
#### Abstract

HUMBER, CARY ROSS. Sparse Regularization for Inverse Problems Governed by Evolution Equations. (Under the direction of Kazufumi Ito.)

The purpose of this thesis is twofold. Firstly, we introduce a novel method for estimating the state of a system governed by a linear evolution equation. The method utilizes the adjoint of the partial differential equation (PDE) and a basis for the Hilbert space to accurately reconstruct the initial condition. The method also provides a filter bank which can be utilized for the purpose of reconstructing initial conditions based on given data. We then extend the method to include source identification and simultaneous state/parameter estimation for a certain class of problems.

Secondly, we develop and analyze the multi-parameter regularization necessary for the accurate approximation of inverse problem solutions. The regularization is essential for both the state estimation method developed in this thesis, as well as for the general inverse problem theory. The multi-parameter regularization allows for solutions which may have a multi-scale profile. Specifically, we address problems involving sparsely distributed measurements. In addition, solutions which are, themselves, locally supported are treated, such as collections of point sources. The method developed is widely applicable and accurate, as demonstrated in this thesis.


(c) Copyright 2011 by Cary Ross Humber

## All Rights Reserved

by
Cary Ross Humber

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina
2011

APPROVED BY:

| $\substack{\text { Alina Chertock }}$ | Ralph Smith |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Zhilin Li |  |
| Chair of Advisory Committee |  |

## DEDICATION

To my wife, Katie LeWare Humber.

## BIOGRAPHY

Cary Ross Humber was born in Memphis, Tennessee in 1983. He earned his Bachelors of Science in Mathematics in 2005. After completing his undergraduate degree, Cary began his graduate studies at North Carolina State University in August of 2006. In September of 2009 Cary began employment with the Naval Surface Warfare Center in Panama City Beach, Florida as a mathematician, while completing his doctoral research.

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Professor Kazufumi Ito, for his guidance and insight while working with him. His encouragement and interest in my work proved to be invaluable.

I am also appreciative for the members of my committee, including Professor Ralph Smith for serving as a last minute substitute.

I would like to express my appreciation for the In-House Laboratory Independent Research (ILIR) program at the Naval Surface Warfare Center, Panama City, Fl, for hosting me during the last research period of my Ph.D. program.

I would also like to express my gratitude to Dr. Quyen Huynh of NSWC-PC for his support and encouragement during the final stages of my Ph.D.

I am very grateful for the support from all of my family and friends. In particular, my wife's loving support has never waned. I am also grateful for the hospitality and support of Jack and Jane Hayes, especially for the accommodations. My mother, father, Reagan and Sarah have been equally supportive and have encouraged me to do things I might otherwise not do.

## TABLE OF CONTENTS

List of Figures ..... vii
Chapter 1 Introduction ..... 1
1.1 Motivation ..... 1
1.2 Examples ..... 5
1.3 Contributions of this thesis ..... 8
1.4 Outline ..... 8
Chapter 2 State Estimation for the Abstract Cauchy Problem ..... 10
2.1 Problem Description ..... 10
2.2 Standard Least-Squares Approach ..... 11
2.3 A Dual Method for approximating the Fourier expansion of the initial condition ..... 15
2.4 Variation of the Dual Control Method ..... 23
2.5 Kalman-Bucy and Time-Reversal methods ..... 25
2.6 Implementation Issues ..... 29
Chapter 3 Extensions of the reconstruction algorithms ..... 33
3.1 Simultaneous State and Parameter Estimation ..... 33
3.2 Source Identification ..... 38
3.2.1 Source Identification in the Reproducing Kernel Hilbert Space Frame- work ..... 40
Chapter 4 Regularization for Inverse Problems ..... 44
4.1 Motivation and Preliminaries ..... 44
4.2 Regularization Methods ..... 46
4.2.1 Standard Tikhonov Regularization ..... 46
4.2.2 Generalized Multi-parameter Approach ..... 50
4.3 Choice rules for the regularization parameter ..... 53
4.3.1 Morozov's Discrepancy Principle ..... 55
4.3.2 Complexity Level ..... 56
4.3.3 Balance principle ..... 56
Chapter 5 Nonsmooth regularization and Sparsity Optimization ..... 62
5.1 Primal-Dual Active Set Method for unilateral constraint ..... 66
Chapter 6 Numerical Tests ..... 71
6.1 1-D Diffusion Equation ..... 71
6.2 2-D Diffusion Equation ..... 80
6.3 2-D Convection-Diffusion Equation ..... 83
Chapter 7 Two real world applications ..... 89
7.1 Application to data classification ..... 89
7.2 Application of Reconstruction to Synthetic Aperture Sonar Imaging ..... 93
7.2.1 Problem formulation ..... 93
7.2.2 Background ..... 94
7.2.3 Parameterized Range-Doppler algorithm ..... 96
7.2.4 PDE migration ..... 97
7.3 Concluding remarks ..... 99
References ..... 101
Appendices ..... 105
Appendix A Semigroup Theory ..... 106
A. 1 Mild Solutions to the Abstract Cauchy Problem ..... 107
Appendix B Definitions and some Background Material ..... 110
Appendix C Sparsity Regularization for a Neural Classification Problem ..... 113
C. 1 The Support Vector Machine ..... 113
C. 2 New Approach ..... 115
C.2.1 Improving separability via inequality constraint ..... 115
C.2.2 Weights for reducing bias ..... 116
C.2.3 Algorithm ..... 117

## LIST OF FIGURES

Figure 5.0.1 Comparison of approximate subdifferentials for $\varepsilon=10 e^{-3}$ and $\varepsilon=10 e^{-2}$. ..... 64
Figure 6.1.1 Measurements taken for 1-D diffusion equation ..... 72
Figure 6.1.2 1-D reconstruction for diffusion equation ..... 73
Figure 6.1.3 Control corresponding to $\sin (2 \pi x)$. ..... 74
Figure 6.1.4 1-D reconstruction for variable coefficient diffusion equation ..... 75
Figure 6.1.5 1-D reconstruction for variable coefficient diffusion equation ..... 75
Figure 6.1.6 Comparison using different basis functions ..... 76
Figure 6.1.7 Simultaneous state/parameter estimation for 1-D diffusion equation 77Figure 6.1.8 Results for simultaneous state/parameter estimation of 1-D diffu-sion equation77
Figure 6.1.9 Point source identification for 1-D diffusion equation ..... 79
Figure 6.2.1 2-D reconstruction for diffusion equation ..... 81
Figure 6.2.2 Results for simultaneous state/parameter estimation of 2-D diffu- sion equation ..... 82
Figure 6.2.3 Measurements taken for 2-D point source identification ..... 83
Figure 6.2.4 2-D point source identification for 2-D diffusion equation ..... 84
Figure 6.3.1 Measurements and locations for 2-D convection-diffusion equation ..... 86
Figure 6.3.2 Reconstruction for 2-D convection-diffusion equation ..... 86
Figure 6.3.3 Comparison of 2-D convection-diffusion and diffusion equations ..... 87
Figure 6.3.4 2-D state/parameter estimation for convection diffusion ..... 88
Figure 7.1.1 Cues for the wrist up action. ..... 91
Figure 7.1.2 Comparison of weights for different $\ell_{p}$ norms. ..... 91
Figure 7.1.3 Comparison of PSVM and nonsmooth formulation. ..... 92
Figure 7.1.4 Comparison of data classifiers ..... 92
Figure 7.2.1 Synthetic Aperture Sonar setup. ..... 93
Figure 7.2.2 Comparison of SAS algorithms ..... 98

## Chapter 1

## Introduction

### 1.1 Motivation

The motivation for this thesis originates from the practical needs of many classically difficult inverse problems in the mathematical sciences, as well as some of the more recent needs throughout the sciences. The primary mathematical focus of this thesis is the solution of inverse problems for the purpose of reconstructing states and identifying parameters of evolution equations. Roughly speaking, inverse problems involve determining a solution to a problem for which the governing dynamics are known, and the measurements, or output, of the system are given. For instance, one may be given a time series, or sampling, of the current heat profile of a metal plate and be asked what the initial heat profile was one hour prior. Of course, if a full history of data is available its just a matter of searching in the right place. Many of the most interesting and relevant scientific problems involve an incomplete time history of data. This is the case either because of storage constraints or due to unavailability/inaccessibility of data. In addition, the data may be sparsely distributed across the "area" being measured.

Inverse problems have both theoretical and numerical difficulties, making them interesting to consider, mathematically. Furthermore, inverse problems are very important scientifically, as there is certainly some sort of inverse problem in every corner of the scientific world. Many scientific efforts are devoted to dealing with constantly growing size and availability of data. On the contrary, the necessity for stably reconstructing solutions of dynamical systems from sparse measurements is paramount. By sparse measurements, we mean either the measurements are either sparsely distributed over the
domain in time, or the measurements are only available for select time values, such as the final time. Not only is it necessary to reconstruct solutions from sparse measurements for obvious reasons, but the ability to reconstruct solutions from a sparse set of data can help mitigate the constant flood of data. This thesis seeks to generalize some of the work already produced in the field of dynamical inverse problems and to improve upon the solution methods for stable reconstruction.

Not only is there increased interest in data reduction/sparsity, but there is increased motivation for robust algorithms that can adapt to varying noise levels in the data. Problems such as numerical weather prediction involve solving large scientific problems repeatedly throughout the day (and night). It is desired for the algorithms to adapt to changes in the data and to fluctuations in noise levels. We utilize the Tikhonov type regularization to stably construct solutions, where the parameters are chosen based on a priori information. Given the needs of the scientific community, the two concepts of sparsity/compression and adaptability/tunability are at the forefront throughout this thesis.

As a consequence of the methods developed in this thesis, we are able to effectively deal with some of the problems involved with large datasets, effectively mining the data. Though the availability of data can certainly be a blessing, there are difficulties that must be overcome when dealing with large datasets. The ever-growing size and availability of data plagues many already difficult problems. Not only that, but it has become increasingly apparent over the last several years that some problems can be solved given very little data in comparison to the size of the problem (i.e., compressed sensing [7]). This is good news for mathematicians tasked with solving difficult problems, but it also provides opportunity for developing hardware and software capable of efficiently utilizing compressed data. Of course, this thesis is void of hardware/software issues, however many questions remain pertaining to compressed sensing, sparsity optimization and general inverse problem theory for systems governed by evolution equations.

In the general framework, we consider the inverse (deconvolution) problem of the form

$$
\begin{equation*}
K v=y^{\delta} \tag{1.1.1}
\end{equation*}
$$

where $K$ is a compact operator. The data, $y^{\delta}$, is assumed to be noisy and possibly incomplete. The operator, $K$, may correspond to a number of operations, including convolution, integration, scattering, et cetera. Numerous applications have been studied
and continue to be studied, while new applications continue to emerge which fall into this framework. An incomplete list includes computerized tomography, inverse scattering, data assimilation for weather forecasting, and so on. A more complete set of examples can be found in the monographs $[9,11,22]$.

As already mentioned, we are particularly motivated by state and parameter estimation problems involving systems governed by partial differential equations (PDE). For instance, in the meteorological sciences, one may wish to determine the atmospheric pressure or velocity from sensor data. The essential features can be modeled by the (linear) convection-diffusion equation

$$
\begin{equation*}
\frac{\partial v}{\partial t}=c(x) \cdot \nabla v+\nabla \cdot(d(x) \nabla v)+f(x) . \tag{1.1.2}
\end{equation*}
$$

For this particular problem, the unknown is the state, $v\left(x, t_{1}\right)$, at some time in the recent past, $t_{1}$. In general, the coefficients $c, d$ may be unknown as well. Actual observations are taken at various sample locations. Using the data and the PDE formulation, an approximate past state is solved for, which is then used as the initial condition for a simplified numerical weather prediction. The full formulation of this problem, based on the Navier-Stokes equations, can be found in the paper [14].

Other related, but simpler, PDE that have similar inverse problem formulations are the diffusion equation

$$
\begin{equation*}
\frac{\partial v}{\partial t}=\nabla \cdot(d(x) \nabla v)+f(x), \tag{1.1.3}
\end{equation*}
$$

and the reaction-diffusion equation

$$
\begin{equation*}
\frac{\partial v}{\partial t}=D \Delta v+f(v) \tag{1.1.4}
\end{equation*}
$$

For the corresponding inverse problems, we are given either a time series of sampled solution values, $v(x, t)$, or a sample of the final time solution $v\left(x, t_{f}\right)$. That is, certain regions are observed in time, such as the boundary or averages over subdomains; or the final state is observed at distributed locations. In either case, for practical applications, the data is noisy and biased. The problem is then to determine the initial condition of the PDE given the noisy partial measurements. The identification of the initial condition of PDE may be ill-posed, especially if the equation is of parabolic type, such as (1.1.2) or (1.1.3). A problem is ill-posed if one or more of the following is not satisfied:

1. For all admissible data, a solution exists,
2. For all admissible data, the solution is unique,
3. The solution depends continuously on the data.

The concept of ill-posedness and methods for dealing with it will be made more clear in Chapter 4. An example of a highly ill-posed problem is the homogeneous backwards heat equation

$$
\begin{align*}
& \frac{\partial v}{\partial t}=\nabla \cdot(c(x) \nabla v), \quad x \in \Omega  \tag{1.1.5}\\
& v\left(x, t_{f}\right)=\phi(x)  \tag{1.1.6}\\
& v_{\Gamma}=0, \quad \Gamma=\partial \Omega \tag{1.1.7}
\end{align*}
$$

due to the amplification of errors by the generalized Fourier expansion of the solution (see [11] for further details). Here, we are solving the heat equation backwards in time, given the final time heat profile. The problem's difficulty is compounded when the final temperature profile is only available on a subdomain. That is, we are given measurements

$$
y=v\left(x, t_{f}\right) \quad x \in \Omega_{s}
$$

for some set $\Omega_{s} \subset \Omega$. For our purposes, we assume the points $x \in \Omega_{s}$ are sparsely distributed in $\Omega$. Solving the heat equation backwards introduces large errors due to integrating small errors from the noisy data. Due to the nature of the fundamental solution of the heat equation, the small errors in the data become exponentially amplified. Identifying the initial heat profile from the full final time heat profile is a more classical inverse problem. Modern day problems of this type have evolved to include more practical concerns. For example, as already mentioned, the data may be a partial time series of the solution. In this case, at each time step, the solution may be available for output at select points in the domain, or the data may be available as averages over one or more subdomain. In the latter case, we are given measurements

$$
y(t)=C v(t)
$$

where

$$
C v(t)=\frac{1}{\mu\left(\Omega_{s}\right)} \int_{\Omega_{s}} v(s) d \mu
$$

for some set $\Omega_{s} \subset \Omega$ having volume $\mu\left(\Omega_{s}\right)$. If the output available is relatively small, this is essentially the idea of sparsity/compression, which has become increasingly important over the last several years in the sciences. For instance, the available data may be sparse in comparison to the size of the problem (i.e., $10 \%$ of a region may be available for measurement). Given very little information, it is then necessary to reconstruct the solution as accurately as possible. Fortunately, the general theory allows us to do so (see [7]). The need to reconstruct solutions from sparsely available data is not only true for PDE inverse problems, but for the general inverse problem (1.1.1).

In the same way, one may also be interested in reconstructing solutions that are sparsely distributed themselves, rather than the data being sparsely distributed. For example, the initial condition of the heat equation may be a collection of point sources, $\delta\left(x-x_{j}\right)$, located at $x_{j}$. Here, one needs to use the a priori information about the solution (i.e., sparsity) to assist in the recovery. Other formulations include identifying the unknown coefficients $c, d$ in addition to the initial condition, $v(0)$, being unknown.

To develop and analyze methods for solving problems of the general form (1.1.3), (1.1.2) we cast the PDE as an abstract Cauchy problem of the form

$$
\begin{align*}
\frac{d v}{d t}(t) & =A v(t)  \tag{1.1.8}\\
y(t) & =C v(t) \tag{1.1.9}
\end{align*}
$$

where $A$ is the generator of a $C_{0}$-semigroup, and $C$ is an observation operator. This formulation, and methods for solving inverse problems derived from this formulation, will be discussed in Chapter 2.

### 1.2 Examples

Solving inverse problems based on the formulation (1.1.8)-(1.1.9) and (4.1.1) is very important for many scientific applications, such as weather forecasting, data classification, image processing, inverse scattering, financial analysis, et cetera. To forecast some of the potential applications of the methods developed in this thesis, we give several examples of

PDE that will fall into the general framework. Note that not all of the examples are actually considered in numerical results of this thesis, however, the methods are applicable to each PDE.

## Convection-Diffusion Equation

The homogeneous convection-diffusion equation with homogeneous Dirichlet boundary condition is given by

$$
\begin{align*}
\frac{\partial v}{\partial t}(x, t) & =c(x) \cdot \nabla v(x, t)+\nabla \cdot(d(x) \nabla v(x, t))  \tag{1.2.1}\\
v(x, 0) & =v_{0}(x)  \tag{1.2.2}\\
v(x, t) & =0 \quad x \in \partial \Omega \tag{1.2.3}
\end{align*}
$$

where the first term on the right-hand side corresponds to convection, while the second term corresponds to diffusion. The coefficients $c, d$ are the convection and diffusion coefficients, respectively. The convection-diffusion equation can be thought of as modeling a simplified weather system or of a mass-transport system. In the case of a simplified weather system, our method allows the reconstruction of a past weather state. This is a linear PDE, unlike the Navier-Stokes, so it falls into the framework presented in this thesis without modification.

## Wave Equation

The second order linear hyperbolic PDE

$$
\begin{align*}
\frac{\partial^{2} v}{\partial t^{2}}(x, t) & =c^{2} \Delta v(x, t)  \tag{1.2.4}\\
\frac{\partial}{\partial t} v(x, 0) & =\psi(x)  \tag{1.2.5}\\
v(x, 0) & =0 \tag{1.2.6}
\end{align*}
$$

describes the propagation of waves with a speed $c$. Given its generality, numerous scientific problems can be modeled by the wave equation or a modified wave equation. For instance, the one way wave equation has been utilized for PDE migration techniques which is outlined in Section 7.2.4 and the report [23].

## Black-Scholes

The Black-Scholes model for pricing European options is described by a parabolic equation of the form

$$
\begin{array}{r}
-\frac{d}{d t} v(t, S)-\left(\frac{\sigma^{2}}{2} S^{2} v_{S S}+(r-\delta) S v_{S}-r v\right)=0 \\
v\left(t_{f}, S\right)=\psi(S) \tag{1.2.8}
\end{array}
$$

Here, $S>0$ denotes the price of the stock, $v$ the value of the share, $r>0$ is the interest rate, $\delta$ the influence of dividends, and $\sigma>0$ is the volatility. Further, $T$ is the maturity date and $\psi$ is the reward function. Typically, the reward function is $\psi(S)=\max (0, K-S)$ for the put option and $\psi(S)=\max (0, S-K)$ for the call option, where $K$ is the strike price. Interesting inverse problems to consider would be to determine the reward function given the stock index, or to recover the unknown strike price $K$. The formulation of such problems is discussed in the paper [21].

## Navier-Stokes

It is beyond the scope of this thesis, due to nonlinearity, however the long-term goal includes the extension to the nonlinear case. In this case, inverse problems involving the incompressible Navier-Stokes equations may be considered:

$$
\begin{align*}
\frac{\partial}{\partial t} v_{i}+\sum_{j=1}^{n} v_{j} \frac{\partial v_{i}}{\partial x_{j}}= & \nu \Delta v_{i}-\frac{\partial p}{\partial x_{i}}+f_{i}(x, t) \quad\left(x \in \mathbb{R}^{n}, t \geq 0\right)  \tag{1.2.9}\\
& \operatorname{div} v=\sum_{i=1}^{n} \frac{\partial v_{i}}{\partial x_{i}}=0 \quad\left(x \in \mathbb{R}^{n}, t \geq 0\right) \tag{1.2.10}
\end{align*}
$$

with initial conditions

$$
v(x, 0)=v_{0}(x) \quad\left(x \in \mathbb{R}^{n}\right)
$$

In this case, the unknowns may be the initial condition $v(x, 0)$, the viscosity $\nu$, or the force $f_{i}$.

The field of inverse problems has been widely studied, but there are still unanswered or partially answered questions. Advances in theoretical mathematics, computer technology and data storage have all contributed to more interesting and important problems being
solved. It is the goal of this thesis to give new insight into the solution of inverse problems from sparsely distributed data or when the solution itself is assumed to be sparsely distributed. Several applications will be considered in the Numerical Tests section of this thesis which demonstrate the applicability of the methods developed.

### 1.3 Contributions of this thesis

The principal contribution of this thesis is twofold. First we introduce a new approach for estimating the initial condition of the abstract Cauchy problem in Chapter 2. Further, we show how the method produces stable approximations and accurately reconstructs the initial state. In short, the method is based on the dual control problem of (1.1.8)

$$
\begin{equation*}
-\frac{d p}{d t}(t)=A^{*} p(t)+C^{*} u(t) \tag{1.3.1}
\end{equation*}
$$

where the control, $u$, is selected so that we can estimate the generalized Fourier coefficients of the initial condition, $x_{0}$, for a chosen basis. Here, we need to solve the corresponding control problem with a certain regularity of the control. Thus, we develop the necessary regularization tools in Chapter 4. The overall method for state estimation involves

1. Integrating the dual equation (1.3.1) against the state equation (1.1.8)
2. An appropriate basis selection for the Hilbert space $X$
3. Tikhonov type regularization for determining the control

Secondly, we develop a generalized regularization framework for solving a wider class of inverse problems in Chapter 4. The development of the regularization framework is motivated by the state estimation problem, but is necessary for numerous ill-posed inverse problems, as will be demonstrated throughout this thesis.

Lastly, we provide demonstrations of the applicability of the method.

### 1.4 Outline

In Chapter 2, we develop the necessary background and tools for the introduction of a novel technique for state estimation in systems governed by PDE. The method utilizes
the adjoint equation (1.3.1) to yield a filter bank of controls which can be utilized for repeated initial condition estimations, based on time-series measurements. We are also able to forecast future states based on the formulation derived. Further, the method has several formulations which are analyzed and compared with the standard linear leastsquares approach for state estimation. Briefly, the least-squares approach does not allow for control of the errors, however, the dual control method has a concrete error estimate. In addition, the methods can be combined with the standard Kalman-Bucy filter and with time-reversal methods, as developed in Chapter 2.

In Chapter 3, we provide two extensions of the methods developed in Chapter 2 to problems of source identification and simultaneous state/parameter estimation for locally constant parameters. The extensions allow for more difficult problems to be considered showing the strength of the methods.

In Chapters 4 and 5, we develop the theory and algorithms for ill-posed inverse problems, with an emphasis on multi-parameter regularization and sparsity optimization. These tools are necessary for the methods developed in Chapters 2 and 3, as well as for more general inverse problems. In particular, the sparsity optimization has become an increasingly important computational tool for obtaining compressed solutions and feature selection.

Finally, in Chapter 6, we give extensive numerical tests. The numerical tests are based on initial condition estimation and source identification for the diffusion and convectiondiffusion equations. Several formulations are considered, including point source identification from sparse measurements. As a consequence of what may be considered, two real world applications are briefly presented in Chapter 7 which are indirectly related to the methods developed in this thesis. However, the applications show the strength of the multi-parameter and sparsity regularization. We conclude with future research directions based on this thesis, including the extension to nonlinear problems.

## Chapter 2

## State Estimation for the Abstract Cauchy Problem

### 2.1 Problem Description

Consider the abstract Cauchy problem

$$
\begin{equation*}
\frac{d x}{d t}(t)=A x(t)+f(t) \tag{2.1.1}
\end{equation*}
$$

where we have measurements

$$
\begin{equation*}
y(t)=C x(t) \tag{2.1.2}
\end{equation*}
$$

for $x$ in a Hilbert space $X$ (e.g., $X=L^{2}(\Omega)$ ). We assume the operator $A$ is linear and generates a strongly continuous semigroup $S_{t}: \mathbb{R} \rightarrow \mathcal{L}(X)$, while $C: X \rightarrow Y$ is an observation operator. Depending on the specific problem, it may be desired to determine the initial condition $x(0)$ or the final state $x\left(t_{f}\right)$, given the incomplete (and possibly noisy) measurements $y$. The primary focus in this thesis is the reconstruction of the initial condition, $x_{0}$, from measurements $y(t), 0 \leq t \leq t_{f}$. The forward problem of forecasting a future state is almost surely better posed than the corresponding backwards problem, however, certain applications require going backward and then forward (or vice versa), such as weather prediction. We are especially interested in the case of partial measurements (i.e., the measurements are sparsely distributed over the domain $\Omega$ ). In the following section, we develop methods for determining the initial state $x(0)$, though the same methods are applicable to forecasting the state $x\left(t_{f}\right)$ with only minor adaptations
as will be shown. We begin by describing three approaches for accomplishing this task in Sections 2.2-2.4. In each case, we assume the Hilbert space $X$ is separable, so that there exists a complete orthonormal sequence $\left\{\varphi_{k}\right\}_{k=0}^{\infty}$.

The three approaches are similar in nature, since the approximation of $x(0)=x_{0}$ is taken to be

$$
x_{0}^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k},
$$

where $x_{0}^{m} \in X_{m} \subset X$ and the coefficients $\alpha_{k}$ are to be determined. The connection between this Chapter and Chapter 4 is that suitable regularization techniques are necessary, whether the initial condition reconstruction is computed directly, or from the methods developed in the following sections. This connection will be discussed in Chapter 4.

