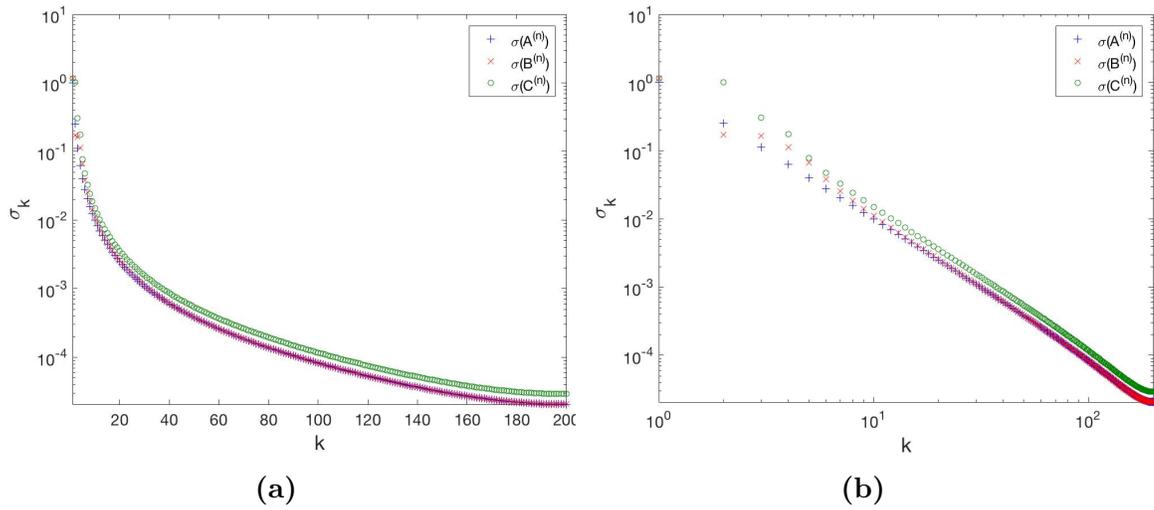
Figure 4.4: Approximate singular values of the individual operators A and B, and the joint operator C on a semi-log scale (a), and a log-log scale (b)

in [22], as there the authors began their study of ill-conditioning with the SVD of already discretized PDEs. The study of ill-posedness in this thesis depends instead on a discretization of the Green's functions which we have shown yields an approximation that converges to the SVE. To our knowledge, this has not been done previously for joint inversion.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In this thesis, we explored inverse problems in both discrete and continuous settings with a focus on ill-conditioning and ill-posedness, respectively. We saw that the primary tool for analysis of discrete linear problems, the singular value decomposition, had an extension to continuous problems with compact linear operators, the singular value expansion. Using the SVE we were able to quantify the ill-posedness of continuous inverse problems by the decay rate of the singular values.

We then considered joint inverse problems. The first mention of joint inversion came in [22], where the authors used the singular value decomposition to determine the degree of ill-conditioning in discrete inverse problems. The authors demonstrated in several examples that combining two models in a joint inversion, and effectively stacking discrete linear models, improved the conditioning of the problem.

We extended these notions to continuous inverse problems with compact linear operators. The analogous operation to stacking matrices is taking the direct sum of operators. We specifically considered the direct sum Green's function solutions to differential equations representing different types of data, and investigated the singular values of this joint operator. Analytical techniques for finding the singular values were presented. These could be useful for simple cases, however, numerical techniques will need to be used in general. To approximate the singular values of the

joint operator, we extended the Galerkin method given in [9, 18]. We the conditions necessary to capture the decay rate of the singular values by this approximation.

The results we have shown in this thesis are for simple cases. To increase the usefulness of the techniques presented here, we will need to construct more realistic and complicated models. The first step towards more realistic models will be the addition of another dimension. Green's function solutions for partial differential equations will naturally require at least two variables related to the dimension of the problem and two associated dummy variables. Therefore, we will need to further extend our numerical techniques.

As we are able to handle more complex models, we can work towards informing experimental design. Our intention is to mathematically test proposed joint inversion schemes based on Green's functions to see what will potentially be gained before discretizing the PDEs. It may not be the case that joint inversion will give substantial improvement over individual inversions. Therefore, these mathematical tests could help prevent unnecessary waste of both experimental and computational resources.

REFERENCES

[1] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier Academic Press, Amsterdam, The Netherlands, 2005.

[2] J.B. Conway. *A Course in Functional Analysis*. Graduate Texts in Mathematics. Springer New York, 1994.

[3] Dean G Duffy. *Green's Function with Applications*. CRC Press, 2015.

[4] Andrew Geary. Setup of a dc resistivity survey, 2016.

[5] New Jersey Geological and Water Survey. Ground penetrating radar, 2005.

[6] Mark Gockenbach. *Linear Inverse Problems and Tikhonov Regularization*. The Mathematical Association of America, 2015.

[7] Mark S Gockenbach. Generalizing the gsvd. *SIAM Journal on Numerical Analysis*, 54(4):2517–2540, 2016.

[8] Paul Richard Halmos and Viakalathur Shankar Sunder. *Bounded integral operators on L2 spaces*. Ergebnisse Mathematik und GrenzGeb. Springer, Berlin, 1978.

[9] Per Christian Hansen. Computation of the singular value expansion. *Computing*, 40(3):185–199, 1988.

[10] Per Christian Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, volume 4. Siam, 1998.

[11] Per Christian Hansen. Regularization tools version 4.0 for matlab 7.3. *Numerical algorithms*, 46(2):189–194, 2007.

[12] Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.

[13] Muneo Hori. Inverse analysis method using spectral decomposition of green's function. *Geophysical Journal International*, 147(1):77–87, 2001.

[14] John D Jackson. *Classical electrodynamics*. John Wiley & Sons, Inc., New York, NY,, 1999.

[15] Michel Kern. *Numerical Methods for Inverse Problems*. John Wiley & Sons, 2016.

[16] Andreas Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer Science & Business Media, 2011.

[17] Rainer Kress, V Maz'ya, and V Kozlov. *Linear integral equations*, volume 82. Springer, 1989.

[18] Rosemary A Renaut, Michael Horst, Yang Wang, Douglas Cochran, and Jakob Hansen. Efficient estimation of regularization parameters via downsampling and the singular value expansion. *BIT Numerical Mathematics*, pages 1–31, 2016.

[19] Bahaa Saleh. *Introduction to subsurface imaging*. Cambridge University Press, 2011.

[20] Gilbert W Stewart. On the early history of the singular value decomposition. *SIAM review*, 35(4):551–566, 1993.

[21] Andreĭ Nikolaevich Tikhonov, Vasiliĭ IAkovlevich Arsenin, and Fritz John. *Solutions of ill-posed problems*, volume 14. Winston Washington, DC, 1977.

[22] K Vozoff and DLB Jupp. Joint inversion of geophysical data. *Geophysical Journal International*, 42(3):977–991, 1975.

[23] Michael S Zhdanov. *Inverse theory and applications in geophysics*, volume 36. Elsevier, 2015.