We also give details of the Kalman-Bucy filter and a Kalman-Bucy based time-reversal method for the case when the operator $A$ is skew-adjoint. This method is iterative in nature in contrast to the method developed in this thesis. Finally, we outline some of the issues necessary for the implementation of the methods developed in this chapter.

### 2.2 Standard Least-Squares Approach

In this section, we describe the standard least-squares approach for solving inverse problems involving the abstract Cauchy problem. The method described is well-known and has a straight forward implementation. The least-squares method is highly applicable to a wide-range of scientific problems, however, there are limitations that will be discussed in this section. These limitations are the motivation for developing the algorithms in the sections to follow, as these limitations will be overcome.

For the system governed by

$$
\begin{equation*}
\frac{d x}{d t}(t)=A x(t)+f(t) \tag{2.2.1}
\end{equation*}
$$

we assume $f$ is known, and we are given the measurements in time

$$
y(t)=C x(t)
$$

The measurements, $y$, are output from the system described by the Cauchy problem. For instance, we could be given temperature profiles, atmospheric pressure, wind velocity, et
cetera. Given the time sampling of the solution profile, it is necessary to estimate the initial solution profile. The theory for the abstract Cauchy problem is well-developed, allowing a systematic formulation of the problem at hand. We know that the solution of (2.2.1) in terms of the initial condition, $x_{0}$, is given by the formula

$$
x(t)=S_{t} x_{0}+\int_{0}^{t} S_{t-s} f(s) d s
$$

where $S_{t}$ is the $C_{0}$-semigroup generated by $A$. Given the formula for $x(t)$ in terms of $x_{0}$, we have the input to output relationship

$$
y(t)=C S_{t} x_{0}+C \xi(t), \quad 0 \leq t \leq t_{f}
$$

where have defined

$$
\xi(t)=\int_{0}^{t} S_{t-s} f(s) d s
$$

Now, for notational convenience throughout this chapter, we define the operator

$$
\begin{equation*}
\mathcal{M}=C S_{t} \quad 0 \leq t \leq t_{f} \tag{2.2.2}
\end{equation*}
$$

where $t_{f}$ is the final sampling time. If $\mathcal{M}$ is invertible, we can simply obtain $x_{0}$ as

$$
x_{0}=\mathcal{M}^{-1} y-\mathcal{M}^{-1} C \xi
$$

however, $\mathcal{M}$ is, in general, a compact operator, which implies an unbounded inverse. Whenever $\mathcal{M}$ is not invertible, a common approach is the linear least-squares(LLS) method given by

$$
\begin{equation*}
\min _{x \in X} \frac{1}{2}\|y-\mathcal{M} x\|_{L^{2}\left(0, t_{f} ; Y\right)}^{2} \tag{2.2.3}
\end{equation*}
$$

for the case when $f \equiv 0$. Since $\mathcal{M}$ is compact, we consider the regularized linear leastsquares(RLLS) method, which is given by

$$
\begin{equation*}
\min _{x \in X} \frac{1}{2}\|y-\mathcal{M} x\|_{L^{2}\left(0, t_{f} ; Y\right)}^{2}+\frac{\beta}{2}\|x\|_{X}^{2} \tag{2.2.4}
\end{equation*}
$$

where $\beta>0$ is a chosen parameter. Due to the compactness of $\mathcal{M}$, the problem of solving

$$
\mathcal{M} x_{0}=y
$$

is ill-posed, so the regularization parameter, $\beta$, is necessary, and must be chosen carefully. The general theory of solving ill-posed linear problems, and the issue of how to select the parameter $\beta$ is discussed in Chapter 4. By the necessary optimality of (2.2.4) we have the closed form solution

$$
x^{\dagger}=\left(\mathcal{M}^{*} \mathcal{M}+\beta I\right)^{-1} \mathcal{M}^{*} y
$$

where $\mathcal{M}^{*}$ is the adjoint of $\mathcal{M}$ and $I$ is the identity operator in $\mathcal{L}(X)$.
A more functional analytic approach, for the case when $f \equiv 0$, is described as follows. We assume $\left\{\varphi_{k}\right\}_{k=0}^{m}$ is an orthonormal basis for $X_{m} \subset X$ and seek an

$$
x_{0}^{m} \in X_{m}=\operatorname{span}\left\{\varphi_{k} ; 0 \leq k \leq m\right\}
$$

such that

$$
\frac{1}{2}\left\|y-\mathcal{M} x_{0}^{m}\right\|_{L^{2}\left(0, t_{f} ; Y\right)}^{2}+\frac{\beta}{2}\left\|x_{0}^{m}\right\|_{X}^{2}
$$

is minimal. In this case, the task is to determine the minimizer $\alpha=\left(\alpha_{0}, \ldots, \alpha_{m}\right)$ of

$$
\frac{1}{2}\left\|y-\mathcal{M} \sum_{k=0}^{m} \alpha_{k} \varphi_{k}\right\|_{L^{2}\left(0, t_{f} ; Y\right)}^{2}+\frac{\beta}{2}\left\|\sum_{k=0}^{m} \alpha_{k} \varphi_{k}\right\|_{X}^{2}
$$

Since $\left\{\varphi_{k}\right\}_{k=0}^{m}$ is orthonormal, the coefficients $\alpha_{k}$ are simply the generalized Fourier coefficients

$$
\left\langle x_{0}, \varphi_{k}\right\rangle_{X}, \quad 0 \leq k \leq m
$$

Now, the minimization

$$
\min _{x_{0}^{m} \in X_{m}} \frac{1}{2}\left\|y-\mathcal{M} x_{0}^{m}\right\|_{L^{2}\left(0, t_{f} ; Y\right)}^{2}+\frac{\beta}{2}\left\|x_{0}^{m}\right\|_{X}^{2}
$$

is equivalent to

$$
\begin{equation*}
\min _{\alpha \in \mathbb{R}^{m+1}} \frac{1}{2} \alpha^{t} W \alpha-\sum_{k=0}^{m} \alpha_{k}\left\langle y, \mathcal{M} \varphi_{k}\right\rangle+\frac{1}{2}\|y\|^{2}+\frac{\beta}{2} \alpha^{t} P \alpha \tag{2.2.5}
\end{equation*}
$$

where $W_{k, l}=\left\langle\mathcal{M} \varphi_{k}, \mathcal{M} \varphi_{l}\right\rangle$. In general, the minimizer of (2.2.5) is determined by

$$
(W+\beta P) \alpha=F
$$

for $F_{k}=\left\langle y, \mathcal{M} \varphi_{k}\right\rangle$, where $P$ is a positive symmetric matrix which depends on the regularization norm we select. If the regularization norm is $\|x\|_{X}$, as in (2.2.4), then $P$ is the identity matrix

$$
P_{i, j}=\left\langle\varphi_{i}, \varphi_{j}\right\rangle_{X}
$$

by the orthonormality of $\left\{\varphi_{k}\right\}_{k=0}^{m}$. With $\alpha$ determined in this manner, we form the approximate initial state by

$$
x_{0}^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k} .
$$

This approach can be summarized by the following:

## Regularized Linear Least-Squares Algorithm:

1. Pick a basis $\left\{\varphi_{k}\right\}_{k=0}^{m}$ for $X_{m} \subset X$, pick $\beta>0$, and a regularization weight $P$.
2. Compute the Gram matrix $W_{k, l}=\left\langle\mathcal{M} \varphi_{k}, \mathcal{M} \varphi_{l}\right\rangle_{Y}$ and the vector $F_{k}=$ $\left\langle\mathcal{M}^{*} y, \varphi_{k}\right\rangle_{X}$, or, with force $f, F_{k}=\left\langle\mathcal{M}^{*}(y-C f), \varphi_{k}\right\rangle_{X}$.
3. Solve $(W+\beta P) \alpha=F$ for $\alpha$.
4. Compute the approximation

$$
x_{0}^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k} .
$$

As has been shown, the LLS and RLLS methods can be seen as methods for approximating the generalized Fourier coefficients of the unknown function $x_{0} \in X$. An advantage of the LLS and RLLS methods are the closed-form solutions, allowing for simple implementation and easy error checking. However, in practice, the methods are not substantially accurate. Another limitation of the methods is the assumption of $L^{2}$ smoothness of the initial condition $x_{0}$. Given these limitations, we are motivated to develop methods for approximating the Fourier coefficients of $x_{0}$ accurately and in a stable manner. Not only that, but it is desired that the method be applicable to a class of solutions that are not necessarily smooth in the $L^{2}$ sense.

### 2.3 A Dual Method for approximating the Fourier expansion of the initial condition

In this section, we develop and analyze a new approach for estimating the initial condition of the abstract Cauchy problem (2.1.1) from time-series data. This method is similar to the LLS/RLLS methods presented in the previous section in that the approach is based on the approximation of the generalized Fourier coefficients. However, the method developed here involves an indirect computation of the generalized Fourier coefficients, based on the dual equation of the Cauchy problem. The dual equation may also be referred to as the adjoint equation. Given noisy data, the accuracy and stability of the method will be demonstrated in this section. The general framework of our method allows numerous PDE inverse problems to fit into this framework. Not only that, but it will be shown that the method can also be applied to the less ill-posed problem of forecasting future states of the system. Thus, our method may be especially beneficial for applications such as weather forecasting or financial futures, where it may be necessary to go both backward and forward.

Our approach for reconstructing $x_{0}$ is based on the dual (adjoint) equation

$$
\begin{equation*}
-\frac{d p}{d t}(t)=A^{*} p(t)+C^{*} u(t) \tag{2.3.1}
\end{equation*}
$$

where $C^{*} \in \mathcal{L}(Y, X)$ corresponds to the adjoint of the observation operator $C$, and, likewise, $A^{*}$ is the adjoint of the generator $A$. Here, $u \in L^{2}\left(0, t_{f} ; Y\right)$ denotes a control or input to the system. The scope of this method is not to control the dynamics in the typical sense. It will be demonstrated that a suitable control can be determined for which the generalized Fourier coefficients can be approximated by the control, $u$, and the data, $y$. This method is closely related to the Kalman-Bucy filter and Luenberger observer, which will be discussed further in Section 2.5.

Now, we proceed to derive the method. Recall the state equation is given by

$$
\begin{equation*}
\frac{d x}{d t}(t)=A x(t)+f(t) \tag{2.3.2}
\end{equation*}
$$

and the measurements

$$
\begin{equation*}
y(t)=C x(t), \quad 0 \leq t \leq t_{f} \tag{2.3.3}
\end{equation*}
$$

are given, for a known source $f$. Multiplying (2.3.2) by $p$, (2.3.1) by $x$, and integrating over $\left(0, t_{f}\right)$ yields

$$
\begin{equation*}
\int_{0}^{t_{f}} \frac{d}{d t}\langle x(t), p(t)\rangle d t=\int_{0}^{t_{f}}\left(\langle A x, p\rangle_{X}-\left\langle A^{*} p, x\right\rangle_{X}-\left\langle C^{*} u, x\right\rangle_{X}+\langle f(t), p(t)\rangle_{X}\right) d t \tag{2.3.4}
\end{equation*}
$$

which implies that

$$
\begin{equation*}
\left\langle x\left(t_{f}\right), p\left(t_{f}\right)\right\rangle_{X}-\langle x(0), p(0)\rangle_{X}=\int_{0}^{t_{f}}\langle f(t), p(t)\rangle_{X}-(u(t), C x(t))_{Y} d t \tag{2.3.5}
\end{equation*}
$$

yielding the relation

$$
\begin{equation*}
\left\langle x\left(t_{f}\right), p\left(t_{f}\right)\right\rangle_{X}-\langle x(0), p(0)\rangle_{X}=\int_{0}^{t_{f}}(u(t), \xi(t)-y(t))_{Y} d t \tag{2.3.6}
\end{equation*}
$$

where

$$
\begin{equation*}
\xi(t)=\int_{0}^{t} C S_{t-s} f(s) d s \tag{2.3.7}
\end{equation*}
$$

The relationship (2.3.6) forms the foundation for our method.
The unique mild solutions of the abstract Cauchy problem and its dual, with conditions $x(0)=x_{0}, p\left(t_{f}\right)=p_{t_{f}}$, are respectively given by

$$
\begin{align*}
& x(t)=S_{t} x_{0}+\int_{0}^{t} S_{t-s} f(s) d s  \tag{2.3.8}\\
& p(t)=S_{t_{f}-t}^{*} p_{t_{f}}+\int_{t}^{t_{f}} S_{s-t}^{*} C^{*} u(s) d s . \tag{2.3.9}
\end{align*}
$$

The method developed here may be applied to forecasting a future state, $x\left(t_{f}\right)$, as well as reconstructing the initial condition, $x(0)$. We first describe the method for reconstructing the initial state $x_{0}$. We assume the controllability of the adjoint system (2.3.1), which is equivalent to the observability of (2.1.1). The pair $(A, C)$ is observable if for all $x \in X$

$$
\begin{equation*}
\int_{0}^{t_{f}}\left\|C S_{t} x\right\|^{2} d t \geq \gamma\|x\|^{2} \tag{2.3.10}
\end{equation*}
$$

for some $\gamma>0$. Note that by the observability assumption (2.3.10), the equation

$$
y=\mathcal{M} x
$$

admits a unique solution, $x^{*}=x_{0}$, for $y \in R(\mathcal{M})$, where $\mathcal{M}$ is the operator defined by (2.2.2). Furthermore, this unique solution depends continuously on $y$ (see [10]). Having assumed the controllability of (2.3.1), we define the operator $\mathscr{L}: L^{2}\left(0, t_{f} ; Y\right) \rightarrow X$ by

$$
\mathscr{L} u:=\int_{0}^{t_{f}} S_{s}^{*} C^{*} u(s) d s
$$

and seek a $u$ satisfying

$$
\begin{equation*}
\mathscr{L} u=p(0) \tag{2.3.11}
\end{equation*}
$$

which means that $p(0) \in R(\mathscr{L})$. Again, by the controllability/observability assumption, we know a unique solution to (2.3.11), $u$, exists for $p(0) \in R(\mathscr{L})$. However, the exact controllability of (2.3.1) is, in general, not true, so we assume the condition (2.3.11) holds approximately, i.e., there exists $u_{\varepsilon}$ such that

$$
\begin{equation*}
\left\|\mathscr{L} u_{\varepsilon}-p(0)\right\|_{X} \leq \varepsilon \tag{2.3.12}
\end{equation*}
$$

for any $\varepsilon>0$.
Now, we proceed by defining a collection of adjoint systems $p_{k}(0)=\varphi_{k}$, such that $\left\{\varphi_{k}\right\}_{k=0}^{\infty}$ forms an orthonormal basis for $X$. Then $\left\langle x(0), \varphi_{k}\right\rangle$ are the generalized Fourier coefficients for $x(0)$. By the controllability assumption (2.3.10) and by utilizing relation (2.3.6), we can determine the Fourier coefficients of $x(0)$ by solving the operator equations

$$
\begin{equation*}
\mathscr{L} u_{k}=\varphi_{k}, \quad 0 \leq k \leq m \tag{2.3.13}
\end{equation*}
$$

for some $m<\infty$. If (2.3.12) holds, we will construct stable approximations $u_{k}$ using a suitable regularization method. An example of such a regularization method (for onedimensional $u$ ) for determining $u_{k}$ is

$$
\min _{u \in L^{2}\left(0, t_{f} ; Y\right)}\|\mathscr{L} u-\varphi\|_{X}^{2}+\eta_{1} \int_{0}^{t_{f}}|u(t)| d t+\frac{\eta_{2}}{2} \int_{0}^{t_{f}}\left|u^{\prime}(t)\right|^{2} d t
$$

where the first term corresponds to the sparsity of the approximate solution $u_{k}(t), t \in$ $\left[0, t_{f}\right]$, while the second term corresponds to the smoothness of $u_{k}$. We note that the smoothness of $u_{k}$ may affect noise dampening (see Remark 2.3.2). Such regularization methods are described in detail in Chapter 4, along with criteria for selecting the regularization parameters $\eta_{1}, \eta_{2}$.

Our approach is based on the fact that for each basis function $\varphi_{k}$ there exists a control $u_{k} \in L^{2}\left(0, t_{f} ; Y\right)$ such that $\mathscr{L} u_{k}=\varphi_{k}\left(\right.$ or $\left.\left\|\mathscr{L} u_{\varepsilon}-p(0)\right\|_{X} \leq \varepsilon\right)$. The controls $u_{k}(t)$ are determined in such a way that each adjoint $p_{k}$ is driven from zero at time $t_{f}$ to $p_{k}(0)=\varphi_{k}$. With the $u_{k}(t)$ determined, we construct the approximation for $x_{0}$ by

$$
x_{0}^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k}
$$

where the generalized Fourier coefficients are approximated by

$$
\begin{equation*}
\left\langle x_{0}, \varphi_{k}\right\rangle_{X} \approx \int_{0}^{t_{f}}\left(u_{k}(t), y(t)-\xi(t)\right)_{Y} d t=\alpha_{k} \tag{2.3.14}
\end{equation*}
$$

using the relation (2.3.6) and condition (2.3.13).
Further analysis of the method is detailed below, including the error analysis in Theorem 2.3.1. The following summarizes the method for estimating $x_{0}$.

## Dual Method for reconstruction of $x_{0}$ :

1. Pick an orthonormal basis, $\left\{\varphi_{k}\right\}_{k=0}^{m}$ for $X_{m} \subset X$
2. For each $k$ solve $\mathscr{L} u_{k}=\varphi_{k}$ for $u_{k} \in L^{2}\left(0, t_{f} ; Y\right)$
3. Form the estimate for $x_{0}$,

$$
x_{0}^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k}
$$

where

$$
\alpha_{k}=\int_{0}^{t_{f}}\left(u_{k}(t), y(t)-\xi(t)\right)_{Y} d t
$$

with $\xi$ defined by (2.3.7)

We note that the algorithm is well-posed under the exact controllability of the dual control system, i.e. there exists $\gamma>0$ such that

$$
\begin{equation*}
\int_{0}^{t_{f}}\left\|C S_{t} x\right\|^{2} d t \geq \gamma\|x\|_{X}^{2} \tag{2.3.15}
\end{equation*}
$$

for all $x \in X$.

## Using the method for forecasting a future state

Now, we briefly introduce how the method is utilized for the purpose of forecasting a future state $x\left(t_{f}\right)$. For this purpose, we assume the adjoint (2.3.1) is null-controllable so that there exists $u \in L^{2}\left(0, t_{f} ; Y\right)$ such that $p(0)=0$ and

$$
\begin{equation*}
\mathscr{L} u=-S_{t_{f}} p\left(t_{f}\right) . \tag{2.3.16}
\end{equation*}
$$

In general, the exact null-controllability may not hold, however we assume the condition (2.3.16) holds approximately, i.e., there exists $u_{\varepsilon}$ such that

$$
\left\|\mathscr{L} u_{\varepsilon}+S_{t_{f}} p\left(t_{f}\right)\right\|_{X} \leq \varepsilon
$$

for any $\varepsilon>0$. With $u_{k}$ determined, the generalized Fourier coefficients are approximated by

$$
\left\langle x\left(t_{f}\right), \varphi_{k}\right\rangle_{X}=-\int_{0}^{t_{f}}\left(u_{k}(t), y(t)-\xi(t)\right)_{Y} d t
$$

where $u_{k}$ is the approximate solution to

$$
\mathscr{L} u_{k}=-S_{t_{f}} \varphi_{k}
$$

For the final state case, the method is well-posed under the assumption of null-controllability of the adjoint control system, i.e.

$$
S_{t_{f}}^{*} X \subseteq R(\mathscr{L})
$$

The method is summarized as follows:

## Dual Method for reconstruction of $x_{t_{f}}$ :

1. Pick an orthonormal basis, $\left\{\varphi_{k}\right\}_{k=0}^{m}$ for $X_{m} \subset X$
2. For each $k$ solve $\mathscr{L} u_{k}=-S_{t_{f}} \varphi_{k}$ for $u_{k} \in L^{2}\left(0, t_{f} ; Y\right)$
3. Form the estimate for $x_{t_{f}}$,

$$
x_{t_{f}}^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k}
$$

where

$$
\alpha_{k}=\int_{0}^{t_{f}}\left(u_{k}(t), y(t)\right)_{Y} d t
$$

or in general with force $f$

$$
\alpha_{k}=\int_{0}^{t_{f}}\left(u_{k}(t), \xi(t)-y(t)\right)_{Y} d t
$$

with

$$
\xi(t)=\int_{0}^{t} C S_{t-s} f(s) d s
$$

The novelty of this method is, in part, due to the fact that it is not necessary to compute the time history of the adjoint, $p$. However, the method utilizes the information available from the adjoint in order to accurately reconstruct $x_{0}$. It will also be demonstrated that we can construct sparse controls, $u$, yielding storage reduction without sacrificing the accuracy of the reconstruction. This aspect is explored in Chapters 4, 6 , where we develop methods for efficiently solving (2.3.13) and discuss numerical results, respectively.

Remark 2.3.1. We also note that there is a stochastic/probabilistic interpretation of this method. Assume $x, p$ are random variables satisfying the linear stochastic differential equations

$$
\begin{align*}
d x & =(A x(t)+f(t)) d t+\sigma d B_{t}  \tag{2.3.17}\\
-d p & =\left(A^{*} p(t)+C^{*} u(t)\right) d t \tag{2.3.18}
\end{align*}
$$

where $B_{t}$ is the Brownian motion, and $\sigma$ is the standard deviation (diffusion coefficient). Then, by the relation (2.3.6) we have

$$
\left\langle x_{0}, \varphi_{k}\right\rangle_{X}=\int_{0}^{t_{f}}\left(u_{k}(t), y(t)\right)_{Y} d t-\int_{0}^{t_{f}}\left\langle f(t), p_{k}(t)\right\rangle_{X} d t+\sigma \int_{0}^{t_{f}} p_{k}(t) d B_{t}
$$

which implies that

$$
E\left[\left|\left\langle x_{0}, \varphi_{k}\right\rangle_{X}-\int_{0}^{t_{f}}\left(u_{k}(t), y(t)-\xi(t)\right)_{Y} d t\right|^{2}\right]=E\left[\sigma^{2}\left|\int_{0}^{t_{f}} p_{k}(t) d t\right|^{2}\right]
$$

Thus, the mean square error in approximating the Fourier coefficients is related to the standard deviation, $\sigma$, of the Brownian motion, regardless of that fact that $p\left(t_{f}\right)=0$ (in the case of estimating $x_{0}$ ). Determining the control, $u_{k}$, can be cast as

$$
\min _{u \in L^{2}\left(0, t_{f} ; Y\right)}\left\|\mathscr{L} u-\varphi_{k}\right\|_{X}^{2}+\beta \sigma^{2} \int_{0}^{t_{f}}|p(t)|^{2} d t \quad 0 \leq k \leq m
$$

where

$$
p(t)=\int_{t}^{t_{f}} S_{s-t}^{*} C^{*} u(s) d s
$$

Thus, we select the parameter $\beta$ so that $\varepsilon^{2}+\beta \sigma^{2}$ is balanced, where $\varepsilon$ is the accuracy of the fidelity term

$$
\mathscr{L} u_{k}-\varphi_{k}=\varepsilon .
$$

The following theorem provides the error estimate of our reconstruction method in the real Hilbert space setting, as well as justification for the method based on regularization. In short, there are two sources of error in approximating the Fourier coefficients. The first source of error is due to solving the equation $\mathscr{L} u_{k}=\varphi_{k}$, while the second source of error is due to the noise, $\delta$, in the observed data. The errors must be balanced to obtain the best possible solution.

Theorem 2.3.1 (Error Estimate). Suppose $\left(A^{*}, C^{*}\right)$ is approximately controllable, there exists $u_{k} \in L^{2}\left(0, t_{f} ; Y\right)$ such that

$$
\left\|\mathscr{L} u_{k}-\varphi_{k}\right\|_{X} \leq \varepsilon_{k}
$$

for each $0 \leq k \leq m$. If we define,

$$
v(t)=y^{\delta}(t)-y(t)
$$

and

$$
\|v(t)\| \leq \delta
$$

then there is a constant $c\left(\delta, t_{f}\right)$ such that

$$
\left\|x_{0}-x_{\delta}^{m}\right\|_{X} \leq\left\|x_{0}-x^{m}\right\|_{X}+\sum_{k=0}^{m}\left(\varepsilon_{k}\left\|x_{0}\right\|_{X}+c\left(\delta, t_{f}\right)\left\|u_{k}(t)\right\|_{Z}\right)
$$

where $Z=L^{2}\left(0, t_{f} ; Y\right)$ and

$$
\left\|x_{0}-x^{m}\right\|_{X}
$$

is the truncation error of the generalized Fourier series. Furthermore, if $x_{0} \in C^{k}(\Omega)$, then

$$
\begin{equation*}
\left\|x_{0}-x_{\delta}^{m}\right\|_{X} \leq \sum_{k=0}^{m}\left(\varepsilon_{k}\left\|x_{0}\right\|_{X}+c\left(\delta, t_{f}\right)\left\|u_{k}(t)\right\|_{Z}\right) \tag{2.3.19}
\end{equation*}
$$

for $m$ sufficiently large.
Proof. By the orhonormality of $\left\{\varphi_{k}\right\}_{k=0}^{m}$ and the Cauchy-Schwarz inequality

$$
\begin{aligned}
\left\|x_{0}-x_{\delta}^{m}\right\|_{X} & \leq\left\|x_{0}-x^{m}\right\|_{X}+\left\|x^{m}-x_{\delta}^{m}\right\|_{X} \\
& \leq\left\|x_{0}-x^{m}\right\|_{X}+\left\|\sum_{k=0}^{m}\left(\left\langle x_{0}, \varphi_{k}\right\rangle_{X}-\left\langle u_{k}(t), y^{\delta}(t)\right\rangle_{Z}\right) \varphi_{k}\right\|_{X} \\
& \leq\left\|x_{0}-x^{m}\right\|_{X}+\left\|\sum_{k=0}^{m}\left(\left\langle\mathscr{L} u_{k}-\varphi_{k}, x_{0}\right\rangle_{X}+\left\langle u_{k}(t), v(t)\right\rangle_{Z}\right) \varphi_{k}\right\|_{X}
\end{aligned}
$$

from which the estimate follows. The estimate (2.3.19) follows from the standard Fourier series analysis.

The error estimate is desired to be independent of the initial condition, $x_{0}$, however this is not realistic since $x_{0}$ is the unknown. It should be noted, that the error estimate is the worst possible error obtained in the estimation. Better error estimates may be realized, however, the results of the Theorem also provide justification for the regularization
methods. By the estimate,

$$
\left\|x_{0}-x_{\delta}^{m}\right\|_{X} \leq\left\|x_{0}-x^{m}\right\|_{X}+\left\|\sum_{k=0}^{m}\left(\left\langle\mathscr{L} u_{k}-\varphi_{k}, x_{0}\right\rangle_{X}+\left\langle u_{k}(t), v(t)\right\rangle_{Z}\right) \varphi_{k}\right\|_{X}
$$

we immediately see the need for appropriately solving $u_{k}$. If the noise level, $\delta$, is large we must obtain controls which are sufficiently regular, so that the term

$$
\left\langle u_{k}(t), v(t)\right\rangle_{Z}
$$

is small, while simultaneously ensuring $\left\|\mathscr{L} u_{k}-\varphi_{k}\right\|_{X}$ is small. The following remark further justifies imposing regularity on $u_{k}$.

Remark 2.3.2. Suppose the noise in the data is highly oscillatory, such as $\cos (l \pi t)$. Then the error in the Fourier coefficients has the term

$$
\begin{equation*}
\int_{0}^{1} u_{k}(t) \cos (l \pi t) d t=\frac{1}{l \pi} \int_{0}^{1} u_{k}^{\prime}(t) \sin (l \pi t) d t \tag{2.3.20}
\end{equation*}
$$

so that highly oscillatory parts may be damped by $l \pi$, if $u_{k}$ is sufficiently smooth. This provides further justification for the use of a penalty term in the cost functional which enforces smoothness on the control $u_{k}$.

It is also apparent that the accuracy, $\varepsilon_{k}$, in solving

$$
\mathscr{L} u_{k}=\varphi_{k}
$$

is necessary for an accurate reconstruction of $x_{0}$. In practice, we must balance the accuracy of solving $\mathscr{L} u_{k}=\varphi_{k}$ and the regularity imposed on $u_{k}$ via the regularization methods. This concern is addressed in Chapter 4 where we discuss how to balance the method to obtain stable but accurate solutions.

### 2.4 Variation of the Dual Control Method

In this section, we outline an alternate procedure for obtaining reconstructions of the initial condition, $x_{0}$. This approach is based on the dual control approach developed
in the previous section. Rather than selecting a collection $\left\{p_{k}(0)\right\}_{k=0}^{\infty}$ to be a basis for $X$, we select $\left\{u_{k}(t)\right\}_{k=0}^{\infty}$ to be a basis(not necessarily orthonormal) for $Z=L^{2}\left(0, t_{f} ; Y\right)$. Assuming the relation (2.3.11) holds, we construct the adjoint set $\left\{\tilde{p}_{k}\right\}$ by the relations

$$
\mathscr{L} u_{k}=\tilde{p}_{k} .
$$

Note that the collection $\left\{\tilde{p}_{k}\right\}_{k=0}^{\infty}$ is linearly independent under the assumption that $(A, C)$ is controllable, i.e.,

$$
R(\mathscr{L})=X \Longrightarrow N(\mathscr{L})=\emptyset
$$

Thus, if $(A, C)$ is exactly controllable, we form an orthogonal(orthonormal) basis by the Gram-Schmidt method. The coefficients of $x_{0}$ are computed by defining the Gram matrix

$$
G_{k, l}=\left\langle\tilde{p}_{k}, \tilde{p}_{l}\right\rangle_{X}
$$

and setting $\boldsymbol{\beta}=\left(\beta_{0}, \ldots, \beta_{m}\right)^{t}$ such that

$$
\boldsymbol{\beta}=G^{-1}\left(\begin{array}{c}
\int_{0}^{t_{f}}\left(u_{0}, y^{\delta}\right) d t \\
\vdots \\
\int_{0}^{t_{f}}\left(u_{m}, y^{\delta}\right) d t
\end{array}\right)
$$

The coefficients $\boldsymbol{\beta}$ can be computed efficiently by the Cholesky decomposition $G=L L^{*}$, since $G$ is symmetric positive definite. Again, the algorithm is well-posed under the exact controllability (2.3.15) of the adjoint system which, in general, may not be true. If the adjoint system is not exactly controllable, care must be exercised to ensure the set $\left\{\mathscr{L} u_{k}\right\}_{k=0}^{m}$ is linearly independent.

## Variation of Dual Control Algorithm:

1. Pick a basis $\left\{u_{k}(t)\right\}_{k=0}^{m}$ for $U_{m} \subset L^{2}\left(0, t_{f} ; Y\right)$
2. Compute $\tilde{p}_{k}$ by $\mathscr{L} u_{k}=\tilde{p}_{k}$
3. Compute the Gram matrix $G_{k, l}=\left\langle\tilde{p}_{k}, \tilde{p}_{l}\right\rangle_{X}$
4. Set $\boldsymbol{y}_{k}=\int_{0}^{t_{f}}\left(u_{k}(t), \xi(t)-y(t)\right)_{Y} d t$ and compute the approximate Fourier coefficients $\boldsymbol{\beta}=G^{-1} \boldsymbol{y}$
5. Compute the approximation

$$
x_{0}^{m}=\sum_{k=0}^{m} \beta_{k} \mathscr{L} u_{k}
$$

There are several potential advantages to this approach. Namely, one can directly regulate the properties of the controls $u_{k}$, such as smoothness or sparsity. Secondly, the operator $\mathscr{L}$ does not need to be inverted. However, since the pair $\left(A^{*}, C^{*}\right)$ is not necessarily controllable, we are not guaranteed linear independence of the set $\left\{\tilde{p}_{k}\right\}_{k=0}^{m}$. Thus, solving

$$
G \boldsymbol{\beta}=\left(\begin{array}{c}
\int_{0}^{t_{f}}\left(u_{0}, y^{\delta}\right) d t  \tag{2.4.1}\\
\vdots \\
\int_{0}^{t_{f}}\left(u_{m}, y^{\delta}\right) d t
\end{array}\right)
$$

for $\boldsymbol{\beta}$ requires regularization. This method only requires the solution of one ill-posed problem, but requires the formation of the $m+1$ adjoints $p_{k}$. Therefore, this method may be less expensive than the dual control method.

### 2.5 Kalman-Bucy and Time-Reversal methods

In this section, we describe an iterative approximation method based on the KalmanBucy filter and Luenberger type observers. The method is based on tracking the state of a system in order to obtain a reasonable estimate at some time. Using a time-reversal process, the method uses the observed estimate as the initial condition for integrating the system backwards in time. In its original context, this method assumes the operator $A$ is skew-adjoint.

In light of the linear system (2.3.2), we consider the problem of estimating the state which is described by the process

$$
\begin{equation*}
d x=A x(t) d t+\sigma d B_{t} \tag{2.5.1}
\end{equation*}
$$

given measurements

$$
\begin{equation*}
d y=C y(t) d t+d v_{t} \tag{2.5.2}
\end{equation*}
$$

We assume $B_{t}, v_{t}$ are independent Brownian motions, with covariances $Q, R$, respectively, that is

$$
\begin{aligned}
E[v(t)] & =0, & E\left[v(t) v^{t}(\tau)\right]=R \delta(t-\tau) \\
E\left[B_{t}\right] & =0, & E\left[B_{t} B_{\tau}\right]=Q \delta(t-\tau) .
\end{aligned}
$$

The Kalman-Bucy filter is defined by

$$
\begin{align*}
& d \hat{x}(t)=A \hat{x}(t) d t+L(t)(y-\hat{y}) d t  \tag{2.5.3}\\
& \hat{y}(t)=C \hat{x}(t) \tag{2.5.4}
\end{align*}
$$

where $L(t)$ is the Kalman filter gain and $\hat{x}$ is an estimate of $x$. This filter is based on the provided mean $\mu_{\hat{x}}(0)$ and the covariance of $\hat{x}(0)$. That is, we have the provided initial conditions

$$
\begin{array}{r}
E[\hat{x}(0)]=\mu_{\hat{x}}(0), \\
E\left[\left(\hat{x}(0)-\mu_{\hat{x}}(0)\right)\left(\hat{x}(0)-\mu_{\hat{x}}(0)\right)^{t}\right]=\Sigma(0) \tag{2.5.6}
\end{array}
$$

The Kalman filter gain is determined by

$$
L(t)=\Sigma(t) C^{*} R^{-1}
$$

where $\Sigma$ is the solution of the differential Riccati equation

$$
\begin{align*}
\frac{d}{d t} \Sigma(t) & =A^{*} \Sigma(t)+\Sigma(t) A-\Sigma(t) C^{*} R^{-1} C \Sigma(t)+\sigma Q \sigma^{t}  \tag{2.5.7}\\
\Sigma(0) & =E\left[\left(x_{0}-E\left[x_{0}\right]\right)^{2}\right]=\operatorname{cov}\left(x_{0}\right) \tag{2.5.8}
\end{align*}
$$

The state estimate $\hat{x}(t)$ provides the maximum likelihood of the state $x(t)$, given observations $y(s), 0 \leq s \leq t_{f}$. The covariance, $\Sigma$, represents the uncertainty in the estimation due to the noise.

It follows from (2.5.1) and (2.5.3) that the error dynamics for $e(t)=x(t)-\hat{x}(t)$ are
given by

$$
\begin{equation*}
d e(t)=(A-L(t) C) e(t) d t-L(t) v(t) d t+\sigma d B_{t} \tag{2.5.9}
\end{equation*}
$$

and the mean error is governed by

$$
\begin{equation*}
d E[e(t)]=(A-L(t) C) E[e(t)] d t \tag{2.5.10}
\end{equation*}
$$

In practice, we often take a stationary gain $L$ so that $A-L C$ generates an exponentially stable semigroup on $X$. This corresponds to driving the mean error (2.5.10) to zero. The covariance, $\Sigma(t)$, represents the error variance $E[(x(t)-\hat{x}(t))(x(t)-\hat{x}(t))]$ of the estimate $\hat{x}$. Whenever the pair $(A, C)$ is observable, we can always determine a filter gain, $L$, such that $A-L C$ generates an exponentially stable semigroup [10].

Now, we describe the time-reversal technique for the case when $A$ is skew-adjoint, i.e. $A^{*}=-A$. We assume the pair $(A, C)$ is exactly observable. Since $A$ is skew-adjoint, we have stable dynamics both forward and backward in time for

$$
\begin{equation*}
\frac{d x}{d t}(t)=A x(t) \tag{2.5.11}
\end{equation*}
$$

The time-reversal technique works by utilizing the Kalman-Bucy filter to form an estimate, $\hat{x}\left(t_{f}\right)$ at the terminal time $t_{f}$. Then, a backwards filter forms the estimate of the initial state, $\hat{x}^{b}(0)$ by integrating (2.5.11) backwards in time. That is, the backward filter is given by

$$
\begin{array}{r}
\frac{d \hat{x}^{b}}{d t}=A \hat{x}^{b}+L\left(y(t)-C \hat{x}^{b}\right) \\
\hat{x}^{b}\left(t_{f}\right)=\hat{x}\left(t_{f}\right) .
\end{array}
$$

where $\hat{x}\left(t_{f}\right)$ is the terminal estimate of the state from the forward Kalman-Bucy filter. Equivalently, the backwards filter is formulated as

$$
\begin{equation*}
-\frac{d \hat{x}}{d t}\left(t_{f}-t\right)=A \hat{x}\left(t_{f}-t\right)+L\left(y\left(t_{f}-t\right)-C \hat{x}\left(t_{f}-t\right)\right) \tag{2.5.12}
\end{equation*}
$$

given $\hat{x}\left(t_{f}\right)$, yielding the estimate, $\hat{x}(0)$, of the initial condition. Then, we forward inte-
grate

$$
\frac{d \hat{x}}{d t}(t)=A \hat{x}(t)
$$

again to obtain a second estimate of $\hat{x}\left(t_{f}\right)$, and repeat the backwards filter (2.5.12). Recursively, this can be cast as

$$
\left\{\begin{array} { l } 
{ \frac { d \hat { x } _ { k } } { d t } = A \hat { x } _ { k } + L ( y - C \hat { x } _ { k } ) , } \\
{ \hat { x } _ { k } ( 0 ) = \hat { x } _ { k - 1 } ^ { b } ( 0 ) , }
\end{array} \quad \left\{\begin{array}{l}
\frac{d \hat{x}_{k}^{b}}{d t}=A \hat{x}_{k}^{b}+L\left(y-C \hat{x}_{k}^{b}\right), \\
\hat{x}_{k}^{b}\left(t_{f}\right)=\hat{x}_{k}\left(t_{f}\right),
\end{array}\right.\right.
$$

setting $\hat{x}_{-1}^{b}=\hat{x}_{0}$. When $A$ is skew-adjoint, this method is of particular interest due to the fact that minimization/optimization is not required. When $A$ is not skew-adjoint, the forward step of this method (Kalman-Bucy filter) can be utilized to obtain a reasonable estimate of the final state, which can be used as the data in the dual control method to estimate the initial condition. A similar iterative forward and backward algorithm can be setup which uses the dual control method for the backward steps, and filtering for the forward steps.

For this particular method, we use the stationary filter gain $L=\gamma C^{*}$ for a chosen gain coefficient, $\gamma$, so that $A-\gamma C^{*} C$ generates an exponentially stable semigroup. That is, the state and state estimate dynamics are given by

$$
\begin{align*}
d x(t) & =A x(t) d t+\sigma d B_{t}  \tag{2.5.13}\\
d \hat{x}(t) & =A \hat{x}(t) d t+\gamma C^{*}(y-C \hat{x}) d t \tag{2.5.14}
\end{align*}
$$

and the filters are given by

$$
\left\{\begin{array} { l } 
{ \frac { d \hat { x } _ { k } } { d t } = A \hat { x } _ { k } + \gamma C ^ { * } ( y - C \hat { x } _ { k } ) , } \\
{ \hat { x } _ { k } ( 0 ) = \hat { x } _ { k - 1 } ^ { b } ( 0 ) , }
\end{array} \quad \left\{\begin{array}{l}
\frac{d \hat{x}_{k}^{b}}{d t}=A \hat{x}_{k}^{b}+\gamma C^{*}\left(y-C \hat{x}_{k}^{b}\right), \\
\hat{x}_{k}^{b}\left(t_{f}\right)=\hat{x}_{k}\left(t_{f}\right) .
\end{array}\right.\right.
$$

The error dynamics are then given by

$$
\begin{equation*}
d e(t)=\left(A-\gamma C^{*} C\right) e(t) d t-\gamma C^{*} v(t) d t+\sigma d B_{t} \tag{2.5.15}
\end{equation*}
$$

for the stochastic system (2.5.1)-(2.5.2). An iterative time-reversal method of this form is formalized and analyzed in [30]. The advantage of this type of method is that optimization is not necessary. However, without modification this algorithm is limited to
skew-adjoint operators $A$ (e.g., the wave equation).

### 2.6 Implementation Issues

In this section, we discuss the necessary numerical issues for the implementation of the methods developed in this chapter. For the numerical implementation for solving the dual control problem we use the Crank-Nicholson scheme

$$
\begin{equation*}
-\frac{p^{k+1}-p^{k}}{\Delta t}=A^{*} \frac{p^{k+1}+p^{k}}{2}+C^{*} u_{k+\frac{1}{2}} \tag{2.6.1}
\end{equation*}
$$

for (2.3.1) where $u_{k+1 / 2}$ is evaluated at the mid-point of the interval $\left[t_{k}, t_{k+1}\right]$ and $t_{k}=$ $k t_{f} \Delta t$. At the time step $k+1$ the solution is computed by

$$
\begin{equation*}
p^{k+1}=-r_{1,1}\left(A^{*} \Delta t\right) p^{k}-\Delta t\left(I-\frac{\Delta t}{2} A^{*}\right)^{-1} C^{*} u_{k+\frac{1}{2}} \tag{2.6.2}
\end{equation*}
$$

where

$$
\begin{equation*}
r_{1,1}\left(A^{*} \Delta t\right)=\left(2-A^{*} \Delta t\right)^{-1}\left(2+A^{*} \Delta t\right)=\left(I-\frac{\Delta t}{2} A^{*}\right)^{-1}\left(I+\frac{\Delta t}{2} A^{*}\right) \tag{2.6.3}
\end{equation*}
$$

is the $(1,1)$ Padé approximation for $\exp \left(A^{*} \Delta t\right)$. The discretized RLLS inverse problem is then formulated as

$$
\begin{equation*}
\min _{x \in X_{m}}\left\|y-M^{n} x\right\|_{Y}+\beta\|x\|_{X} \tag{2.6.4}
\end{equation*}
$$

where we have defined the controllability(observability) matrix

$$
\begin{equation*}
M^{n}=\left[C, C r_{1,1}(A \Delta t), \ldots, C r_{1,1}(A \Delta t)^{n-1}\right] . \tag{2.6.5}
\end{equation*}
$$

and $X_{m}$ is a finite-dimensional subspace of $X$. In the dual control formulation, the discretized problem for each control $u$ is formulated as

$$
\min _{u \in U_{m}}\left\|L^{n} u-\varphi\right\|_{Y}+\beta \psi(u)
$$

where $L^{n}=\left(M^{n}\right)^{*}, U_{m}$ is a finite-dimensional subspace of $L^{2}\left(0, t_{f} ; Y\right)$ and $\psi$ is a chosen penalty term.

If necessary, higher order Padé approximations may be considered, which are of the form

$$
\begin{equation*}
r_{m, n}(z)=\frac{P_{m}}{Q_{n}}(z)=\frac{a_{0}+a_{1} z+\ldots+a_{m} z^{m}}{b_{0}+b_{1} z+\ldots+b_{n} z^{n}} \tag{2.6.6}
\end{equation*}
$$

where the degree of $P, Q$ is not more than $m, n$ respectively. Higher order Padé approximations of semigroups are discussed in detail in the paper [36].

## Operator Splitting for the Convection-Diffusion Equation

The Crank-Nicholson scheme works well for the diffusion dominant case, however, for the convection dominant case it is necessary to solve the problem more accurately. In this section, we describe the numerics for the initial condition estimation of the convectiondiffusion equation

$$
\begin{align*}
& \frac{\partial v}{\partial t}(x, t)=c(x) \cdot \nabla v(x, t)+\nabla \cdot(d(x) \nabla v)(x, t)  \tag{2.6.7}\\
& v(x, 0)=v_{0}(x) \tag{2.6.8}
\end{align*}
$$

where $c(x), d(x)$ are the convection and diffusion coefficients, respectively.
The reconstruction methods have a natural extension to such problems, using a differential operator splitting

$$
\frac{\partial v}{\partial t}=L v(t)=(A+B) v(t)
$$

where $A, B \in \mathcal{L}(X)$
For the numerical solution of the convection-diffusion equation, we consider the two stage Strang operator splitting

$$
v(x, t+\Delta t)=S_{\frac{\Delta t}{2}}^{h} S_{\Delta t}^{p} S_{\frac{\Delta t}{2}}^{h} v(x, t)
$$

where $S_{t}^{p}, S_{t}^{h}$ are the semigroups corresponding to the parabolic and hyperbolic subproblems, respectively. That is, $S_{t}^{p}, S_{t}^{h}$ are the $C_{0}$-semigroup semigroups generated by $A, B$ respectively.

Assuming a constant convection coefficient $c$, we solve the hyperbolic subproblem via the method of characteristics $v\left(t_{n+1}, x\right)=v\left(t_{n}, x-c \Delta t\right)$ where the right hand side is evaluated via cubic interpolation. As in the previous section, we use the Crank-Nicholson
method for solving the parabolic subproblem, using the approximating polynomial

$$
r_{1,1}(z)=\frac{2+z}{2-z}
$$

for the approximation of $\exp (A \Delta t)$.

## Discrete Kalman Filter

We now briefly describe the discrete Kalman filter in its predictor corrector form, which is useful for the approximation of continuous dynamics in time. In our framework, we define $S=r_{1,1}(A \Delta t)$. The discrete dynamics are described by

$$
\begin{equation*}
x_{k}=S x_{k-1}+w_{k-1} \tag{2.6.9}
\end{equation*}
$$

given measurements

$$
\begin{equation*}
y_{k}=C_{k} x_{k}+v_{k} . \tag{2.6.10}
\end{equation*}
$$

We assume the noises $w_{k}, v_{k}$ are zero-mean with covariances $Q_{k-1}, R_{k}$, respectively, i.e.

$$
\begin{align*}
w_{k} & \approx \mathcal{N}\left(0, Q_{k-1}\right)  \tag{2.6.11}\\
v_{k} & \approx \mathcal{N}\left(0, R_{k}\right) \tag{2.6.12}
\end{align*}
$$

We assume the probability density function (pdf) of the initial state, $f_{x_{0}}$, is known. Our goal is to construct the posterior pdf, $f_{x_{k} \mid y_{k}}$, given the known prior pdf $f_{x_{k-1} \mid y_{k-1}}$. Since the dynamics are known, we can update the state by

$$
\begin{equation*}
x_{k,-}^{*}=S x_{k-1}^{*} \tag{2.6.13}
\end{equation*}
$$

and update the error covariance by

$$
\begin{equation*}
V_{k,-}=S V_{k-1} S^{t}+Q_{k-1} \tag{2.6.14}
\end{equation*}
$$

The linear Kalman filter can be summarized by the two steps:

## Prediction:

$$
\begin{align*}
x_{k,-}^{*} & =S x_{k-1}^{*}  \tag{2.6.15}\\
V_{k,-} & =S V_{k-1} S^{t}+Q_{k-1} \tag{2.6.16}
\end{align*}
$$

## Correction:

$$
\begin{align*}
K_{k} & =V_{k,-} C_{k}^{t}\left(C_{k} V_{k,-} C_{k}^{t}+R_{k}\right)^{-1}  \tag{2.6.17}\\
x_{k}^{*} & =x_{k,-}^{*}+K_{k}\left(y_{k}-C_{k} x_{k,-}^{*}\right)  \tag{2.6.18}\\
V_{k} & =\left(I-K_{k} C_{k}\right) V_{k,-} \tag{2.6.19}
\end{align*}
$$

This formulation is useful for the implementation of the time-reversal technique.

## Chapter 3

## Extensions of the reconstruction algorithms

In this section, we propose two extensions of the methods presented in Chapter 2. First, we formulate the simultaneous parameter identification/initial state estimation problem. As an example, a parameter such as the diffusion coefficient or convection coefficient may be unknown. Secondly, the methods are extended to identifying the unknown source, $f$, rather than the initial condition $x_{0}$.

### 3.1 Simultaneous State and Parameter Estimation

In this brief section, we provide an extension of the algorithms provided in Chapter 2 to simultaneously estimating the initial condition and parameter. The parameter-dependent abstract Cauchy problem is given by

$$
\begin{align*}
& \frac{d x}{d t}(t)=A(p) x(t)+f(t)  \tag{3.1.1}\\
& x(0)=x_{0} \tag{3.1.2}
\end{align*}
$$

based on the given measurements

$$
\begin{equation*}
y(t)=C x(t) \tag{3.1.3}
\end{equation*}
$$

where $A(p): \mathbb{R}^{m} \rightarrow \mathcal{L}(X, X)$ is a closed linear operator on the Hilbert space $X$, for each $p \in \mathcal{Q} \subset \mathbb{R}^{m}$. The set $\mathcal{Q}$ is the set of admissible parameters. As an example, a parameter dependent linear operator may be of the form

$$
A(p) x=\operatorname{div}(p(\xi) \nabla x)
$$

where $p(\xi)=\sum_{i=1}^{n} p_{i} \phi_{i}(\xi)$ and $\left\{\phi_{i}(\xi)\right\}_{i=1}^{n}$ is a basis for $X$, such as a spline basis. For instance, in groundwater filtration [13], $p$ represents hydraulic permittivity, while in impedance tomography, $p$ represents the conductivity. In the respective problems, the state $x$ represents the pressure of water and the voltage. The parameter identification problem consists of reconstructing the parameter $p$ from knowledge of the systems output. Thus, for the simultaneous state and parameter estimation, we must reconstruct the initial condition $x_{0}$ and determine the parameter(s) of the system.

Taking the regularized linear least-squares (RLLS) approach given in Section 2.2, for $p \in \mathcal{Q} \subset \mathbb{R}^{m}$, we define

$$
\begin{equation*}
\mathcal{J}_{\eta}(p)=\frac{1}{2}\left\|y-\mathcal{M}(p) x_{p}\right\|_{L^{2}(0, T ; Y)}^{2}+\frac{\eta}{2}\left\|x_{p}\right\|_{X}^{2} \tag{3.1.4}
\end{equation*}
$$

and seek

$$
\min _{p \in \mathcal{Q}} \mathcal{J}_{\eta}(p)
$$

where

$$
\begin{equation*}
x_{p}=\underset{x \in X}{\arg \min }\left\{\frac{1}{2}\|y-\mathcal{M}(p) x\|_{L^{2}\left(0, t_{f} ; Y\right)}^{2}+\frac{\eta}{2}\|x\|_{X}^{2}\right\} . \tag{3.1.5}
\end{equation*}
$$

Here, the parameter-dependent input-output map is defined by

$$
\mathcal{M}(p)=C S_{t}(p)
$$

where $S_{t}(p)$ is the $C_{0}$-semigroup generated by $A(p)$ on $X$. We assume $\operatorname{dom}(A(p))=$ $\operatorname{dom}(A)$ is independent of $p \in Q \subset \mathbb{R}^{m}$. Define

$$
\dot{A}(p) \psi:=\lim _{s \rightarrow 0} \frac{A(p+s h)-A(p)}{s} \psi
$$

in the direction $h \in \mathbb{R}^{m}$, for all $\psi \in \operatorname{dom}(A)$. Then,

$$
\begin{equation*}
\dot{\mathcal{M}}(p) \psi=\lim _{s \rightarrow 0} \frac{\mathcal{M}(p+s h) \psi-\mathcal{M}(p) \psi}{s}=C S_{t}(p)(\dot{A}(p)) h \psi \tag{3.1.6}
\end{equation*}
$$

for all $\psi \in \operatorname{dom}(A)$. The parametric solution of (3.1.5) is given by

$$
\begin{equation*}
x_{p}=\left(\mathcal{M}^{*}(p) \mathcal{M}(p)+\eta I\right)^{-1} \mathcal{M}^{*}(p) y \tag{3.1.7}
\end{equation*}
$$

We use the gradient like method to minimize the cost functional (3.1.4) over $p \in \mathcal{Q}$. The Gateaux derivative of the cost functional (3.1.4) is given by

$$
\frac{d}{d p} \mathcal{J}_{\eta}(p) h=\lim _{s \rightarrow 0} \frac{\mathcal{J}_{\eta}(p+s h)-\mathcal{J}_{\eta}(p)}{s}
$$

for the direction $h \in \mathbb{R}^{m}$. From (3.1.4), the Gateaux derivative can be computed by

$$
\begin{align*}
\frac{d}{d p} \mathcal{J}_{\eta}(p) & =\left\langle\dot{\mathcal{M}}(p) x_{p}, \mathcal{M}(p) x_{p}-y\right\rangle_{Y}+\left\langle\dot{x}_{p},\left(\mathcal{M}^{*}(p) \mathcal{M}(p)+\eta I\right) x_{p}-\mathcal{M}^{*}(p) y\right\rangle_{X}  \tag{3.1.8}\\
& =\left\langle\dot{\mathcal{M}}(p) x_{p}, \mathcal{M}(p) x_{p}-y\right\rangle_{Y} \tag{3.1.9}
\end{align*}
$$

since $x_{p}$ satisfies

$$
\left(\mathcal{M}^{*}(p) \mathcal{M}(p)+\eta I\right) x_{p}=\mathcal{M}^{*}(p) y
$$

From (3.1.6) we have

$$
\begin{equation*}
\frac{d}{d p} \mathcal{J}_{\eta}(p)=\left\langle C S_{t}(p) \dot{A}(p) x_{p}, \mathcal{M}(p) x_{p}-y\right\rangle_{Y} \tag{3.1.10}
\end{equation*}
$$

Given this approach, we can compute the minimizer $p^{\dagger}$ by the Gradient method iteration

$$
p_{n+1}=p_{n}-\gamma \frac{d}{d p} \mathcal{J}_{\eta}
$$

for a chosen initial guess $p_{0}$ and step size $\gamma$. If the second variation $\ddot{\mathcal{M}}(p)$ exists then

$$
\begin{equation*}
\frac{d^{2}}{d p^{2}} \mathcal{J}_{\eta}=\left\langle\ddot{\mathcal{M}}(p) x_{p}+\dot{\mathcal{M}}(p) \dot{x}_{p}, \mathcal{M}(p) x_{p}-y\right\rangle_{Y}+\left\langle\dot{\mathcal{M}}(p) x_{p}, \dot{\mathcal{M}}(p) x_{p}+\mathcal{M}(p) \dot{x}_{p}\right\rangle_{Y} \tag{3.1.11}
\end{equation*}
$$

To compute $\dot{x}_{p}$ we take the derivative of

$$
\mathcal{M}^{*}(p) \mathcal{M}(p) x_{p}+\eta x_{p}-\mathcal{M}^{*}(p) y=0
$$

to obtain

$$
\left(\dot{\mathcal{M}}^{*}(p) \mathcal{M}(p)+\mathcal{M}^{*}(p) \dot{\mathcal{M}}_{p}\right) x_{p}+\eta \dot{x}_{p}+\mathcal{M}^{*}(p) \mathcal{M}(p) \dot{x}_{p}=\dot{\mathcal{M}}^{*}(p) y
$$

Solving for $\dot{x}_{p}$ yields

$$
\dot{x}_{p}=\left(\mathcal{M}^{*}(p) \mathcal{M}(p)+\eta I\right)^{-1}\left(\dot{\mathcal{M}}^{*}(p) y-\dot{\mathcal{M}}(p) \mathcal{M}(p) x_{p}-\mathcal{M}^{*}(p) \dot{\mathcal{M}}(p) x_{p}\right)
$$

where the derivative of the adjoint of $\mathcal{M}(p)$ is defined by

$$
\dot{\mathcal{M}}^{*}(p) y=\dot{A}^{*} \mathcal{M}^{*}(p) y
$$

Given this Jacobian, we may use the Newton method

$$
p_{n+1}=p_{n}-\gamma\left(\frac{d^{2}}{d p^{2}} \mathcal{J}_{\eta}\right)^{-1} \frac{d}{d p} \mathcal{J}_{\eta}
$$

to find the minimizer $p^{\dagger}$.
In summary, the algorithm for simultaneously estimating the initial condition and the unknown parameter $p$ using the regularized linear least-squares is given by

## Simultaneous state/parameter estimation using RLLS:

1. Select initial parameter $p_{0}$ and set $p_{n}=p_{0}$
2. Determine $x_{p_{n}}$ via method described by 2.2.4, i.e.,

$$
x_{p_{n}}=\left(\mathcal{M}^{*}\left(p_{n}\right) \mathcal{M}\left(p_{n}\right)+\eta I\right)^{-1} \mathcal{M}^{*}\left(p_{n}\right) y
$$

3. Compute the gradient by (3.1.10) and Jacobian by (3.1.11)
4. Update the parameter $p_{n+1}$ by

$$
p_{n+1}=p_{n}-\gamma \frac{d}{d p} \mathcal{J}_{\eta}
$$

or

$$
p_{n+1}=p_{n}-\left(\frac{d^{2}}{d p^{2}} \mathcal{J}_{\eta}\right)^{-1} \frac{d}{d p} \mathcal{J}_{\eta}
$$

5. Set $p_{n}=p_{n+1}$ and return to 2

In practice, upon convergence to $p^{\dagger}$, we use the dual method 2.3 to refine the estimate of the initial condition, $x_{0}$. This approach utilizes the ease of computing the minimum for the RLLS approach, and subsequently reconstructing the initial condition using the dual approach in order to obtain a more accurate initial condition estimate.

We now formulate a direct approach for the simultaneous state/parameter estimation based on the dual control method. We assume the controls $u_{k}(p)$ are computed by the RLLS, i.e,

$$
\begin{equation*}
u_{k}(p)=\underset{u}{\arg \min }\left\{\left\|\mathscr{L}(p) u-\varphi_{k}\right\|_{X}^{2}+\eta\langle P u, u\rangle_{L^{2}\left(0, t_{f} ; Y\right)}^{2}\right\} \tag{3.1.12}
\end{equation*}
$$

and we utilize the cost functional (3.1.4). In general, other penalty terms may be considered. The parameter dependent approximation for $x_{0}$ is given by

$$
x(p)=\sum_{k=0}^{m} \int_{0}^{t_{f}}\left(u_{k}(p), y^{\delta}\right)_{Y} d t \varphi_{k} .
$$

If we use the exact derivative, it follows from (3.1.4),(3.1.12) that

$$
\begin{equation*}
\frac{d}{d p} \mathcal{J}_{\eta}(p)=\left\langle\dot{\mathcal{M}}(p) x_{p}, \mathcal{M}(p) x_{p}-y\right\rangle_{Y}+\left\langle\dot{x}_{p},\left(\mathcal{M}^{*}(p) \mathcal{M}(p)+\eta I\right) x_{p}-\mathcal{M}^{*}(p) y\right\rangle_{X} \tag{3.1.13}
\end{equation*}
$$

where

$$
\dot{x}_{p}=\sum_{k=0}^{m} \int_{0}^{t_{f}}\left(\dot{u}_{k}(p), y^{\delta}\right)_{Y} d t \varphi_{k}
$$

The derivative $\dot{u}_{k}(p)$ satisfies

$$
\left(\dot{\mathscr{L}}^{*}(p) \mathscr{L}(p)+\mathscr{L}^{*}(p) \dot{\mathscr{L}}(p)\right) u_{k}(p)+\left(\mathscr{L}^{*}(p) \mathscr{L}(p)+\eta P\right) \dot{u}_{k}(p)=\dot{\mathscr{L}}^{*}(p) \varphi_{k}
$$

The algorithm is summarized below.

## Simultaneous state/parameter estimation using the dual control formulation

1. Select step-size, $\gamma$, initial parameter $p_{0}$ and set $p_{n}=p_{0}$
2. Compute

$$
x_{0}\left(p_{n}\right)=\sum_{k=0}^{m} \int_{0}^{t_{f}}\left(u_{k}\left(p_{n}\right), y^{\delta}\right)_{Y} d t \varphi_{k}
$$

with $u_{k}\left(p_{n}\right)$ determined from (3.1.12)
3. Compute $\frac{d}{d p} \mathcal{J}_{\eta}$ by (3.1.13) where

$$
\dot{x}_{0}\left(p_{n}\right)=\sum_{k=0}^{m} \int_{0}^{t_{f}}\left(\dot{u}_{k}\left(p_{n}\right), y^{\delta}\right)_{Y} d t \varphi_{k}
$$

4. Update the parameter $p_{n+1}$ by

$$
p_{n+1}=p_{n}-\gamma \frac{d}{d p} \mathcal{J}_{\eta} .
$$

5. Set $p_{n}=p_{n+1}$ and return to 2

### 3.2 Source Identification

In this section, we consider the source identification for the abstract Cauchy problem

$$
\begin{align*}
\frac{d x}{d t}(t) & =A x(t)+f  \tag{3.2.1}\\
x(0) & =0  \tag{3.2.2}\\
y(t) & =C x(t) \tag{3.2.3}
\end{align*}
$$

where the source $f$ is assumed to be time-homogeneous.

Identifying sources is an important problem for nondestructive evaluation, contaminant localization, remote sensing applications, et cetera. Related formulations can be found in the papers [35,51]. Since the force is time-homogeneous, the source identification problem fits into our framework by differentiating the state and observation equations in time. The system obtained by differentiating (3.2.1)-(3.2.3) is given by

$$
\begin{gather*}
\frac{d \dot{x}}{d t}=A \dot{x}  \tag{3.2.4}\\
\dot{x}(0)=f  \tag{3.2.5}\\
\dot{y}=C \dot{x} \tag{3.2.6}
\end{gather*}
$$

which transforms the source identification problem into the initial condition reconstruction problem. If it is reasonable to differentiate the data, $y^{\delta}$, then the generalized Fourier coefficients of $f$ can be approximated by

$$
\begin{equation*}
\left\langle f, p_{0}\right\rangle_{X}=\int_{0}^{t_{f}}\left(u_{k}(t), \dot{y}^{\delta}(t)\right)_{Y} d t \tag{3.2.7}
\end{equation*}
$$

as in the case of the initial condition estimation. In most cases, it is desirable to avoid differentiation of the data $y^{\delta}$, especially if the noise level is high, or the measurements are inaccurate. To avoid this problem, we simply integrate (3.2.7) by parts to obtain

$$
\begin{align*}
\int_{0}^{t_{f}}\left(u_{k}(t), \dot{y}^{\delta}(t)\right) d t= & -\int_{0}^{t_{f}}\left(\dot{u}_{k}(t), y^{\delta}(t)\right)_{Y} d t \\
& +\left(u_{k}\left(t_{f}\right), y^{\delta}\left(t_{f}\right)\right)_{Y}-\left(u_{k}(0), y^{\delta}(0)\right)_{Y} \tag{3.2.8}
\end{align*}
$$

alleviating this concern, since by assumption $u_{k} \in L^{2}\left(0, t_{f} ; Y\right)$. The following summarizes the approach for estimating the source, $f$.

## Dual Method for reconstruction of time-homogeneous source:

1. Pick an orthonormal basis, $\left\{\varphi_{k}\right\}_{k=0}^{m}$ for $X$
2. For each $k$ solve $\mathscr{L} u_{k}=\varphi_{k}$ for $u_{k} \in L^{2}\left(0, t_{f} ; Y\right)$
3. Form the estimate for $f$,

$$
f^{m}=\sum_{k=0}^{m} \alpha_{k} \varphi_{k}
$$

where

$$
\alpha_{k}=-\int_{0}^{t_{f}}\left(\dot{u}_{k}, y^{\delta}\right)_{Y} d t+\left(u_{k}\left(t_{f}\right), y^{\delta}\left(t_{f}\right)\right)_{Y}-\left(u_{k}(0), y^{\delta}(0)\right)_{Y}
$$

Though not detailed here, this approach can be extended to the case when $f$ is timeinhomogeneous, such that

$$
\frac{d x}{d t}(t)=A x(t)+h(t) f
$$

where $h$ is a known function.

### 3.2.1 Source Identification in the Reproducing Kernel Hilbert Space Framework

In this section, we discuss a specific formulation of the reconstruction methods developed in Chapter 2, for estimating the source in (2.1.1). Specifically, we provide the details for the case when the state space $X=\mathcal{H}$ is a reproducing kernel Hilbert space, the definition of which follows. Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle\cdot, \cdot\rangle_{\mathcal{H}}$, where each $f \in \mathcal{H}$ is defined in $E \subset \mathbb{R}^{n}$, for $E$ arbitrary and non-empty. A symmetric function $\Phi: E \times E \rightarrow \mathbb{R}$ is termed a kernel. Such a kernel, $\Phi$, is called positive definite if for all pairwise distinct points $\tilde{X}=\left\{\xi_{1}, \ldots, \xi_{n}\right\} \subset E$ the Gram matrix $\mathcal{W}_{k, j}=\Phi\left(\xi_{k}, \xi_{j}\right)$ is positive definite. Now, we define the following concept.

Definition 3.2.1. A function $\Phi: E \times E \rightarrow \mathbb{R}$ is called a reproducing kernel for $\mathcal{H}$ if

1. $\Phi(\cdot, \xi) \in \mathcal{H}$ for all $\xi \in E$
2. $f(\xi)=\langle f, \Phi(\cdot, \xi)\rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and all $\xi \in E$.

A space of functions $\mathcal{H}$ admitting a reproducing kernel is termed a reproducing kernel Hilbert space(RKHS). For a function $f \in \mathcal{H}$ the norm is defined by

$$
\|f\|_{\mathcal{H}}=\langle f, f\rangle_{\mathcal{H}}^{\frac{1}{2}},
$$

and by the second property 3.2.1-2

$$
\|f\|_{\mathcal{H}}^{2}=\sum_{k=0}^{n} \sum_{j=0}^{n} \alpha_{k} \alpha_{j} \Phi\left(\xi_{k}, \xi_{j}\right)
$$

whenever $f$ is of the form

$$
f=\sum_{k=0}^{n} \alpha_{k} \Phi\left(\cdot, \xi_{k}\right), \quad \xi_{k} \in E .
$$

By the previous section, if the source $f$ is time-homogeneous, we can transform the source identification problem into an initial condition identification problem. Thus, we formulate the method in terms of the initial condition identification. Under this framework, we can form approximations by seeking coefficients $\alpha_{k}$ such that

$$
x_{0}=\sum_{k=0}^{m} \alpha_{k} \Phi\left(\cdot, \xi_{k}\right)
$$

for some $\xi_{k}$. The coefficients $\left\{\alpha_{k}\right\}_{k=1}^{m}$ are uniquely determined by

$$
\sum_{k=0}^{m} \alpha_{k} \Phi\left(\xi_{j}, \xi_{k}\right)=x_{0}\left(\xi_{j}\right), \quad 0 \leq j \leq m
$$

since the matrix $\mathcal{W}$ is positive definite.
Common choices for the kernel $\Phi$ include

$$
\begin{array}{ll}
\Phi(t, s)=e^{-\mu\|t-s\|^{2}} & \text { Gaussian } \\
\Phi(t, s)=\sqrt{\|t-s\|^{2}+\mu^{2}} & \text { Multiquadric } \\
\Phi(t, s)=(1-\|t-s\|)_{+}^{3}(3\|t-s\|+1) & \text { Wendlund. }
\end{array}
$$

The Gaussian kernel $\Phi(t, s)=\frac{1}{2}=e^{-\|t-s\|}$ produces the Sobolev space $\mathcal{H}=H^{1}(\mathbb{R})$, while the kernel $\Phi(t, s)=(1-\|t-s\|)_{+}^{3}(3\|t-s\|+1)$ is compactly supported [50].

In the context of the regularized linear least-squares approach, we can reconstruct the initial condition by determining the coefficients $\alpha_{k}$ by

$$
\min _{\alpha \in \mathbb{R}^{m+1}}\left\|y^{\delta}-\sum_{k=0}^{m} \alpha_{k} C S_{t_{f}} \Phi\left(\cdot, \xi_{k}\right)\right\|_{Y}+\beta \sum_{k=0}^{m} \sum_{j=0}^{m} \alpha_{k} \alpha_{j} \Phi\left(\xi_{k}, \xi_{j}\right),
$$

where $C: \mathcal{H} \rightarrow Y, S_{t_{f}}: \mathcal{H} \rightarrow \mathcal{H}$. By the necessary optimality, $\alpha=\left(\alpha_{0}, \ldots, \alpha_{m}\right)^{t}$ can be determined by solving

$$
(A+\beta B) \alpha=F
$$

where

$$
A_{i, j}=\left\langle C S_{t_{f}} \Phi\left(\cdot, \xi_{i}\right), C S_{t_{f}} \Phi\left(\cdot, \xi_{j}\right)\right\rangle_{Y}, \quad B_{i, j}=\Phi\left(\xi_{i}, \xi_{j}\right)
$$

and

$$
F_{i}=\left\langle C S_{t_{f}} \Phi\left(\cdot, \xi_{i}\right), y^{\delta}\right\rangle_{Y}
$$

With the $\alpha_{k}$ 's determined, we have the approximation

$$
x_{0}^{m}=\sum_{k=0}^{m} \alpha_{k} \Phi\left(\cdot, \xi_{k}\right) .
$$

This formulation is similar to the one developed in [48].
The extension to the dual control formulation (2.3) is straightforward for $X=\mathcal{H}$. We note that the relation

$$
\langle x(0), p(0)\rangle_{X}=\int_{0}^{t_{f}}(u(t), y(t))_{Y} d t
$$

holds regardless if $p$ is chosen to be an orthogonal basis. By the assumption that

$$
x_{0}=\sum_{k=0}^{m} \alpha_{k} \Phi\left(\cdot, \xi_{k}\right)
$$

we have

$$
\begin{equation*}
\left\langle\sum_{k=0}^{m} \alpha_{k} \Phi\left(\cdot, \xi_{k}\right), p\right\rangle_{\mathcal{H}}=\int_{0}^{t_{f}}(u(t), y(t))_{Y} d t \tag{3.2.9}
\end{equation*}
$$

Thus, by choosing a set of functions $\left\{p_{k}\right\}_{k=0}^{m}$ we can determine $\alpha_{k}$ such that (3.2.9) holds where

$$
\mathscr{L} u_{k}=p_{k}
$$

for each $k$. Note that the collection $\left\{p_{k}\right\}_{k=0}^{m}$ is not required to be of any specific form as long as $p_{k} \in \mathcal{H}$ for which $\Phi$ is the reproducing kernel. However, if $p_{k}=\Phi\left(\cdot, \xi_{k}\right) \in \mathcal{H}$, we have

$$
\left\langle\sum_{k=0}^{m} \alpha_{k} \Phi\left(\cdot, \xi_{k}\right), \Phi\left(\cdot, \xi_{j}\right)\right\rangle_{\mathcal{H}}=\int_{0}^{t_{f}}\left(u_{j}(t), y(t)\right)_{Y} d t \quad 0 \leq j \leq m
$$

and the $\left\{\alpha_{k}\right\}_{k=0}^{m}$ are uniquely determined since $\Phi$ is a positive definite kernel. That is, by the properties of the reproducing kernel, we solve

$$
\begin{equation*}
\sum_{k=0}^{m} \alpha_{k} \Phi\left(\xi_{k}, \xi_{j}\right)=\int_{0}^{t_{f}}\left(u_{j}(t), y^{\delta}(t)\right)_{Y} d t, \quad 0 \leq j \leq m \tag{3.2.10}
\end{equation*}
$$

The solutions of

$$
\mathscr{L} u_{k}=p_{k}
$$

can be approximated by solving

$$
\min _{u \in Y}\left\|\mathscr{L} u-\Phi\left(\cdot, \xi_{k}\right)\right\|_{\mathcal{H}}+\beta\|u\|_{L^{2}\left(0, t_{f} ; Y\right)}
$$

for each $k$. Other suitable methods are developed in Chapter 4.

## Chapter 4

## Regularization for Inverse Problems

### 4.1 Motivation and Preliminaries

In Chapters 2 and 3, each of the proposed methods leads to an inverse problem of the form

$$
\begin{equation*}
K x=y \tag{4.1.1}
\end{equation*}
$$

where $x \in X, y \in Y$ with $X, Y$ Banach spaces and $K: X \rightarrow Y$ a compact operator. The objective is to recover the function $x$ given noise corrupted data, $y^{\delta}$. That is,, $\left\|y^{\delta}-y\right\| \leq \delta$ where the noise level, $\delta$, may not be known a priori. Due to the compactness of $K$, the problem (4.1.1) is ill-posed, meaning at least one of the following criteria for well-posedness is not met:

Definition 4.1.1 (Well-posed).

1. For all admissible data, a solution exists,
2. For all admissible data, the solution is unique,
3. The solution depends continuously on the data.

The compactness of $K$ signifies ill-posedness due to the following proposition from [11].

Proposition 4.1.1. Let $K: X \rightarrow Y$ be compact, $\operatorname{dim} R(K)=\infty$. Then $K^{\dagger}$ is a densely defined, unbounded linear operator with closed graph.

That is, the minimum norm solution to (4.1.1) will be unbounded. Considerable challenges are faced when solving such inverse problems, due to their ill-posedness. Even if the operator $K$ is invertible, it may be numerically ill-conditioned, which can yield meaningless solutions for highly ill-conditioned problems. The condition number of $K$ (or its discretization) is given by

$$
\begin{equation*}
\operatorname{cond}(K)=\frac{\sigma_{1}}{\sigma_{n}} \tag{4.1.2}
\end{equation*}
$$

where the $\left\{\sigma_{i}\right\}_{i=1}^{n}$ are the ordered singular values of $K$, in decreasing order. Due to the ill-posedness, approximate solutions of (4.1.1) are obtained by regularizing the problem, meaning that the problem is perturbed in some manner, so that the new problem is better posed. When an approximate solution $x$ to (4.1.1) is attainable, one would like to know how closely this solution approximates the proper solution, $x^{\dagger}$. That is, given what information is known, can we obtain an a priori estimate of the error, and if not, what kind of a posteriori estimate can be obtained. The most well known regularization method is the Tikhonov regularization [49], which will be discussed in the section that follows. We also develop a more general regularization framework that is employed for the solution of the inverse problems considered in this thesis.

Inverse problems specific to this thesis include the solution of

$$
\mathscr{L} u_{k}=\varphi_{k}, \quad 0 \leq k \leq m,
$$

where $\mathscr{L}$ is defined by (2.3.11), in order to determine the controls, $u_{k}$, for approximating the generalized Fourier coefficients (2.3.14). In this case, the basis functions, $\varphi_{k}$, which represent the data, are noiseless. However, in the more direct reconstruction method of linear-least squares, the inverse problem is to determine $x_{0}$ such that

$$
\mathcal{M} x_{0}=y^{\delta}
$$

given the noisy data $y^{\delta}$.
In this chapter, we develop methods for problems of the form (4.1.1) that can be applied to a wide range of problems, in addition to the abstract Cauchy problem (2.1.1). Thus, we develop the theory in the general context of (4.1.1), with the abstract Cauchy problem in mind. Problems of this form arise in all areas of science. An incomplete list
of applications includes computerized tomography, inverse scattering theory, and signal processing to name a few.

### 4.2 Regularization Methods

### 4.2.1 Standard Tikhonov Regularization

A classical technique for regularizing ill-posed problems is that of Tikhonov, which we describe in this section. Let $K: X \rightarrow Y$ where $X$ and $Y$ are Banach spaces, with $K$ being a compact operator. Typically, $K$ is an integral operator of the form

$$
(K x)(s)=\int_{\Omega} k(s, \xi) x(\xi) d \xi
$$

where $k$ is the kernel of the operator. The task is to solve the equation

$$
\begin{equation*}
K x=y^{\delta} \tag{4.2.1}
\end{equation*}
$$

for $x \in X$ given $y \in Y$, and $y^{\delta}$ is the noise contaminated data. In general, the problem may be very ill-posed, by which we mean the solution(if it exists) does not depend continuously on the data, as described in definition 4.1.1. Due to the fact that $K$ (or its discretization) may not be invertible, we seek a minimum norm solution to (4.2.1), that is, we seek $x^{*}$ such that

$$
\begin{equation*}
\left\|K x^{*}-y^{\delta}\right\|_{X}^{2}=\min _{x \in C}\left\{\left\|K x-y^{\delta}\right\|_{X}^{2}\right\} . \tag{4.2.2}
\end{equation*}
$$

However, this does not mitigate the unboundedness of the solutions due to the compactness of $K$. In order to obtain a better posed problem, one seeks a solution to

$$
\begin{equation*}
\min _{x \in C}\left\|K x-y^{\delta}\right\|_{X}^{2}+\beta\|x\|_{X}^{2} \tag{4.2.3}
\end{equation*}
$$

for some $\beta>0$. The set $C$ is a closed, convex subset of $X$, representing constraints for the solution (e.g., lower/upper bounds for solution). This technique, known as Tikhonov regularization, has shown remarkable applicability since its introduction in [49]. The term $\|x\|_{2}^{2}$ in the Tikhonov regularization is a penalty term aimed at ensuring the approximate
solution remains bounded. This technique has close ties with a regularized Singular Value Decomposition and with Bayesian maximum likelihood estimators, as will be discussed below. The general idea of Tikhonov regularization is to regularize the operator $K$ by the introduction of the positive parameter $\beta$, effectively shifting the singular values of $K$ away from zero. To see this, consider the minimization of the functional

$$
\begin{equation*}
\min _{x \in C} \frac{1}{2}\left\|K x-y^{\delta}\right\|_{X}^{2}+\frac{\beta}{2}\|x\|_{X}^{2} \tag{4.2.4}
\end{equation*}
$$

where the operator $K$ has the singular value decomposition $K=U \Sigma V^{*}$ where $U, V$ are unitary operators and $\Sigma$ has a diagonal matrix representation. Then, assuming a real Banach space and letting $v=V^{*} x$, the minimization problem is equivalent to minimizing

$$
\begin{aligned}
\frac{1}{2}\left\langle U \Sigma V^{*} x-y^{\delta}, U \Sigma V^{*} x-y^{\delta}\right\rangle+\frac{\beta}{2}\langle x, & x\rangle \\
& =\frac{1}{2}\left\langle\Sigma v-U^{*} y^{\delta}, \Sigma v-U^{*} y^{\delta}\right\rangle+\frac{\beta}{2}\langle v, v\rangle \\
& =\frac{1}{2}\left\langle v, \Sigma^{*} \Sigma v\right\rangle-\left\langle\Sigma^{*} U^{*} y^{\delta}, v\right\rangle+\frac{1}{2}\left\langle y^{\delta}, y^{\delta}\right\rangle+\frac{\beta}{2}\langle v, v\rangle \\
& =\frac{1}{2}\left\langle v,\left(\Sigma^{*} \Sigma+\beta I\right) v\right\rangle-\left\langle\Sigma^{*} U^{*} y^{\delta}, v\right\rangle+\frac{1}{2}\left\langle y^{\delta}, y^{\delta}\right\rangle
\end{aligned}
$$

By the necessary optimality, the minimum occurs when

$$
\begin{equation*}
v_{\beta}^{\delta}=\left(\Sigma^{*} \Sigma+\beta I\right)^{-1} \Sigma^{*} U^{*} y^{\delta}, \tag{4.2.5}
\end{equation*}
$$

denoting the dependence of the solution on $\beta, \delta$ by $v_{\beta}^{\delta}$. By this analysis, one can notice that the singular values are shifted by the value $\beta>0$. In fact, if the operator $K$ has singular system $\left\{\sigma_{n} ; v_{n} ; u_{n}\right\}$, then the solution (4.2.5) has the form

$$
v_{\beta}^{\delta}=\sum_{n=1}^{\infty} \frac{\sigma_{n}}{\sigma_{n}^{2}+\beta}\left\langle y^{\delta}, u_{n}\right\rangle v_{n}
$$

Henceforth, we define the cost functional

$$
\begin{equation*}
\mathcal{J}_{\beta}(x):=\frac{1}{2}\left\|K x-y^{\delta}\right\|_{X}^{2}+\frac{\beta}{2}\langle x, P x\rangle_{X} \tag{4.2.6}
\end{equation*}
$$

and denote the solution of (4.2.6) by

$$
\begin{equation*}
x_{\beta}^{\delta}=\left(K^{*} K+\beta P\right)^{-1} K^{*} y^{\delta} . \tag{4.2.7}
\end{equation*}
$$

where $P$ is a positive, self-adjoint operator on $X$, and $\langle P x, x\rangle^{\frac{1}{2}}$ defines a norm on $X$.
The following Theorem gives the uniqueness of the minimizer of (4.2.6).
Theorem 4.2.1. Let $x_{\beta}^{\delta}$ be the solution denoted by (4.2.7). Then $x_{\beta}^{\delta}$ is the unique minimizer of

$$
\frac{1}{2}\left\|K x-y^{\delta}\right\|_{X}^{2}+\frac{\beta}{2}\langle x, P x\rangle_{X}
$$

Proof. Since $P$ is positive, for $\beta>0$ we have that the cost functional (4.2.6) is strictly convex and has a unique minimizer. The minimizer is characterized by

$$
\left\langle K x-y^{\delta}, K z\right\rangle_{X}+\beta\langle P x, z\rangle_{X}=\left\langle\left(K^{*} K+\beta P\right) x-K^{*} y^{\delta}, z\right\rangle_{X}=0 \quad \forall z \in X,
$$

which is equivalent to (4.2.7).
Furthermore, the following convergence result can be established.
Theorem 4.2.2. Let $x_{\beta}^{\delta}$ be defined by (4.2.7), $y \in R(K),\left\|y-y^{\delta}\right\| \leq \delta$. If $\beta=\beta(\delta)$ satisfies

$$
\lim _{\delta \rightarrow 0} \frac{\delta^{2}}{\beta(\delta)}=0
$$

then

$$
\lim _{\delta \rightarrow 0} x_{\beta(\delta)}^{\delta}=x^{\dagger}=K^{\dagger} y
$$

is the minimum norm solution, i.e. $K^{\dagger} y=x^{\dagger}$ minimizes $\langle x, P x\rangle_{X}^{\frac{1}{2}}$ over all solutions to $K x=y$.

Proof. We first show that the sequence $\left\{x_{n}\right\}$ is bounded, where $x_{n}$ is the unique minimizer of $\mathcal{J}_{n}(x):=\mathcal{J}_{\beta_{n}}(x)$. Since $x_{n}$ minimizes (4.2.6) and $P$ is self-adjoint we have

$$
\begin{aligned}
\beta_{n}\left\langle x_{n}, P x_{n}\right\rangle_{X} & \leq \beta_{n}\left\langle x_{n}, x_{n}\right\rangle_{X} \leq \mathcal{J}_{n}\left(x_{n}\right) \leq \mathcal{J}_{n}\left(x^{\dagger}\right) \\
& \leq \delta_{n}^{2}+\beta_{n}\|P\|\left\langle x^{\dagger}, x^{\dagger}\right\rangle_{X}
\end{aligned}
$$

which implies

$$
\begin{equation*}
\left\|x_{n}\right\|_{X}^{2} \leq \frac{\delta_{n}^{2}}{\beta_{n}}+\|P\| \cdot\left\|x^{\dagger}\right\|_{X}^{2} \tag{4.2.8}
\end{equation*}
$$

Thus, there exists a subsequence $\left\{x_{n_{k}}\right\}$ such that

$$
x_{n_{k}} \rightharpoonup x \in X
$$

and

$$
K x_{n_{k}} \rightharpoonup K x
$$

since $K$ is a bounded linear operator on $X$. Then,

$$
\left\|K x_{n_{k}}-y_{n}^{\delta}\right\|^{2} \leq \mathcal{J}_{n_{k}}\left(x_{n_{k}}\right) \leq \delta_{n_{k}}^{2}+\beta_{n_{k}}\left\langle x^{\dagger}, P x^{\dagger}\right\rangle_{X} \rightarrow 0 \quad \text { as } k \rightarrow \infty
$$

Combining the above yields

$$
K x=y .
$$

It remains to show that $x=x^{\dagger}$ and $x_{n} \rightarrow x^{\dagger}$ strongly. By the minimizing property of $x_{n}$ we have

$$
\left\langle K x_{n}-y_{n}^{\delta}, K z\right\rangle_{X}+\beta\left\langle P x_{n}, z\right\rangle_{X}=0, \quad \forall z \in X,
$$

which implies

$$
\beta\left\langle P x_{n}, z\right\rangle_{X}=0
$$

by taking $z \in N(K)$. We note that if $z \in N(K)$ then $P z \in N(K)$ so that $x_{n} \in N(K)^{\perp}$ using the self-adjointness of $P$. Hence, we also have that $z \in N(K)^{\perp}$. By uniqueness of $x^{\dagger}=K^{\dagger} y$, we have $x^{\dagger}=z$ and $x_{n_{k}} \rightharpoonup x^{\dagger}$. As a consequence, we have

$$
x_{n} \rightharpoonup x^{\dagger},
$$

which yields the weak convergence of $x_{n}$ to $x^{\dagger}$. To show the strong convergence, let us assume there exists $\varepsilon>0$ and a subsequence $\left\{x_{n_{k}}\right\}$ such that for all $k,\left\|x_{n_{k}}\right\| \leq\left\|x^{\dagger}\right\|-\varepsilon$. By the boundedness of this subsequence, there would be a further subsequence of $\left\{x_{n}\right\}$ such that

$$
x_{n_{k_{l}}} \rightharpoonup z \in X
$$

and

$$
\|z\| \leq\left\|x^{\dagger}\right\|-\varepsilon
$$

However, this contradicts the minimizing property of the original sequence, from which it follows that

$$
\liminf _{n \rightarrow \infty}\left\|x_{n}\right\| \geq\left\|x^{\dagger}\right\|
$$

By the inequality (4.2.8), we have

$$
\limsup _{n \rightarrow \infty}\left\|x_{n}\right\| \leq\left\|x^{\dagger}\right\|
$$

which implies

$$
x_{n} \rightarrow x^{\dagger}
$$

Letting $\delta_{n} \rightarrow 0$, we obtain

$$
x_{\beta(\delta)}^{\delta} \rightarrow K^{\dagger} y .
$$

The selection of the parameter $\beta$ reflects a balance between ensuring the fidelity $\left\|K x-y^{\delta}\right\|_{X}^{2}$ is small and ensuring the solution does not become unbounded. Thus, $\beta$ must be chosen carefully in order to obtain an acceptable approximation. Selecting the regularization term forms the basis for many modified regularization methods. These two aspects of the regularization method are discussed and analyzed in the following sections.

### 4.2.2 Generalized Multi-parameter Approach

In a more general context, the Tikhonov regularization (4.1.1) can be recast as the minimization of

$$
\begin{equation*}
\mathcal{J}_{\beta}(x)=\phi\left(x, y^{\delta}\right)+\beta \psi(x), \tag{4.2.9}
\end{equation*}
$$

over $x \in C$, where the fidelity term, $\phi$, is chosen based on the noise statistic, while $\psi$ is chosen based on which class the solution $x$ should belong to. When $\phi\left(x, y^{\delta}\right)=\|K x-y\|_{X}^{2}$ and $\psi(x)=\|x\|_{X}^{2}$ this formulation is the classical Tikhonov regularization (4.2.3). The main drawback to this method is the single regularization term $\psi$. Modern day scientific problems typically involve applications where the standard Tikhonov regularization fails to capture the full set of distinct features in the physical solution. Many research efforts
have been devoted to improving the standard regularization techniques for a wide range of applications(see [8, 28, 41, 46] for example). Especially in the field of image processing, the solution often exhibits a multi-scale structure typically described by multi-resolution analysis. In such applications, single parameter regularization can oversmooth the solution in the case of $\psi=\|\cdot\|_{L^{2}}^{2}$ or exhibit stair-case effects in the case of $\psi=\|\cdot\|_{T V}$. Here, the total variation (TV) of a function $f$ over $[a, b]$, is defined by

$$
\sup \sum_{i=1}^{k}\left|f\left(x_{i+1}\right)-f\left(x_{i}\right)\right|
$$

where the supremum is taken over all possible subdivisions of the interval $[a, b]$. In order to capture the multi-scale structure of solutions without introducing oversmoothing or staircasing, many research efforts have focused on mixed regularization approaches, such as combining the $L^{2}$ penalty term with the $T V$ penalty:

$$
\begin{equation*}
\min _{x \in C} \frac{1}{2} \int_{\Omega}\left|K x-y^{\delta}\right|^{2} d \xi+\frac{\eta_{1}}{2} \int_{\Omega}|x|^{2} d \xi+\eta_{2} \int_{\Omega}|\nabla x| d \xi \tag{4.2.10}
\end{equation*}
$$

In general, we cast this as the multi-parameter Tikhonov regularization technique, i.e., we minimize

$$
\begin{equation*}
\mathcal{J}_{\boldsymbol{\eta}}(x)=\phi\left(x, y^{\delta}\right)+\boldsymbol{\eta} \cdot \boldsymbol{\psi}(x) \tag{4.2.11}
\end{equation*}
$$

The terms $\phi, \boldsymbol{\psi}$ are known as the fidelity and regularization terms, respectively. Here, $\left\{\psi_{k}\right\}_{k=1}^{n}$ is the set of regularization terms, $\left\{\eta_{k}\right\}_{k=1}^{n}$ are the regularization parameters, and we take the dot product

$$
\boldsymbol{\eta} \cdot \boldsymbol{\psi}(x)=\sum_{k=1}^{n} \eta_{k} \psi_{k}(x)
$$

for $\boldsymbol{\eta}=\left(\eta_{1}, \eta_{2}, \ldots, \eta_{n}\right)$ and $\boldsymbol{\psi}(x)=\left(\psi_{1}(x), \psi_{2}(x), \ldots, \psi_{n}(x)\right)$. The functionals $\phi, \boldsymbol{\psi}$ can be chosen based on any a priori information about the problem and its exact solution. Then,

$$
x_{\boldsymbol{\eta}}=\underset{x \in C}{\arg \min } \mathcal{J}_{\boldsymbol{\eta}}(x)
$$

is taken as the regularized solution. For instance, in the case of a multi-scale image with a smooth region and a stepped region, one may consider the $L^{2}-T V$ regularization (4.2.10).

A common theme in the literature and throughout this thesis, is the introduction of sparsity. Due to the ever increasing size of today's most challenging and interesting problems, it is often necessary to reduce the size of a problem and/or obtain a solution with as few nonzeros as possible. This idea corresponds to minimizing the size of a dataset or solution by keeping only what information is necessary in order to describe the solution accurately. In the case of obtaining solutions with few nonzeros, we consider regularization terms such as

$$
\psi_{k}(x)=\|x\|_{\ell_{p}}^{p}:=\sum_{k=1}^{\infty}\left|x_{k}\right|^{p} \quad \text { for } p \leq 1 .
$$

This particular choice of $\psi$ is examined in Chapter 5, along with treatment of the associated numerical difficulties.

Naturally, when considering problems of the form (4.2.9) two questions arise as to the solution. Firstly, does this method always "work"? That is, does this method always yield a close approximation to the proper solution $x^{\dagger}$ ? Secondly, how does one go about selecting the optimal parameter $\beta$ ?

Before proceeding, we briefly describe the probabilistic interpretation of the minimization of (4.2.9). A Bayesian model is formulated as deducing the distribution of the unknown $x$ conditioned on the data $y^{\delta}$. That is, we must deduce the posterior probability density function (PPDF). According to Bayes' rules, the PPDF is given by

$$
\begin{equation*}
p\left(x \mid y^{\delta}\right)=\frac{p\left(y^{\delta} \mid x\right) p(x)}{\int p\left(y^{\delta} \mid x\right) p(x) d x} \tag{4.2.12}
\end{equation*}
$$

where $p\left(y^{\delta} \mid x\right)$ is the conditional density of $y^{\delta}$ given $x \in X$ and $p(x)$ is the prior probability density function of $x \in X$. The denominator in (4.2.12) is the normalizing constant. Thus, the unnormalized PPDF is described by

$$
\begin{equation*}
p\left(x \mid y^{\delta}\right) \propto p\left(y^{\delta} \mid x\right) p(x) \tag{4.2.13}
\end{equation*}
$$

In terms of the single term regularization (4.2.9) we have

$$
p\left(x \mid y^{\delta}\right) \propto \exp \left(-\phi\left(x, y^{\delta}\right)\right) \exp (-\beta \psi(x))
$$

Here, we see that the fidelity gives the conditional density, while the regularization gives
the prior density. As a particular example,

$$
p\left(x \mid y^{\delta}\right) \propto \exp \left(-\frac{\left\|K x-y^{\delta}\right\|_{2}^{2}}{2 \sigma^{2}}\right) \exp \left(-\frac{\lambda}{2}\langle x, P x\rangle\right)
$$

when $\phi\left(x, y^{\delta}\right)=\left\|K x-y^{\delta}\right\|_{2}^{2}$ and $\psi(x)=\langle x, P x\rangle_{X}$ for the self-adjoint operator $P$ (i.e., the $H^{1}$ semi-norm). In the probabilistic interpretation, the parameter is chosen based on how much we trust the prior information. We examine this interpretation further through the development of parameter choice rules in the section to follow.

### 4.3 Choice rules for the regularization parameter

To obtain a stable solution, $x_{\beta}^{\delta}$, which closely approximates the true solution, $x^{\dagger}$, the regularization parameter(s) must be chosen properly. If the regularization parameter(s) is too small, then the approximate solution will be unstable due to the ill-posedness of the operator $K$. However, if the parameter(s) is too large, the approximate solution may be oversmoothed. In this section, we briefly describe the Morozov's Discrepancy Principle, as well as a new choice rule developed in $[25,26]$.

The analysis of the parameter choice rules is carried out using the value functional defined by

$$
\begin{equation*}
F(\beta)=\inf _{x} \phi\left(x, y^{\delta}\right)+\beta \psi(x) . \tag{4.3.1}
\end{equation*}
$$

We briefly analyze the properties of the value functional for the single parameter case, which is to be used in the remainder of this section. Further analysis, including the multi-parameter case, can be found in [25, 26]. With $F$ defined by (4.3.1), we have the following basic results.

Lemma 4.3.1. The functional $F$ defined by (4.3.1) is monotonically increasing, assuming $\psi$ is a nonnegative function.

Proof. To show this, let $\beta_{1}<\beta_{2}$ so that we have

$$
\begin{equation*}
\phi\left(x, y^{\delta}\right)+\beta_{1} \psi(x)<\phi\left(x, y^{\delta}\right)+\beta_{2} \psi(x) \tag{4.3.2}
\end{equation*}
$$

which yields the monotonicity of $F$ by taking the infimum over all $x$.
It should also be clear that $F$ is concave, nonetheless the result is provided.

Lemma 4.3.2. The functional $F$ is concave.
Proof. Let $\alpha \in \mathbb{R}$, then

$$
\begin{aligned}
F\left((1-\alpha) \beta_{1}+\alpha \beta_{2}\right) & =\inf \left\{\phi\left(x, y^{\delta}\right)+(1-\alpha) \beta_{1} \psi(x)+\alpha \beta_{2} \psi(x)\right\} \\
& =\inf \left\{(1-\alpha)\left(\phi\left(x, y^{\delta}\right)+\beta_{1} \psi(x)\right)+\alpha\left(\phi\left(x, y^{\delta}\right)+\beta_{2} \psi(x)\right)\right\} \\
& \geq \inf \left\{(1-\alpha)\left(\phi\left(x, y^{\delta}\right)+\beta_{1} \psi(x)\right)\right\}+\inf \left\{\alpha\left(\phi\left(x, y^{\delta}\right)+\beta_{2} \psi(x)\right)\right\} \\
& =(1-\alpha) F\left(\beta_{1}\right)+\alpha F\left(\beta_{2}\right),
\end{aligned}
$$

which shows that $F$ is concave.
Note that if $F$ is differentiable at $\beta>0$, then

$$
\begin{equation*}
F^{\prime}\left(x_{\beta}\right)=\psi\left(x_{\beta}\right) \tag{4.3.3}
\end{equation*}
$$

since, assuming $x_{\beta}$ is differentiable in the classical sense,

$$
\begin{aligned}
\frac{d}{d \beta}\left(\phi\left(u_{\beta}, y^{\delta}\right)+\beta \psi\left(u_{\beta}\right)\right) & =\phi^{\prime}\left(x_{\beta}, y^{\delta}\right) \frac{d}{d \beta} x_{\beta}+\psi\left(x_{\beta}\right)+\beta \psi^{\prime}\left(x_{\beta}\right) \frac{d}{d \beta} x_{\beta} \\
& =\frac{d}{d \beta} x_{\beta}\left(\phi^{\prime}\left(x_{\beta}, y^{\delta}\right)+\beta \psi^{\prime}\left(x_{\beta}\right)\right)+\psi\left(x_{\beta}\right) \\
& =\psi\left(x_{\beta}\right)
\end{aligned}
$$

due to the necessary optimality $\phi^{\prime}\left(x_{\beta}, y^{\delta}\right)+\beta \psi^{\prime}\left(x_{\beta}\right)=0$. When $F$ is differentiable, we also have the relation

$$
\begin{equation*}
\phi(\beta)=F(\beta)-\beta F^{\prime}(\beta) . \tag{4.3.4}
\end{equation*}
$$

The following argument shows that (4.3.3) holds under much weaker conditions. Given $\beta \geq 0$ we let

$$
M_{\beta}=\left\{x_{\beta} \in X: \mathcal{J}_{\beta}(x) \text { is minimized }\right\}
$$

be the set of all minimizers for the given $\beta$. Then, for $0 \leq \beta \leq \hat{\beta}$

$$
\phi\left(x_{\beta}\right) \leq \phi\left(x_{\hat{\beta}}\right) \quad \text { and } \quad \psi\left(x_{\beta}\right) \geq \psi\left(x_{\hat{\beta}}\right)
$$

for all minimizers $x_{\beta} \in M_{\beta}$ and $x_{\hat{\beta}} \in M_{\hat{\beta}}$. Moreover, for all $h>0$ sufficiently small and
all minimizers $x_{\beta} \in M_{\beta}$,

$$
\begin{equation*}
\frac{F(\beta+h)-F(\beta)}{h} \leq \psi\left(x_{\beta}\right) \leq \frac{F(\beta)-F(\beta-h)}{h .} \tag{4.3.5}
\end{equation*}
$$

The left-hand side of the previous inequality (4.3.5) is monotonically increasing and the right-hand side is monotonically decreasing as $h \rightarrow 0^{+}$. Thus, the left and right one-sided derivatives exist and the following inequality holds:

$$
D^{+} F(\beta) \leq \psi\left(x_{\beta}\right) \leq D^{-} F(\beta) .
$$

### 4.3.1 Morozov's Discrepancy Principle

In this section, we formulate the variational justification for the Morozov's discrepancy principle [38]. The Morozov's discrepancy principle seeks $\beta>0$ such that

$$
\begin{equation*}
\beta \rightarrow \frac{F(\beta)-\sigma}{\beta} \tag{4.3.6}
\end{equation*}
$$

is maximized, where $F$ is the value functional (4.3.1) and $\sigma:=c \delta$ for some $c \geq 1$. The left and right derivatives of (4.3.6) are respectively given by

$$
\begin{aligned}
& D^{-}\left(\frac{F(\beta)-\sigma}{\beta}\right)=\left(\beta D^{-} F(\beta)-(F(\beta)-\sigma)\right) \frac{1}{\beta^{2}} \geq 0 \\
& D^{+}\left(\frac{F(\beta)-\sigma}{\beta}\right)=\left(\beta D^{+} F(\beta)-(F(\beta)-\sigma)\right) \frac{1}{\beta^{2}} \leq 0
\end{aligned}
$$

Thus, if $F$ is differentiable, from (4.3.3),(4.3.4) we have

$$
\beta \psi\left(x_{\beta}\right)-(F(\beta)-\sigma)=\phi\left(x_{\beta}\right)-\sigma=0
$$

by the equality (4.3.3). In short, the Morozov's discrepancy principle seeks $\beta>0$ such that

$$
\phi\left(x_{\beta}, y\right)=\sigma
$$

where $\delta$ is the noise level (or performance level), $x_{\beta}$ is the optimal solution, and $c \geq 1$. Obviously, one needs a priori information about the noise level in order to effectively compute the regularization parameter $\beta$.

### 4.3.2 Complexity Level

A similar approach can be taken based on the complexity level (or sparsity level) $\gamma>0$ where it is desired that

$$
\psi(x) \leq \gamma
$$

Similarly, this principle has the variational formulation, where one seeks $\beta>0$ such that

$$
\beta \rightarrow F(\beta)-\gamma \beta
$$

is maximized. That is, if $F$ is differentiable at the optimal $\beta$ then from (4.3.3),(4.3.4)

$$
\psi\left(x_{\beta}\right)=\gamma
$$

### 4.3.3 Balance principle

We present here the balance principle for the single-term regularization, as well as some theoretical results, however similar results can be obtained for multi-term regularization as can be found in [25]. The previous selection rules were based on either the performance level (noise) or the complexity level. The selection rule developed in this section is based on balancing the performance level and the complexity level. Consider maximizing the conditional density $p((x, \tau, \lambda) \mid y) \sim p(y \mid(x, \tau, \lambda)) p(x, \tau, \lambda)$ where $(\tau, \lambda)$ are density functions for $\phi, \psi$, respectively, both having Gamma distribution. The balancing principle is derived from the Bayesian inference [33]

$$
\min _{(x, \tau, \lambda)} \tau \phi(x, y)+\lambda \psi(x)+\tilde{\beta}_{0} \lambda-\tilde{\alpha}_{0} \ln \lambda+\tilde{\beta}_{1} \tau-\tilde{\alpha}_{1} \ln \tau
$$

Letting $\beta=\frac{\lambda}{\tau}$, the necessary optimality is given by

$$
\begin{aligned}
& x_{\beta}=\underset{x}{\arg \min }\{\phi(x, y)+\beta \psi(x)\} \\
& \lambda=\frac{\tilde{\alpha}_{0}}{\psi\left(x_{\beta}\right)+\tilde{\beta}_{0}} \\
& \tau=\frac{\tilde{\alpha}_{1}}{\phi\left(x_{\beta}\right)+\tilde{\beta}_{1}}
\end{aligned}
$$

or

$$
\begin{gather*}
x_{\beta}=\underset{x}{\arg \min }\{\phi(x, y)+\beta \psi(x)\}  \tag{4.3.7}\\
\beta=\frac{1}{\mu} \frac{\phi\left(x_{\beta}\right)+\tilde{\beta}_{0}}{\psi\left(x_{\beta}\right)+\tilde{\beta}_{1}}, \quad \mu=\frac{\tilde{\alpha}_{1}}{\tilde{\alpha}_{0} .} \tag{4.3.8}
\end{gather*}
$$

The Bayesian inference selection corresponds to minimization of the functional

$$
\begin{equation*}
\frac{F\left(x_{\beta}\right)^{1+c}}{\beta} \tag{4.3.9}
\end{equation*}
$$

The minimum occurs when $\frac{d}{d \beta} F\left(x_{\beta}\right)=0$ so that

$$
\begin{equation*}
\frac{F\left(x_{\beta}\right)^{c}\left((1+c) F^{\prime}\left(x_{\beta}\right) \beta-F\left(x_{\beta}\right)\right)}{\beta^{2}}=0 \tag{4.3.10}
\end{equation*}
$$

which implies

$$
\begin{equation*}
\phi\left(x_{\beta}, y^{\delta}\right)=c \beta \psi\left(x_{\beta}\right) \tag{4.3.11}
\end{equation*}
$$

since $F^{\prime}\left(x_{\beta}\right)=\psi\left(x_{\beta}\right)$. The relation (4.3.11) is what we refer to as the balance principle for choosing the parameter $\beta$.

The natural choice for updating the parameter $\beta$ is the fixed point iterate

$$
\begin{equation*}
\beta^{+}=\frac{1}{\gamma} \frac{\phi\left(x_{\beta}, y^{\delta}\right)}{\psi\left(x_{\beta}\right)} \tag{4.3.12}
\end{equation*}
$$

where $x_{\beta}$ is the solution with the previous value for $\beta$. Here, $\gamma$ is selected by the two-step procedure proposed in [25, 26] :

- Choose $\gamma_{0}, \beta_{0}$
- Compute $x_{\beta}$ with chosen parameters
- Set $\gamma$ by

$$
\begin{equation*}
\gamma=\gamma_{0}\left(\frac{\phi\left(x_{\beta}, y^{\delta}\right)}{.05 \phi\left(0, y^{\delta}\right)}\right)^{d} \tag{4.3.13}
\end{equation*}
$$

with $d, \gamma_{0}$ heuristically chosen as $d=\frac{1}{4}, \gamma_{0}=10$.

The multi-parameter balance principle is obtained by minimizing

$$
\Phi_{\gamma}(\boldsymbol{\eta})=\frac{\gamma^{\gamma}}{(\gamma+2)^{\gamma+2}} \frac{F^{2+\gamma}(\boldsymbol{\eta})}{\prod_{i} \eta_{i}}
$$

where

$$
F(\boldsymbol{\eta})=\mathcal{J}_{\boldsymbol{\eta}}\left(x_{\boldsymbol{\eta}}\right)
$$

with $x_{\boldsymbol{\eta}}$ the minimizer of (4.2.11).
In $[25,26]$ a posterior estimate of the error using the balance principle in the dualparameter regularization is analyzed. Here, we are working in a Hilbert space setting and $\phi\left(x, y^{\delta}\right)=\frac{1}{2}\left\|K x-y^{\delta}\right\|^{2}, \psi_{1}(x)=\frac{1}{2}\left\|L_{1} x\right\|^{2}, \psi_{2}(x)=\frac{1}{2}\left\|L_{2} x\right\|^{2}$. For this purpose, we define the weighted norm

$$
\|x\|_{\boldsymbol{\eta}}^{2}=\frac{\eta_{1}}{\eta_{1}+\eta_{2}}\left\|L_{1} x\right\|^{2}+\frac{\eta_{1}}{\eta_{1}+\eta_{2}}\left\|L_{2} x\right\|^{2} .
$$

For notational convenience, we also define $Q_{\boldsymbol{\eta}}=\frac{\eta_{1}}{\eta_{1}+\eta_{2}} L_{1}^{*} L_{1}+\frac{\eta_{2}}{\eta_{1}+\eta_{2}} L_{2}^{*} L_{2}$ and $L_{\boldsymbol{\eta}}=$ $Q_{\boldsymbol{\eta}}^{\frac{1}{2}}, \tilde{K}_{\boldsymbol{\eta}}=K L_{\boldsymbol{\eta}}^{-1}$. We also take $\beta=\eta_{1}+\eta_{2}$ to simplify notation. We will utilize the interpolation inequality

$$
\begin{equation*}
\left\|\left(K^{*} K\right)^{r} x\right\| \leq\left\|\left(K^{*} K\right)^{q} x\right\|^{\frac{r}{q}}\|x\|^{1-\frac{r}{q}} . \tag{4.3.14}
\end{equation*}
$$

Theorem 4.3.1. Assume that the exact solution $x^{\dagger}$ satisfies

$$
L_{\boldsymbol{\eta}} x^{\dagger}=\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\eta}\right)^{\mu} w
$$

for some $w \in Y$. Suppose further that the regularization parameter $\boldsymbol{\eta}$ is chosen by the balance principle. Let $\delta_{*}=\left\|K x_{\eta^{*}}-y^{\delta}\right\|$. Then

$$
\left\|x_{\boldsymbol{\eta}^{*}}^{\delta}-x^{\dagger}\right\|_{\boldsymbol{\eta}^{*}} \leq C\left(1+\frac{F^{1+\frac{\gamma}{2}}\left(\delta^{\frac{1}{2 \mu+1}} \boldsymbol{e}\right)}{F^{1+\frac{\gamma}{2}}\left(\boldsymbol{\eta}^{*}\right)}\right) \max \left\{\delta_{*}, \delta\right\}^{\frac{2 \mu}{2 \mu+1}}
$$

Proof. As is often done, we decompose the error into propagation error and approximation error

$$
x_{\boldsymbol{\eta}^{*}}^{\delta}-x^{\dagger}=\left(x_{\boldsymbol{\eta}^{*}}^{\delta}-x_{\boldsymbol{\eta}^{*}}\right)+\left(x_{\boldsymbol{\eta}^{*}}-x^{\dagger}\right) .
$$

We first estimate the error, $x_{\boldsymbol{\eta}}^{\boldsymbol{\eta}}-x_{\boldsymbol{\eta}}$, due to noise propagation. By the necessary optimality, we have

$$
\left(K^{*} K+\eta_{1} L_{1}^{*} L_{1}+\eta_{2} L_{2}^{*} L_{2}\right)\left(x_{\boldsymbol{\eta}}-x_{\boldsymbol{\eta}}^{\delta}\right)=K^{*}\left(y-y^{\delta}\right) .
$$

From this relation, we multiply each side by $x_{\boldsymbol{\eta}}^{\boldsymbol{\eta}}-x_{\boldsymbol{\eta}}$ to obtain

$$
\begin{aligned}
&\left\|K\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right)\right\|^{2}+\eta_{1}\left\|L_{1}\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right)\right\|^{2}+\eta_{2}\left\|L_{2}\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right)\right\|^{2}=\left\langle K\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right), y-y^{\delta}\right\rangle \\
& \leq\left\|K\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right)\right\|^{2}+\frac{1}{4}\left\|y-y^{\delta}\right\|^{2} \\
& \Longrightarrow \eta_{1}\left\|L_{1}\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right)\right\|^{2}+\eta_{2}\left\|L_{2}\left(x_{\boldsymbol{\eta}}^{\delta}-x_{\boldsymbol{\eta}}\right)\right\|^{2} \leq \frac{1}{4}\left\|y-y^{\delta}\right\|^{2} .
\end{aligned}
$$

by use of the Cauchy-Schwarz inequality and the Young's inequality. Hence,

$$
\left\|x_{\boldsymbol{\eta}}\right\|_{\boldsymbol{\eta}} \leq \frac{\left\|y-y^{\delta}\right\|}{2 \sqrt{\eta_{1}+\eta_{2}}} \leq \frac{\delta}{2 \sqrt{\max \left\{\eta_{1}, \eta_{2}\right\}}}
$$

Since $\boldsymbol{\eta}^{*}$ minimizes the value functional, we have

$$
\frac{F^{2+\gamma}\left(\boldsymbol{\eta}^{*}\right)}{\max \left\{\eta_{1}^{*}, \eta_{2}^{*}\right\}^{2}} \leq \frac{F^{2+\gamma}\left(\boldsymbol{\eta}^{*}\right)}{\eta_{1}^{*} \eta_{2}^{*}} \leq \frac{F^{2+\gamma}(\boldsymbol{\eta})}{\eta_{1} \eta_{2}}
$$

for any $\boldsymbol{\eta}$. Selecting $\boldsymbol{\eta}=\left(\delta^{\frac{1}{2 \mu+1}}, 1\right)$ we have

$$
\begin{aligned}
& \frac{1}{\max \left\{\eta_{1}^{*}, \eta_{2}^{*}\right\}} \leq \frac{F^{2+\gamma}(\boldsymbol{\eta})}{F^{2+\gamma}\left(\boldsymbol{\eta}^{*}\right)} \delta^{-\frac{1}{2 \mu+1}} \\
& \Longrightarrow\left\|x_{\boldsymbol{\eta}^{*}}^{\delta}-x_{\boldsymbol{\eta}}\right\| \leq \frac{\left.F^{1+\frac{\gamma}{2}}\left(\delta^{\frac{1}{2 \mu+1}}, 1\right)\right)}{F^{1+\frac{\gamma}{2}}\left(\boldsymbol{\eta}^{*}\right)} \delta^{\frac{2 \mu}{2 \mu+1}} .
\end{aligned}
$$

Now, we estimate the approximation error $x_{\boldsymbol{\eta}}-x^{\dagger}$. We have

$$
\begin{aligned}
x_{\boldsymbol{\eta}}-x^{\dagger} & =\left(K^{*} K+\eta_{1} L_{1}^{*} L_{1}+\eta_{1} L_{2}^{*} L_{1}\right)^{-1}\left(\eta_{1} L_{1}^{*} L_{1}+\eta_{2} L_{2}^{*} L_{2}\right) x^{\dagger} \\
& =\beta\left(K^{*} K+\beta Q_{\boldsymbol{\eta}}\right)^{-1} Q_{\boldsymbol{\eta}} x^{\dagger} \\
& =\beta L_{\boldsymbol{\eta}}^{-1}\left(L_{\boldsymbol{\eta}}^{-1} K^{*} K L_{\boldsymbol{\eta}}^{-1}+\beta I\right)^{-1} L_{\boldsymbol{\eta}} x^{\dagger}
\end{aligned}
$$

which yields

$$
L_{\boldsymbol{\eta}}\left(x_{\boldsymbol{\eta}}-x^{\dagger}\right)=\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right) L_{\boldsymbol{\eta}} x^{\dagger} .
$$

Using the interpolation inequality (4.3.14), we have

$$
\begin{aligned}
\left\|x_{\boldsymbol{\eta}}-x^{\dagger}\right\|_{\boldsymbol{\eta}} & =\left\|L_{\boldsymbol{\eta}}\left(x_{\boldsymbol{\eta}}-x^{\dagger}\right)\right\|=\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right) L_{\boldsymbol{\eta}} x^{\dagger}\right\| \\
& =\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right)^{-1}\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}\right)^{\mu} w\right\| \\
& \leq\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right)^{-1}\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}\right)^{\frac{1}{2}+\mu} w\right\|^{\frac{2 \mu}{2 \mu+1}}\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right)^{-1} w\right\|^{\frac{1}{2 \mu+1}} \\
& =\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right)^{-1} \tilde{K}_{\boldsymbol{\eta}} L_{\boldsymbol{\eta}} x^{\dagger}\right\|^{\frac{2 \mu}{2 \mu+1}}\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right)^{-1} w\right\| \\
& \leq c\left(\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}} \tilde{K}_{\boldsymbol{\eta}}^{*}+\beta I\right)^{-1} y^{\delta}\right\|+\left\|\beta\left(\tilde{K}_{\boldsymbol{\eta}} \tilde{K}_{\boldsymbol{\eta}}^{*}+\beta I\right)^{-1}\left(y^{\delta}-y\right)\right\|\right)^{\frac{2 \mu}{2 \mu+1}}\|w\|^{\frac{1}{2 \mu+1}},
\end{aligned}
$$

where $c$ depends on the maximum of $r_{\beta}(t)=\frac{\beta}{\beta t}$ over $\left[0,\left\|\tilde{K}_{\eta}\right\|^{2}\right]$. Further, by noting that

$$
\begin{aligned}
\beta\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right)^{-1} y^{\delta} & =y^{\delta}-\left(\tilde{K}_{\boldsymbol{\eta}} \tilde{K}_{\boldsymbol{\eta}}^{*}+\beta I\right) \tilde{K}_{\boldsymbol{\eta}} \tilde{K}_{\boldsymbol{\eta}}^{*} y^{\delta} \\
& =y^{\delta}-\tilde{K}\left(\tilde{K}_{\boldsymbol{\eta}}^{*} \tilde{K}_{\boldsymbol{\eta}}+\beta I\right) \tilde{K}_{\boldsymbol{\eta}}^{*} y^{\delta} \\
& =y^{\delta}-K\left(K^{*} K+\beta Q_{\boldsymbol{\eta}}\right)^{-1} K^{*} y^{\delta} \\
& =y^{\delta}-K x_{\boldsymbol{\eta}}^{\delta}
\end{aligned}
$$

we have

$$
\left\|x_{\boldsymbol{\eta}^{*}}-x^{\dagger}\right\| \leq c\left(\delta_{*}+\delta\right)^{\frac{1 \mu}{2 \mu+1}}\|w\|^{\frac{1}{2 \mu+1}} \leq c_{1} \max \left\{\delta_{*}, \delta\right\}^{\frac{2 \mu}{2 \mu+1}}
$$

Thus,

$$
\left\|x_{\boldsymbol{\eta}^{*}}^{\delta}-x^{\dagger}\right\|_{\boldsymbol{\eta}^{*}} \leq C\left(1+\frac{F^{1+\frac{\gamma}{2}}\left(\delta^{\frac{1}{2 \mu+1}} \boldsymbol{e}\right)}{F^{1+\frac{\gamma}{2}}\left(\boldsymbol{\eta}^{*}\right)}\right) \max \left\{\delta_{*}, \delta\right\}^{\frac{2 \mu}{2 \mu+1}}
$$

A more general a posteriori error estimate is proven in [24] based on the Bregman distance. The following corollary follows from the above theorem and the error estimate obtained in Theorem 2.3.1 for the solution of $\mathscr{L} u_{k}=\varphi_{k}$.

Corollary 4.3.1. Take $\phi\left(u_{k}, \varphi_{k}\right)=\frac{1}{2}\left\|\mathscr{L} u_{k}-\varphi_{k}\right\|^{2}, \psi_{1}\left(u_{k}\right)=\frac{1}{2}\left\|L_{1} u_{k}\right\|^{2}, \psi_{2}\left(u_{k}\right)=\frac{1}{2}\left\|L_{2} u_{k}\right\|^{2}$. Let $\delta_{*}=\left\|\mathscr{L} u_{k}^{*}-\varphi_{k}\right\|$ where $u_{k}^{*}$ is the approximate solution to (2.3.13) corresponding to $\boldsymbol{\eta}^{*}$ chosen by the balance principle. Then the estimated initial condition of (2.3.2) satisfies
the error estimate

$$
\left\|x_{0}-x_{\delta}^{m}\right\| \leq\left\|x_{0}-x^{n}\right\|_{X}+\sum_{k=0}^{m}\left(\delta\left\|u_{k}^{\dagger}\right\|+c_{1} \delta_{*}^{\frac{2 \mu}{2 \mu+1}}\left\|x_{0}\right\|\right)
$$

where $u_{k}^{\dagger}$ is the exact solution satisfying the source condition

$$
L_{\eta} u_{k}^{\dagger}=\left(\tilde{\mathscr{L}}_{\eta}^{*} \tilde{\mathscr{L}}_{\eta}\right)^{\mu} w
$$

for some $w \in X, \mu>0$, and $\delta$ is the noise level of $y$.
This error estimate shows that the overall error using the dual control method coupled with the multi-parameter regularization is affected by the accuracy $\delta_{*}$ and the noise level $\delta$ in $y$, since the "data" for $u$ is noiseless (i.e., we know the basis exactly).

## Chapter 5

## Nonsmooth regularization and Sparsity Optimization

In this chapter, we formulate methods for solving the inverse problem for cases involving nonsmooth functionals $\psi$ in the formulation of the regularization methods 4.2.9 and 4.2.11. We do not describe the methods in full detail, however, the computation of the minimizer of a nonsmooth functional is necessary for the application of the proposed methods. Further details can be found in [29]. An example of a nonsmooth functional for the multi-parameter case corresponds to the choices of the $L^{1}$ norm and $H^{1}$ semi-norm for $\psi_{1}, \psi_{2}$, that is, when we seek a minimum of

$$
\begin{equation*}
\mathcal{J}_{\beta}(x)=\frac{1}{2} \int_{\Omega}\left|K u-y^{\delta}\right|^{2} d \xi+\eta_{1} \int_{\Omega}|u| d \xi+\frac{\eta_{2}}{2} \int_{\Omega}|\nabla u|^{2} d \xi \tag{5.0.1}
\end{equation*}
$$

for which a minimum occurs when the (formal) necessary optimality

$$
\begin{equation*}
K^{*}\left(K u-y^{\delta}\right)+\eta_{1} \frac{u}{|u|}-\eta_{2} \Delta u=0 \tag{5.0.2}
\end{equation*}
$$

is satisfied.
We begin by describing the sparsity optimization, that is, formulations involving the $\ell_{p}$ norm for $0 \leq p \leq 1$. As already pointed out, one advantage of the standard Tikhonov regularization is the closed form solution. However, sparsity optimization is becoming increasingly important for practical applications. Due to the nonsmoothness of the norms $\|\cdot\|_{p}, 0 \leq p \leq 1$, methods capable of handling the nonsmoothness are necessary
for computing minimizers involving these norms. We use the word sparsity due to the fact

$$
\|u\|_{p} \rightarrow \# \text { of nonzero elements of } u \text { as } p \rightarrow 0^{+}
$$

Thus, the use of $\|\cdot\|_{p}$ enhances sparsity in the solution $u$ as $p \rightarrow 0^{+}$. In other words, the choice of the $p$-norm with $p \leq 1$ removes information in the approximate solution that is unnecessary.

One can develop the necessary optimality condition for

$$
\begin{equation*}
\min \left\|K u-y^{\delta}\right\|_{2}+\beta\|u\|_{p}, \quad 0 \leq p \leq 1 \tag{5.0.3}
\end{equation*}
$$

despite the fact that $\|u\|_{p}$ is not differentiable at $u=0$. Specifically, let us consider the case when $p=1$ for a scalar $u$. For $p=1$, we have

$$
\partial\|u\|=\frac{u}{\|u\|}
$$

at $u \neq 0$, however for $u=0$ we have the subdifferential

$$
\partial\|u\|=[-1,1]
$$

where the subdifferential of a functional $f: X \rightarrow(-\infty, \infty]$ at $u \in X$ is defined as the set

$$
\begin{equation*}
\left\{u^{*} \in X^{*} \mid f(z) \geq f(u)+\left\langle u^{*}, z-u\right\rangle, \forall z \in X\right\} \tag{5.0.4}
\end{equation*}
$$

which is more precisely defined in [29, 44] for example. However, for $p<1$ we have $\partial\|u\|_{p}^{p}=\emptyset$ when $u=0$. To remedy this, for $\varepsilon \ll 1$, we take the approximation

$$
\begin{equation*}
\partial_{\varepsilon}\|u\|_{p}^{p}=\frac{p u}{\max \left(\varepsilon^{2-p},\|u\|^{2-p}\right)} \approx \partial\|u\|_{p}^{p} \tag{5.0.5}
\end{equation*}
$$

which approximates the formal derivative

$$
\begin{equation*}
\partial\|u\|_{p}^{p}=\frac{u}{\|u\|^{2-p}} \tag{5.0.6}
\end{equation*}
$$

for any value of $u$. One can see a depiction of the approximate derivative $\partial_{\varepsilon}\|u\|$ in Figure 5.0.1 with $\varepsilon=10 e^{-3}$ and $\varepsilon=10 e^{-2}$.

It can be shown that this sequence converges to the minimizer of the appropriate cost


Figure 5.0.1: Comparison of approximate subdifferentials for $\varepsilon=10 e^{-3}$ and $\varepsilon=10 e^{-2}$.
functional. We consider the regularized problem of the form

$$
\begin{equation*}
J_{\varepsilon}(u)=\frac{1}{2}\left\|K u-y^{\delta}\right\|^{2}+\Psi_{\varepsilon}\left(\|u\|^{2}\right) \tag{5.0.7}
\end{equation*}
$$

where for $t \geq 0$

$$
\Psi_{\varepsilon}(t)= \begin{cases}\frac{p}{2} \frac{t}{\varepsilon^{2-p}}+\left(1-\frac{p}{2}\right) \varepsilon^{p} & t \leq \varepsilon^{2} \\ t^{\frac{p}{2}} & t \geq \varepsilon^{2}\end{cases}
$$

For $\varepsilon>0$, consider the iterative algorithm of the form

$$
\begin{equation*}
K^{*} K u^{k+1}+\frac{\beta p}{\max \left(\varepsilon^{2-p},\left\|u^{k}\right\|^{2-p}\right)} u^{k+1}=K^{*} y^{\delta} \tag{5.0.8}
\end{equation*}
$$

to find the minimizer of (5.0.7). Multiplying (5.0.8) by $u^{k+1}-u^{k}$, we obtain

$$
\begin{aligned}
& \left.\frac{1}{2}\left(K u^{k+1}, u^{k+1}\right)-\left(K u^{k}, u^{k}\right)+\left(K\left(u^{k+1}-u^{k}\right), u^{k+1}-u^{k}\right)\right) \\
& \quad+\frac{\beta p}{\max \left(\varepsilon^{2-p},\left\|u^{k}\right\|^{2-p}\right)} \frac{1}{2}\left(\left\|u^{k+1}\right\|^{2}-\left\|u^{k}\right\|^{2}+\left\|u^{k+1}-u^{k}\right\|^{2}\right)+\left(y^{\delta}, u^{k+1}-u^{k}\right)
\end{aligned}
$$

Then,

$$
\frac{1}{\max \left(\varepsilon^{2-p},\left\|u^{k}\right\|^{2-p}\right)} \frac{p}{2}\left(\left\|u^{k+1}\right\|^{2}-\left\|u^{k}\right\|^{2}\right)=\Psi_{\varepsilon}^{\prime}\left(\left\|u^{k}\right\|^{2}\right)\left(\left\|u^{k+1}\right\|^{2}-\left\|u^{k}\right\|^{2}\right)
$$

Since $t \rightarrow \Psi_{\varepsilon}(t)$ is concave, we have

$$
\Psi_{\varepsilon}\left(\left\|u^{k+1}\right\|^{2}\right)-\Psi_{\varepsilon}\left(\left\|u^{k}\right\|^{2}\right)-\frac{1}{\max \left(\varepsilon^{2-p},\left\|u^{k}\right\|^{2-p}\right)} \frac{p}{2}\left(\left\|u^{k+1}\right\|^{2}-\left\|u^{k}\right\|^{2}\right) \leq 0
$$

and thus

$$
\begin{equation*}
\left.J_{\varepsilon}\left(u^{k+1}\right)+\frac{1}{2}\left(K\left(u^{k+1}-u^{k}\right), u^{k+1}-u^{k}\right)\right)+\frac{\beta p}{\max \left(\varepsilon^{2-p},\left\|u^{k}\right\|^{2-p}\right)} \frac{1}{2}\left\|u^{k+1}-u^{k}\right\|^{2} \leq J_{\varepsilon}\left(u^{k}\right) \tag{5.0.9}
\end{equation*}
$$

shows that $J_{\varepsilon}$ is non-increasing. Now, we give the following result.
Theorem 5.0.2. For $\varepsilon>0$ let $\left\{u_{k}\right\}$ be generated by (5.0.8). Then, $J_{\varepsilon}\left(u^{k}\right)$ is monotonically non increasing and $u_{k}$ converges to the minimizer of $J_{\varepsilon}$ defined by (5.0.7).

Proof. The monotonicity of $J_{\varepsilon}$ has already been shown. Thus, we show that $\left\{u^{k}\right\}$ converges to the minimizer of $J_{\varepsilon}$. It follows from (5.0.7) that $\left|u^{k}\right|_{\infty}<\infty$ and

$$
\sum_{k=0}^{\infty}\left\|u^{k+1}-u^{k}\right\|_{2}^{2}<\infty
$$

and thus there exists a subsequence of $\left\{u^{k}\right\}$ and $u^{*} \in \ell^{p}$ such that

$$
\lim _{k \rightarrow \infty} u_{k}=\lim _{k \rightarrow \infty} u^{k+1}=u^{*}
$$

It follows from (5.0.8) that

$$
K^{*} K u^{*}+\frac{\beta p}{\max \left(\varepsilon^{2-p},\left\|u^{*}\right\|^{2-p}\right)} u^{*}=K^{*} y^{\delta}
$$

i.e., $u^{*}$ minimizes $J_{\varepsilon}$.

By the assumption of sparsity, the solution $u=0$ everywhere except on a set

$$
\begin{equation*}
S=\{u: u \neq 0\} \tag{5.0.10}
\end{equation*}
$$

where $\mu(S)<\delta$ for some measure $\mu$ and a small number $\delta$. That is, $S=\operatorname{supp}(u)$, and $u$ has compact support. Since $K^{*} K$ may be ill-conditioned over $S \cup S^{c}$, we develop methods for solving (4.2.10) so that when the problem is restricted to the set $S$ we obtain a better-posed problem.

We are motivated to consider functions $u$ with compact support for two reasons. Considering the applications of interest, we assume that the set $S$ is of very small dimension compared with the dimension of the original problem. Hence, we will be solving a system of linear equations of reduced order. Therefore, we seek algorithms that will take advantage of the reduced order with respect to the set $S$. Secondly, the restriction of the operator $K K^{*}$ to $S$ will be less ill-conditioned, which will improve the computational accuracy of the approximate solution.

### 5.1 Primal-Dual Active Set Method for unilateral constraint

In this section, we discuss the basic concepts of the Primal-Dual Active set method, and show how the sparsity optimization can be cast in this framework. This method is not a contribution of this thesis, however, it is a useful tool for computing the minimizer of a nonsmooth functional. In general, the Primal-Dual Active Set method allows for approximations to numerous nonsmooth optimization problems to be efficiently computed. The method is especially useful for large scale problems due to the reduced computational expense associated with the inactive set.

Prior to applying the method, we briefly discuss the Primal-Dual Active set method for a general class of problems given by

$$
\left\{\begin{array}{l}
\min _{u \in X} \frac{1}{2}\langle A u, u\rangle-\langle a, u\rangle  \tag{5.1.1}\\
\text { subject to } u \leq \psi
\end{array}\right.
$$

This method is formulated in detail, with more general constraints, in the monograph [29],the paper [20] and the references therein. For the inequality constraint in (5.1.1), we define the complementarity condition as

$$
\begin{equation*}
\mu=\max \{0, \mu+c(u-\psi)\} \tag{5.1.2}
\end{equation*}
$$

for the Lagrange multiplier $\mu$, so that we have the system

$$
\begin{align*}
& A u+\mu=a  \tag{5.1.3}\\
& \mu=\max \{0, \mu+c(u-\psi)\} \tag{5.1.4}
\end{align*}
$$

For this general class of problems, the active set is given by

$$
\begin{equation*}
\mathcal{A}=\{\mu+c(u-\psi)>0\} \tag{5.1.5}
\end{equation*}
$$

which, if known, yields the system of equations

$$
\begin{align*}
& A u+\mu=a  \tag{5.1.6}\\
& u=\psi \text { in } \mathcal{A} \quad \text { and } \quad \mu=0 \text { in } \mathcal{A}^{c} \tag{5.1.7}
\end{align*}
$$

The following Newton-like method(which can be shown to be a semi-smooth Newton method) is given in [29]

1. Initialize $u^{0}, \mu^{0}$. Set $k=0$.
2. Set $\mathcal{I}_{k}=\left\{\mu^{k}+c\left(u^{k}-\psi\right) \leq 0\right\}, \mathcal{A}_{k}=\left\{\mu^{k}+c\left(u^{k}-\psi\right)>0\right\}$
3. Solve for $\left(u^{k+1}, \lambda^{k+1}, \mu^{k+1}\right)$ :

$$
\begin{align*}
& A u^{k+1}+\mu^{k+1}=a,  \tag{5.1.8}\\
& u^{k+1}=\psi \operatorname{in} \mathcal{A}_{k} \quad \text { and } \quad \mu^{k+1}=0 \operatorname{in} \mathcal{I}_{k} \tag{5.1.9}
\end{align*}
$$

4. Stop, or set $k=k+1$, and return to 2 .

## Primal-Dual Active set method for $L^{1}-H^{1}$ minimization

Consider minimizing the cost functional

$$
\left\|K u-y^{\delta}\right\|_{L^{2}}^{2}+\eta_{1}\|u\|_{L^{1}}+\eta_{2}\|\nabla u\|_{L^{2}}^{2}
$$

where $\|\cdot\|_{L^{2}}^{2}$ corresponds to the $H^{1}$ semi-norm. As an example, we formulate the PrimalDual method in terms of one-dimensional $u$. Extending the method to higher dimensions
is straight-forward. We must formally solve the Euler-Lagrange necessary optimality

$$
\begin{equation*}
K K^{*} u-K^{*} y^{\delta}+\eta_{1} \frac{u}{|u|}-\eta_{2} u_{x x}=0 \tag{5.1.10}
\end{equation*}
$$

More precisely, this can be formulated as

$$
\begin{align*}
K K^{*} u-K^{*} y^{\delta}+\eta_{1} \lambda-\eta_{2} u_{x x} & =0  \tag{5.1.11}\\
\lambda|u| & =u \quad|\lambda| \leq 1 \tag{5.1.12}
\end{align*}
$$

where the selection of $\lambda$ is equivalent to the complementarity condition

$$
\begin{equation*}
\lambda=\frac{\lambda+c u}{\max \{1,|\lambda+c u|\}} . \tag{5.1.13}
\end{equation*}
$$

To see this, first suppose that $|\lambda+c u| \leq 1$ so that $\max \{|\lambda+c u|, 1\}=1$. Then we have that

$$
\lambda=\lambda+c u \Longrightarrow u=0
$$

That is, we obtain the set $S^{c}$ given by (5.0.10) whenever $|\lambda+c u| \leq 1$. Now, let $|\lambda+c u|>1$, so that

$$
\lambda=\frac{\lambda+c u}{|\lambda+c u|} \Longrightarrow \lambda\left(|\lambda+c u|_{1}-1\right)=c u
$$

which implies that

$$
\alpha \lambda=c u
$$

where $\alpha>0$ is given by $\alpha=|\lambda+c u|-1$. Thus, we have that $\lambda$ is proportional to $u$. This yields

$$
\lambda=\frac{\lambda+c u}{|\lambda+c u|}=\frac{\lambda(1+\alpha)}{|\lambda|(1+\alpha)}=\frac{\lambda}{|\lambda| .}
$$

Finally, since $c$ is arbitrary this yields

$$
\begin{equation*}
\lambda=\frac{u}{|u| .} \tag{5.1.14}
\end{equation*}
$$

We can make use of the Primal-Dual Active set method, in which we solve for the primal and dual functions, $u$ and $\lambda$, respectively, when we are in the active set. Now, consider
solving the complementarity condition

$$
\begin{equation*}
\lambda=\frac{\lambda+c u}{\max \{1,|\lambda+c u|\}} \tag{5.1.15}
\end{equation*}
$$

by Newton's method. That is, for the solution of the equation $F(y)=0$ we compute the roots of

$$
\begin{equation*}
F^{\prime}(y) \delta y+F(y)=0 \tag{5.1.16}
\end{equation*}
$$

In our case, we look for roots of the equation

$$
\begin{equation*}
\delta \lambda \max \{1,|\lambda+c u|\}+\lambda \cdot 0-\delta \lambda-c \delta u+\lambda \max \{1,|\lambda+c u|\}-(\lambda+c u) . \tag{5.1.17}
\end{equation*}
$$

We will consider the Newton's method for two cases. Firstly, let $|\lambda+c u|<1$ so that $\max \{1,|\lambda+c u|\}=1$ which implies that

$$
\delta u+u=0
$$

from (5.1.17). When discretized, the condition becomes

$$
u^{k+1}-u^{k}+u^{k}=0
$$

so that $u^{k+1}=0$. That is, we can set $u=0$ whenever we are solving in the inactive set. Now, for the case when we are in the set $A_{k}=\left\{\left|\lambda^{k}+c u^{k}\right|>1\right\}$ we discuss methods for computing $\lambda$ based on the Primal-Dual Active set formulation. Define

$$
\begin{equation*}
d^{k}=\left|\lambda^{k}+c u^{k}\right|, \quad F^{k}=a^{k}\left(b^{k}\right)^{t}, \quad a^{k}=\frac{\lambda^{k}}{\max \left\{1,\left|\lambda^{k}\right|\right\}}, \quad b=\frac{\lambda^{k}+c u^{k}}{\left|\lambda^{k}+c u^{k}\right|} \tag{5.1.18}
\end{equation*}
$$

For more details concerning the Primal-Dual Active Set method one may refer to the monograph [29] or the paper [20]. The method is developed in a much more general context in the references, which also include the theoretical analysis of the method.

With these terms (5.1.18) defined, the Primal-Dual Active Set method for the $L^{1}-H^{1}$ regularization is summarized by the following:

## Primal-Dual Active Set method for $L^{1}-H^{1}$ regularization:

1. Initialize $u^{0}, \lambda^{0}$. Set $k=0$.
2. Set $I_{k}=\left\{\left|\lambda^{k}+c u^{k}\right|<1\right\}, \quad A_{k}=\left\{\left|\lambda^{k}+c u^{k}\right|>1\right\}$.
3. Solve for $u^{k+1}, \lambda^{k+1}$ :

$$
\begin{aligned}
& \lambda^{k+1}=\frac{1}{d^{k}-1}\left(I-F^{k}\right) u^{k+1}+\frac{\lambda^{k}}{\max \left\{1,\left|\lambda^{k}\right|\right\}} \\
& K K^{*} u^{k+1}+\eta_{1}\left(\frac{1}{d^{k}-1}\left(I-F^{k}\right) c u^{k+1}\right)+\frac{\lambda^{k}}{\max \left\{1,\left|\lambda^{k}\right|\right\}}-\eta_{2} u_{x x}^{k+1}=K^{*} y^{\delta}
\end{aligned}
$$

4. Stop or set $k=k+1$ and return to 2

## Chapter 6

## Numerical Tests

### 6.1 1-D Diffusion Equation

In this section, we consider inverse problems involving the 1-D diffusion equation

$$
\begin{align*}
& \frac{\partial v}{\partial t}=\frac{\partial}{\partial x}\left(d(x) \frac{\partial v}{\partial x}\right)+f(x) \quad x \in \Omega \subset \mathbb{R}  \tag{6.1.1}\\
& v(0, t)=0=v(1, t)  \tag{6.1.3}\\
& y(t)=C v(t)
\end{align*}
$$

with Dirichlet boundary conditions, where the measurements are restricted to a subinterval $\Omega_{s} \subset \Omega$, for the time $0 \leq t \leq 1$. For all the 1-D simulations presented here, the domain is $\Omega=[0,1]$. The thermal conductivity, $d$, is potentially variable in space, but known. The 1-D diffusion equation is formulated as an abstract Cauchy problem (2.2.1) where

$$
A v=\frac{d}{d x}\left(d(x) \frac{d v}{d x}\right)
$$

and

$$
\begin{aligned}
& \operatorname{dom}(A)=\left\{v \in L^{2}(0,1) \mid v, \frac{d v}{d x}\right. \text { are absolutely continuous, } \\
&\left.\frac{d^{2} v}{d x^{2}} \in L^{2}(0,1) \text { and } v(0)=0=v(1)\right\}
\end{aligned}
$$

It can be seen in the Appendix A that this generates a strongly continuous semigroup.

## Example 1 : Simple 1-D example

For this simple 1-D case, we assume a Gaussian initial condition of the form

$$
v_{0}(x)=e^{-10^{2}(x-.7)^{2}}
$$

and we take the basis $\left\{\varphi_{k}\right\}_{k=0}^{m}=\{\sin (k \pi x)\}_{k=0}^{m}$. The observation operator is defined by

$$
C v(t)=\frac{1}{\mu\left(\Omega_{s}\right)} \int_{\Omega_{s}} v(s) d \mu
$$

where $\Omega_{s}=\Omega_{1} \cup \Omega_{2}$. Here, $\mu\left(\Omega_{s}\right)$ is the volume of the set $\Omega_{s}$, so that the measurements represent averages over the two regions $\Omega_{1}, \Omega_{2}$. Specifically, we take $\Omega_{1}=\left[\frac{3}{10}, \frac{4}{10}\right], \Omega_{2}=$ $\left[\frac{6}{10}, \frac{7}{10}\right]$ so that

$$
C v(t)=10 \int_{.3}^{.4} v(x, t) d x+10 \int_{.6}^{.7} v(x, t) d x
$$

An example of the measurement locations and the corresponding measurements can be seen in Figure 6.1.1. We then solve for the controls $u_{k}$ using the $L^{1}-H^{1}$ regularization


Figure 6.1.1: (a) Location of measurements; (b) Corresponding measurements, $y$.
method. For the case when the data is exact $(\delta=0)$ we obtain a nearly exact reconstruction with the parameters $\eta_{1}=1 \times 10^{-8}, \eta_{2}=1 \times 10^{-16}$, by computing $m=10$ coefficients. Contaminating the data with $10 \%$ noise, we obtain a reasonable reconstruction with the parameters $\eta_{1}=1 \times 10^{-7}, \eta_{2}=1 \times 10^{-7}$, by computing $m=6$ coefficients. As one might expect, computation of Fourier coefficients at high frequencies is more sensitive when the noise in the data is significant. Fortunately, the dual control method works well with few coefficients. Plots of the approximate solutions with the exact solution can be seen in Figures 6.1.2a, 6.1.2b. Note that the results are based on unfiltered data. The tuning of the parameters $\eta_{1}, \eta_{2}$ provides the filtering based on the smoothness constraint of the control $u$. As discussed in Chapter 2, the control set $\left\{u_{k}\right\}_{k=0}^{m}$ provides a filter bank that can be stored for estimating solutions based on given data $y^{\delta}$. An example of a control is depicted in Figure 6.1.3, corresponding to the basis function $\sin (2 \pi x)$.


Figure 6.1.2: (a) Reconstruction with exact measurements; (b) Reconstruction with $10 \%$ noise in measurements.

## Example 2: Spatially varying diffusion coefficient

For this example, we consider the case when the thermal conductivity is spatially variable. In particular, we take

$$
d(x)=1.0625-\left(x-\frac{1}{2}\right)^{4}
$$



Figure 6.1.3: Control corresponding to $\sin (2 \pi x)$.
and the initial condition is given by

$$
v_{0}(x)=e^{-200\left(x-\frac{1}{2}\right)^{4}}
$$

As in the first example, we take the basis $\varphi_{k}=\{\sin (k \pi x)\}_{k=0}^{m}$ and solve for the controls $u_{k}$ using the $L^{1}-H^{1}$ regularization method. Assuming a noise level of $10 \%$, we obtain a reasonable reconstruction with the parameters $\eta_{1}=5 \times 10^{-8}, \eta_{2}=1 \times 10^{-10}$, by computing $m=8$ coefficients. The corresponding results are depicted in Figures 6.1.4. It should be pointed out that the abstract Cauchy based dual control method does not make any assumptions on the coefficients of the PDE.

## Comparison of basis choices

Here, we compare the reconstructions obtained by two different basis choices. For this example, we take

$$
d(x)= \begin{cases}1 \frac{5}{16}-5\left(x-\frac{1}{2}\right)^{4}, & 0 \leq x<\frac{1}{2} \\ 1 \frac{3}{16}+\frac{1}{8+e^{-50(x-.65)}}, & \frac{1}{2} \leq x \leq 1\end{cases}
$$

as depicted in Figure 6.1.5, and the initial condition is given by

$$
v_{0}(x)=e^{-200\left(x-\frac{1}{2}\right)^{4}}
$$



Figure 6.1.4: (a) Thermal conductivity; (b) Reconstruction with $10 \%$ noise in measurements versus Exact initial condition.

We solve for the controls $u_{k}$ using the $L^{1}-H^{1}$ regularization method. Assuming a relative noise level of $5 \%$, we obtain a reasonable reconstruction with the parameters $\eta_{1}=5 \times 10^{-7}, \eta_{2}=1 \times 10^{-10}$, by computing $m=8$ coefficients, using Daubechies wavelets. In Figure 6.1.6, one can see a comparison of two reconstructions using a standard sine


Figure 6.1.5: Thermal conductivity.
basis and Daubechies-18 wavelets. The reconstruction obtained using the Daubechies
wavelets is much better than the sine basis reconstruction, using the same regularization parameters. This example illustrates how the basis choice affects the resulting reconstruction.


Figure 6.1.6: (a) Reconstructed initial condition using Daubechies- 18 wavelets with $\eta_{1}=$ $5 \times 10^{-7}, \eta_{2}=1 \times 10^{-10}$; (b) Reconstructed initial condition using sine basis with $\eta_{1}=$ $5 \times 10^{-7}, \eta_{2}=1 \times 10^{-10}$.

## Example 3 : Simultaneous reconstruction and parameter identification

We now provide results for simultaneously reconstructing the initial condition and the unknown constant $d(x) \equiv d$ for the diffusion equation. The exact initial condition is given by

$$
e^{-100(x-.7)^{2}}+.4 e^{-20(x-.3)^{2}} .
$$

The exact value of $d$ is taken to be .025 , and we assume an initial parameter of $d_{0}=.01$ to compute the minimizer of 3.1.4 via the gradient method, with the approximations 3.1.7. The gradient method converges with a tolerance of $9.94 \times^{-8}$ in 272 iterations, using a step size of $\gamma=\frac{1}{2}$. The results are depicted in Figures 6.1.7, 6.1.8 where one can see that the gradient method converges rapidly in this example. The results are promising, however
it is necessary to have a good initial guess for the parameter $d$ due to local minima.


Figure 6.1.7: (a) Convergence of $\nabla J$ to 0 ; (b) Convergence of $d$ to .025 .


Figure 6.1.8: (a) Reconstructed solution at convergence of $d$; (b) Cost functional $\mathcal{J}_{\beta}$ depicted with optimal value $d=.025$.

## Example 4 : Point source identification from randomly distributed measurements

In this example, we consider the identification of point sources from randomly distributed point measurements of the solution at the final time. The approach is based on the regularized least squares method utilizing the sparsity optimization, where we represent the source in a Haar basis. We consider the problem of determining the unknown source location and intensity for the function, $f$, in the diffusion equation (6.1.1), where we assume $f$ is of the form

$$
f(x)=\sum_{k=1}^{N} \beta_{k} \delta\left(x-a_{k}\right)
$$

and the diffusivity, $d$, is known. Here, $a_{1}, \ldots, a_{N} \in \mathbb{R}$ are mutually distinct and the intensities $\beta_{1}, \ldots, \beta_{N} \in \mathbb{R}$ are unknown. For our purposes, $\delta$ is defined by

$$
\delta\left(x-a_{k}\right)= \begin{cases}1, & x=a_{k} \\ 0, & \text { elsewhere }\end{cases}
$$

The problem is to determine the $N$ unknown source locations $\left\{a_{k}\right\}_{k=1}^{N}$ and the unknown source intensities $\left\{\beta_{k}\right\}_{k=1}^{N}$ from the observation data $y(t)=C v(t)$. We transform the source identification problem into an initial condition estimation problem, as described in Section 3.2.

For point source reconstruction, we consider the method developed in Section 2.3 with a sparse basis, such as the Haar or Daubechies D2, generated by the mother wavelet

$$
\psi(x)= \begin{cases}1 & 0 \leq x<\frac{1}{2} \\ -1 & \frac{1}{2} \leq x<1 \\ 0 & \text { otherwise }\end{cases}
$$

That is, we formulate the problem as

$$
\min _{v \in V_{m}} \frac{1}{2}\left\|\mathcal{M} v-y^{\delta}\right\|_{2}^{2}+\eta\|v\|_{1} .
$$

where $V_{m}=\operatorname{span}\left\{\psi_{j, k}(x)\right\}$. It is well known that $\left\{\psi_{j, k}(x)\right\}$ is an orthonormal basis for $L^{2}(\mathbb{R})$ [19].

In this simulation, the exact point source locations are taken at $a_{1}=.4$ and $a_{2}=.7$
with the respective intensities $\beta_{1}=10, \beta_{2}=30$. We only consider two point sources for clarity in the results, however similar results have been obtained with more point sources. Here, we take the observation operator

$$
C v(t)= \begin{cases}0 & 0 \leq t<t_{f} \\ \left.v\right|_{\Omega_{s}} & t=t_{f}\end{cases}
$$

where $\Omega_{s}=\left\{x_{1}, \ldots, x_{M}\right\}$ for $M=16$ randomly selected observation points. Note that we are assuming measurements are only available at the final time $t_{f}=1$, so that this inverse problem coincides with the backward heat equation with partially available final time data. The parameter, $\eta$, for the $L^{1}$ minimization is taken as $1 \times 10^{-4}$. This example is representative of the potential for sparsity optimization to improve results for state estimation. The sparse optimization allows for the resolution of the point sources without smearing, as well as accurately estimating the intensities.


Figure 6.1.9: (a) Point source identification with exact measurements; (b) Point source identification with $10 \%$ noise in measurements.

### 6.2 2-D Diffusion Equation

In this section, we consider inverse problems involving the 2-D diffusion equation

$$
\begin{align*}
& \frac{\partial v}{\partial t}(\boldsymbol{x}, t)=d \Delta v(\boldsymbol{x}, t)+f(\boldsymbol{x})  \tag{6.2.1}\\
& v(\boldsymbol{x}, t)=0 \quad x \in \partial \Omega  \tag{6.2.2}\\
& v(\boldsymbol{x}, 0)=v_{0}(\boldsymbol{x}) \tag{6.2.3}
\end{align*}
$$

where $\boldsymbol{x}=(x, y) \in \Omega \subset \mathbb{R}^{2}$. As in the 1-D case, we work on the time interval $0 \leq t \leq 1$.
The 2-D diffusion equation can be cast in the abstract Cauchy framework where $A$ coincides with the closure of the Laplace operator, defined by

$$
\Delta f(\boldsymbol{x}):=\frac{\partial^{2} f}{\partial x^{2}}+\frac{\partial^{2} f}{\partial y^{2}}
$$

Details for casting the diffusion equation in the abstract Cauchy framework are detailed in Appendix A.

## Example 1 : Simple 2-D example

Here, we assume a Gaussian initial condition of the form

$$
v_{0}(x, y)=e^{-10^{2}\left((x-.3)^{2}+(y-.3)^{2}\right)}
$$

and we take the basis $\varphi_{k, l}=\{\sin (k \pi x) \sin (l \pi y)\}_{k, l=0}^{m}$. We then solve for the controls $u_{k}$ using the $L^{1}-H^{1}$ regularization method. Assuming a noise level of $10 \%$, we obtain a reasonable reconstruction with the parameters $\eta_{1}=3 \times 10^{-10}, \eta_{2}=1 \times 10^{-7}$, by computing $m=12$ coefficients such that $(k+l)^{2} \leq 60$. In practice, the two-dimensional problems may be more sensitive to the noise in the data. In some cases, it may be desired to filter the data prior to computing the controls. To filter the data, we formulate the minimization problem

$$
\begin{equation*}
\min _{y \in Y_{m}} \frac{1}{2}\left\|y-y^{\delta}\right\|_{Y}^{2}+\eta\langle y, P y\rangle_{Y} \tag{6.2.4}
\end{equation*}
$$

where $P$ is often taken as a differential operator to enforce smoothing, such as the $H^{1}$ semi-norm, and $Y_{m}$ is a finite-dimensional subspace of $Y$. The results are depicted in Figure 6.2.1. Here, one can see that the location and intensity of the estimate is quite accurate, where the small oscillations are due to the basis choice.


Figure 6.2.1: (a) Exact initial condition; (b) Reconstruction with $10 \%$ noise in measurements.

## Example 2: Simultaneous state/parameter estimation

We now provide results for simultaneously reconstructing the initial condition and the unknown constant $d(x) \equiv d$ for the 2-D diffusion equation. The exact value of $d$ is taken to be .01 and

$$
v_{0}(x, y)=e^{-10^{2}\left((x-.5)^{2}+(y-.5)^{2}\right)}
$$

We assume an initial parameter of $d_{0}=.02$ to compute the minimizer of 3.1.4 via the MATLAB routine fminunc, with the approximations 3.1.7. The cost functional is depicted in Figure 6.2.2a, for which a minimum occurs at the value .01. The results are depicted in Figure 6.2.2b where the estimation is seen to be reasonably accurate.


Figure 6.2.2: (a) Cost functional for simultaneous RLLS method ; (b) Reconstructed initial condition at minimum $d=.01$.

## Example 3 : Point source identification

For this simulation, we consider the 2-D diffusion equation (6.2.1) where we assume $f$ is of the form

$$
f(x)=\sum_{k=1}^{N} \beta_{k} \delta\left(x-a_{k}\right) \in \mathscr{S}\left(\mathbb{R}^{2}\right)
$$

The task is to determine the source locations, $a_{k}$, and the source intensities $\beta_{k}$. We proceed by selecting the radial basis functions

$$
\varphi\left(r_{k}\right)=e^{-\mu\left\|x_{k}-c_{k}\right\|^{2}}
$$

for large $\mu$, in order to obtain a basis of approximate point sources. Using these basis functions, we obtain a positive definite kernel $\Phi$ and utilize the RKHS method presented in Section 3.2.1. Since the point sources are known to be discrete "delta" distributions (point sources), it is reasonable to use the $\ell_{p}$ norm, $0 \leq p \leq 1$ in the formulation (4.2.9). In our tests, we obtained similar results using the $\ell_{1}$ norm versus other $\ell_{p}$ norms for $0 \leq p<1$. Since there are many well established algorithms for $\ell_{1}$ minimization, we proceed by selecting $\psi=\|\cdot\|_{1}$ in (4.2.9). For clarity in the presentation, we assume only two point source locations, however, by superposition, we expect similar results for more point sources. In practice, this has been noticed as well. The measurements
are taken at 65 randomly selected observation locations at the final time, $t_{f}=1$. The measurement locations and noisy measurements are depicted in Figures 6.2.3a, 6.2.3b. Two representations of the approximate solution are shown in Figures 6.2.4a, 6.2.4b, where one can see that the two point sources are observed quite well, with only small values away from the actual locations. As in the one-dimensional case, this example really shows the strength of the method when the solution and the data are sparsely distributed.


Figure 6.2.3: (a) Randomly selected observation locations displayed in red; (b) Measurements with $10 \%$ noise.

### 6.3 2-D Convection-Diffusion Equation

In this section, we present severeal numerical results for inverse problems involving the convection-diffusion equation

$$
\begin{align*}
& \frac{\partial v}{\partial t}=c(x) \cdot \nabla v+\nabla \cdot(d(x) \nabla v)+f(x)  \tag{6.3.1}\\
& v(x, 0)=v_{0}(x) . \tag{6.3.2}
\end{align*}
$$



Figure 6.2.4: (a) Reconstruction with $\eta_{1}=1 e-3, \eta_{2}=1 e-14$; (b) Reconstruction depicted with exact solution.

For the results presented here, we assume $c(x) \equiv c, d(x) \equiv d$ are constant (or at least locally constant), and we take $f(x) \equiv 0$. For both simulations, the domain is taken as the unit square $\Omega=[0,1] \times[0,1]$. We consider this problem for several reasons. Determining the initial condition for the convection-diffusion equation is important in its own right, however, one can also consider this as determining the initial condition of the heat equation with moving sensors or measurement locations. Hence, we are able to analyze the performance of reconstructing the initial condition of the heat equation with moving sensors, as well as for the convection-diffusion equation itself.

## Example 1 : Initial condition reconstruction

We first consider the initial condition reconstruction problem with $d=.1, c=\left(\frac{1}{2}, \frac{1}{2}\right)$ known, where we have the observation operator defined by

$$
C v(t)=\frac{1}{\mu\left(\Omega_{s}\right)} \int_{\Omega_{s}} v(s) d \mu
$$

where $\mu\left(\Omega_{s}\right)$ is the volume of the set $\Omega_{s}$. That is, we take average measurements over a sample set $\Omega_{s} \subset \Omega$. For both simulations, we take nine measurement locations equally spaced over the domain, each location of size $\frac{1}{10} \times \frac{1}{10}$. The corresponding contaminated
measurements are depicted in Figure 6.3.1b, which are filtered via (6.2.4) prior to implementing the method. Using the operator splitting technique outline in Section 2.6, the convection-diffusion equation fits into the abstract framework (2.1.1). We take the exact initial condition

$$
v_{0}(x, y)=e^{-100\left((x-.55)^{2}+(y-.5)^{2}\right)}
$$

and we solve the corresponding inverse problem using the $L_{1}-H_{1}$ regularization, with basis functions

$$
\varphi_{k, l}(x, y)=\sin (k \pi x) \sin (l \pi y)
$$

As can be seen by comparing Figure 6.3.2a, 6.3.2b, the method for reconstructing the initial condition performs well with the parameters $\eta_{1}=.03, \eta_{1}=1 \times 10^{-8}$. Depending on the basis choice, small errors are expected due to the truncation of the generalized Fourier series. In this case, we have small oscillations due to the sinusoidal basis. We also consider the convection-diffusion equation to investigate how adding convection affects the estimation results. We compare the reconstruction of the same initial condition from the convection-diffusion equation and the diffusion equation using the same measurement locations. As can be seen in Figure 6.3.3, the addition of convection has the potential to sharpen results depending on measurement location. In this case, there is the appearance of a hump over one measurement location when no convection is present. This is likely due to difficulty in capturing the direction in which information(heat flow) is propagating.


Figure 6.3.1: (a) Nine measurement locations depicted in red; (b) Noisy measurements used for reconstruction compared with exact measurements.


Figure 6.3.2: (a) Exact initial condition; (b) Reconstruction with $10 \%$ noise in measurements.


Figure 6.3.3: (a) Reconstruction with $c_{x}=0=c_{y}$; (b) Reconstruction with $c_{x}=\frac{1}{2}=c_{y}$.

## Example 2: Simultaneous state/parameter estimation

We now provide results for simultaneously reconstructing the initial condition and the unknown constants $c(x) \equiv c, d(x) \equiv d$ for the 2-D convection-diffusion equation. The exact values of the parameters are $d=.01, c_{x}=\frac{1}{2}, c_{y}=\frac{1}{2}$, so that the problem is convection dominant. We assume initial parameters of $d_{0}=.007, c_{x, 0}=.4, c_{y, 0}=.3$ to compute the minimizer of 3.1.4 via the MATLAB routine fminunc, with the approximations (3.1.7). The results are depicted in Figure 6.3.4 where one can see that the method reasonably reconstructs the initial condition and identifies the convection/diffusion coefficients.


Figure 6.3.4: (a) Exact solution; (b) Reconstruction with $d=.0074901, c_{x}=.43712, c_{y}=$ . 48074.

## Chapter 7

## Two real world applications

In this chapter, we present results for the application of the nonsmooth regularization to two real world applications. This section justifies how the methods can be applied to numerous studies.

### 7.1 Application to data classification

In this section, we briefly describe the application of the nonsmooth regularization to a data classification problem. Our general classifier can be written as the optimization problem

$$
\begin{align*}
& \min _{(w, \gamma)} \phi(y)+\beta \psi(w, \gamma),  \tag{7.1.1}\\
& y=e-H u
\end{align*}
$$

where we assume $\phi, \psi$ are lower semi-continuous, as in Chapter 4. An example of such a classifier is the proximal support vector machine (PSVM)[17], where one takes $\phi(y)=$ $\|y\|_{2}^{2}$ and $\psi(w, \gamma)=\frac{1}{2}\left(\|w\|_{2}^{2}+|\gamma|^{2}\right)$, for $X=\mathbb{R}^{n}$, with $\beta=\frac{1}{\nu}$. The choice of the 2-norm is often chosen for ease of computation and to guarantee the closed form solution, however, statistically we should consider other norms. To obtain a sparse classifier, we take

$$
\begin{equation*}
\psi(w, \gamma)=\|w\|_{p}^{p}+\frac{1}{2}|\gamma|^{2} \quad \text { where } \quad\|w\|_{p}^{p}=\sum_{i=1}^{m}\left|w_{i}\right|^{p} \tag{7.1.2}
\end{equation*}
$$

for $0<p \leq 1$. Further details can be found in Appendix C.

We now provide numerical results from the application of our approach to a neural classification problem. We first provide an outline of how the experiment is conducted, as it will be important for deciphering the results provided. The experiment proceeds by asking a person who is incapable of certain movements to think about performing a specific movement such as wrist flexion, elbow extension, or closing the hand. The data obtained is neural firing rate data, meaning that this is a time dynamical data classification problem. It is desired to determine the correlation between imagining a movement and the neural response to imagining the movement. The experiment consists of periods of rest and periods where the person is cued to think about a certain movement. For this particular experiment, there are five wrist movements consisting of wrist extension, wrist flexion, wrist radial deviation, wrist ulnar deviation, and closing hand. The patient is cued to imagine one of the movements consecutively, with periods of rest between each cue. The person is then cued for another of the five movements consecutively, again with periods of rest in between each cue. The period of time after the cues for wrist up consists of both rest and cues for other movements. This is the nature of how the experiment is conducted. Data is collected for a specified interval of time, which consists of periods of rest and periods of cues for each movement. The goal of the data classification is to sharply separate the data for each movement. For example, we must separate the data corresponding to wrist "up" (extension) from the data corresponding to both the rest periods and the periods for other movements. Thus, the results given here are only for one particular movement, however the method produces similar results for the other five movements.

The results provided are for a patient who was left paralyzed after a stroke. The cues for the action of 'wrist up' can be seen in Figure 7.1.1 for this particular patient.

By taking the $\ell_{p}$ minimization with $p=.2$ versus the $\ell_{2}$ minimization we are able to increase the number of coefficients such that $\left|w_{i}\right| \leq 1 e-4$ from 1 to 18 . The increase in sparsity is depicted in Figure 7.1.2 where one can see how taking successively smaller values of $p$ reduces the number of nonzero weights, $w$. We now illustrate how changing $\beta$ can improve the force. A comparison of the nonsmooth formulation can be seen in Figure 7.1.3. It should be noted that the performance is improved, in addition to increased sparsity

As a test of the classifier, we train on the first three cues for a specified movement and test the classifier on the fourth cue. The test results can be seen in Figure 7.1.4 for


Figure 7.1.1: Cues for the wrist up action.


Figure 7.1.2: Comparison of weights for different $\ell_{p}$ norms.
two movements. As one can see, the classification is nearly equal for both the PSVM and our nonsmooth formulation.


Figure 7.1.3: Comparison of PSVM and nonsmooth formulation.


Figure 7.1.4: (a) Test of nonsmooth formulation for wrist down for $\alpha=.1, p=2$; (b) Test of nonsmooth formulation for wrist right for $\alpha=.1, p=2$.

### 7.2 Application of Reconstruction to Synthetic Aperture Sonar Imaging

### 7.2.1 Problem formulation

Synthetic aperture sonar(SAS) operates by sending acoustic signals at predefined stops along a track or path. The returned signals from the objects(scatterers) are then combined to simulate the signal from a real aperture sonar. Due to this process, techniques are required in order to obtain image estimates from the synthesized signal. The setup of the process for obtaining the data can be seen in Figure 7.2.1.


Figure 7.2.1: Synthetic Aperture Sonar setup.

The most widely used technique for obtaining image estimates in SAS is the $\omega$ - $k$
algorithm, which will briefly describe here. The fully details of the method are developed in the theses [6],[18]. The $\omega$ - $k$ algorithm is a fast Fourier domain technique and works well under certain assumptions. The predecessor to the $\omega-k$ algorithm is the Range-Doppler algorithm, which is still commonly used. In their most basic forms, the $\omega$ - $k$ algorithm outperforms the Range-Doppler algorithm. A PDE migration technique is developed in [23], which utilizes a regularization technique for refining the image estimate. This technique works well and produces a more physically meaningful estimate, however, the speed is much slower when compared with frequency domain techniques.

### 7.2.2 Background

Given a signal(source) function $p$ and the raw echo signal $e$, we form the pulse-compressed echo signal by

$$
\begin{equation*}
s(t, u)=\int e(s, u) p^{*}(s-t) d s \tag{7.2.1}
\end{equation*}
$$

where $k_{u}$ is the wavenumber in the along-track direction.

## Range-Doppler algorithm

For the Range-Doppler algorithm, we take the Fourier transform in the range, $u$, that is, the along-track direction(see Figure 7.2.1). Here, we have the 1-D Fourier transform of the pulse-compressed signal

$$
\begin{equation*}
s\left(t, k_{u}\right)=\iint e\left(s, k_{u}\right) p^{*}(s-t) e^{-i k_{u}} d s d k_{u} \tag{7.2.2}
\end{equation*}
$$

The Range-Doppler representation of the pulse-compressed signal is given by

$$
\begin{equation*}
s\left(t, k_{u}\right) \approx \frac{f\left(x_{0}, k_{u}\right)}{\sqrt{4 k_{0}^{2}-k_{u}^{2}}} \delta\left(t-\frac{2}{c} x_{0} C_{s}\left(k_{u}\right)\right) \exp \left(i\left|x_{0}\right| \sqrt{4 k_{0}^{2}-k_{u}^{2}}\right) \tag{7.2.3}
\end{equation*}
$$

where

$$
\begin{equation*}
C_{s}\left(k_{u}\right)=\frac{1}{\sqrt{1-\left(\frac{k_{u}}{2 k_{0}}\right)^{2}}}-1 \tag{7.2.4}
\end{equation*}
$$

for a target located at $x_{0}$, where $k_{0}$ is the carrier wavenumber. Note that we use the exact curvature factor (7.2.4) rather than the approximation

$$
\begin{equation*}
C_{s}\left(k_{u}\right) \approx \frac{1}{2}\left(\frac{k_{u}}{2 k_{0}}\right)^{2} \tag{7.2.5}
\end{equation*}
$$

which is often utilized. Since, we are only working in the along-track direction, we decouple the range and cross-range via the transformation $T$ defined by

$$
\begin{align*}
x & \rightarrow \frac{c}{2}\left(x_{0}-x_{0} C_{s}\left(k_{u}\right)\right)  \tag{7.2.6}\\
k_{y} & \rightarrow k_{u} . \tag{7.2.7}
\end{align*}
$$

Next, we apply a narrowband propagation filter

$$
\begin{equation*}
q\left(x, k_{y}\right)=\sqrt{4 k_{0}^{2}-k_{y}^{2}} \exp \left(i|x| \sqrt{4 k_{0}^{2}-k_{y}^{2}}\right) \tag{7.2.8}
\end{equation*}
$$

after applying the transformation $T$. Thus, we have

$$
\begin{equation*}
\hat{f}\left(x, k_{y}\right)=\tilde{s}\left(x, k_{y}\right) q\left(x, k_{y}\right) \tag{7.2.9}
\end{equation*}
$$

where $\tilde{s}=T s$. The final image estimate is given by the inverse Fourier transform of $\hat{f}\left(x, k_{y}\right)$, in the along-track direction, so that the Range-Doppler algorithm can be summarized by

$$
\begin{equation*}
\hat{f}(x, y)=\mathcal{F}_{k_{y}}^{-1}\left\{q\left(x, k_{y}\right) T\left(s\left(t, k_{u}\right)\right)\right\} . \tag{7.2.10}
\end{equation*}
$$

## Wavenumber algorithm

The $\omega$ - $k$ (wavenumber) algorithm is very similar to the Range-Doppler algorithm, however, in order to reverse the effects of the SAS imaging system, a coordinate transform is utilized in the wavenumber domain. This requires that we take the Fourier transform in the range, as well as the cross-range(along-track). This is the main difference between the two algorithms. The wavenumber algorithm can be summarized by

$$
\begin{equation*}
\hat{f}\left(k_{x}, k_{y}\right)=S^{-1}\left\{\sqrt{4 k^{2}-k_{u}^{2}} \hat{s}\left(\omega, k_{u}\right) \exp \left(i\left|x_{0}\right| \sqrt{4 k^{2}-k_{u}^{2}}\right)\right\} \tag{7.2.11}
\end{equation*}
$$

where the inverse Stolt map, $S^{-1}$, is defined by

$$
\begin{align*}
k_{x} & =\sqrt{4 k^{2}-k_{u}^{2}}  \tag{7.2.12}\\
k_{y} & =k_{u} \tag{7.2.13}
\end{align*}
$$

and $\hat{s}\left(\omega, k_{u}\right)$ is the 2-D Fourier transform of (7.2.1). The 2-D inverse Fourier transform of (7.2.11) gives the image estimate.

For the approximation of the integral

$$
\begin{equation*}
\int \exp \left(-i 2 k \sqrt{x^{2}+(y-u)^{2}}-i k_{u} u\right) d u=\int \exp (i \psi(u)) d u \tag{7.2.14}
\end{equation*}
$$

which arises in taking the Fourier transform of (7.2.1), we utilize the principle of stationary phase or other approximations of highly oscillatory integrals [40].

### 7.2.3 Parameterized Range-Doppler algorithm

We briefly describe the parameterization of the Range-Doppler and/or $\omega$ - $k$ algorithms for intrinsically sharpening the image estimates. We effectively window the cross-range, $u$, by taking a modified wavespeed

$$
\tilde{c}=\alpha c
$$

where $c$ is the original wavespeed. This acts as a window for the computation of the Fourier transform, thus reducing the computational expense. The parameter $\alpha$ must be chosen appropriately so that the sharpness of the image is maintained(or improved), while reducing the computational cost of the algorithm. Given this parameterization of the Range-Doppler algorithm, it turns out that the parameter corresponding to the sharpest image coincides with the minimization of the total variation, and likewise, the maximization of the entropy. Thus, we cast the problem as

$$
\min _{I(\alpha) \in X}\|I\|_{T V}=\min _{I \in X} \int_{X}|\nabla I| d x
$$

subject to

$$
K(\alpha) I=y^{\delta}
$$

where $\alpha$ is the parameter to be determined. Here, $K(\alpha)$ is the migration operator corresponding the Range-Doppler algorithm or the $\omega-k$ algorithm and $I$ is the image estimate.

In Figure 7.2.2, on can see a comparison of images produced using the $\omega$ - $k$ algorithm and the improved Range-Doppler algorithm. Note that the improved Range-Doppler algorithm maintains the sharp results of the $\omega-k$ algorithm while reducing the appearance of striations.

One advantage of the filtered Range-Doppler method is the ability to localize the method. One may window in on a specific region of interest prior to applying the method. Also, if the wave speed is variable in the along-track direction then the parameter $\beta$ may be adjusted, so that the method can be adapted for the case of non-homogeneous materials.

### 7.2.4 PDE migration

We now briefly describe a PDE migration technique for forming image estimates from SAS data via the one-way wave equation. Let $s(x, t)$ be the raw data, then we denote the Fourier transform of $s(x, t)$ by

$$
\begin{equation*}
A(\psi, \omega)=\iint_{\omega \xi} s(x, t) e^{-i(\omega t+\xi x)} d \xi d \omega \tag{7.2.15}
\end{equation*}
$$

We assume the plane wave extrapolation

$$
\begin{equation*}
D(\xi, \eta, \omega)=A(\xi, \omega) e^{i(\omega t+\xi x+\eta y)} \tag{7.2.16}
\end{equation*}
$$

where

$$
\omega^{2}=\frac{c^{2}}{4}\left(k_{x}^{2}+k_{y}^{2}\right) .
$$

The inverse Fourier transform of $D$ is given by

$$
\hat{d}(x, y, t)=\frac{1}{(2 \pi)^{3}} \int D\left(k_{x}, k_{y}, \omega\right) d k_{x} d k_{y} d \omega
$$

and it satisfies

$$
\begin{align*}
\frac{4}{c^{2}} \frac{\partial^{2} \hat{d}}{\partial t^{2}} & =\frac{1}{2} \frac{\partial^{2} \hat{d}}{\partial y^{2}}+\frac{2}{c} \frac{\partial^{2} \hat{d}}{\partial x \partial t}  \tag{7.2.17}\\
\hat{d}(x, 0, t) & =s(x, t) \tag{7.2.18}
\end{align*}
$$

where $s$ is the SAS data. Solving (7.2.18) yields an image estimate, which can be sharpened via regularization techniques such as those developed in Chapter 4. Further details can be found in [23].


Figure 7.2.2: (a) $\omega$ - $k$ image estimate; (b) Improved Range-Doppler image estimate.

### 7.3 Concluding remarks

The abstract Cauchy problem provides a unified framework for the analysis of systems governed by PDE. The methods developed in this thesis allow for the systematic reconstruction of initial conditions of the abstract Cauchy problem. In particular, the dual control method coupled with the multi-parameter regularization yields a method that is very tunable and robust. By an appropriate basis selection for the problem at hand, and by selecting the parameters in the regularization framework based on the balance principle, a reconstruction filter is determined based on the governing PDE. Depending on the problem size, there may be significant overhead in computing the controls (2.3.11). However, once computed, the controls can be banked (or stored) for future use. Thus, if one carefully selects the basis and the parameters are tuned to the noise and a priori information about the solution, the method can potentially be implemented in real time, simply by integrating the controls against the data.

Diffusion processes and parabolic equations fit particularly well into this framework, due to the necessity for stabilizing the dynamics backward in time. The method accurately reconstructs both initial conditions and point sources of diffusion processes, and allows the forecasting of future states. Thus, the tool provided is valuable for problems where numerous calculations are required based on sensor data, and for problems where integrating forward and backward in time is important.

Based on the multi-parameter regularization, the methods developed are particularly suited for problems involving a locally supported source, such as point sources, as well as those with sparsely distributed data. The sparsity optimization works well for both identifying initial conditions/sources that are locally supported, as well as for selecting the necessary control profile.

Certain questions still remain and extensions to more difficult problems can be realized. Specifically, nonlinear problems can be treated in a similar manner, through the development of nonlinear dual control filters. Preliminary results are promising for the one-dimensional viscous Burger's equation

$$
u_{t}+\left(\frac{u^{2}}{2}\right)_{x}=\varepsilon u_{x x}
$$

As mentioned in the introductory comments, the long term goal is to extend the method
to nonlinear problems, including the incompressible Navier-Stokes equations

$$
\begin{aligned}
\frac{\partial}{\partial t} v_{i}+\sum_{j=1}^{n} v_{j} \frac{\partial v_{i}}{\partial x_{j}}= & \nu \Delta v_{i}-\frac{\partial p}{\partial x_{i}}+f_{i}(x, t) \quad\left(x \in \mathbb{R}^{n}, t \geq 0\right) \\
& \operatorname{div} v=\sum_{i=1}^{n} \frac{\partial v_{i}}{\partial x_{i}}=0 \quad\left(x \in \mathbb{R}^{n}, t \geq 0\right)
\end{aligned}
$$

with initial conditions

$$
v(x, 0)=v_{0}(x) \quad\left(x \in \mathbb{R}^{n}\right)
$$

The consideration of the Navier-Stokes equations also involves the necessity for considering higher-dimensional problems. In this case, the solution for the controls must be performed efficiently, though, once computed, this framework may be ideal for large problems since the filter can simply be banked. Thus, future research for this method involves addressing computational efficiency.

## REFERENCES

[1] N. Akhiezer and I. Glazman. Theory of Linear Operators in Hilbert Space. Dover, New York, 1993.
[2] P. Antsaklis and A. Michel. Linear Systems. Birkhäuser, New York, second edition, 2006.
[3] A.V. Balakrishnan. Kalman Filtering Theory. Optimization Software, New York, 1987.
[4] N. Boussetila and F. Rebbani. A modified quasi-reversibility method for a class of ill-posed cauchy problems. Georgian Mathematical Journal, 14(4):627-642, 2007.
[5] K. Brammer and G. Siffling. Kalman-Bucy Filters. Artech House, Norwood, MA, 1989.
[6] Hayden J. Callow. Signal Processing for Synthetic Aperture Sonar Image Enhancement. PhD thesis, University of Canterbury, NZ, 2003.
[7] E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52:489-509, 2006.
[8] F. Catté, P. Lions, J. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. SIAM J. Numer. Anal., 29:182-193, 1992.
[9] D.L. Colton. Surveys on solution methods for inverse problems. Springer mathematics. Springer, 2000.
[10] R.F. Curtain and H.J. Zwart. An introduction to infinite-dimensional linear systems theory. springer-verlag, New York, NY, 1995.
[11] H. Engl, M. Hanke, and A. Neubauer. Regularization of Inverse Problems. Kluwer Academic, Boston, 2000.
[12] H. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularisation of non-linear ill-posed problems. Inverse Problems, 5:23-540, 1989.
[13] H.W. Engl, W. Rundell, Gesellschaft für Angewandte Mathematik und Mechanik, Society for Industrial, and Applied Mathematics. Inverse problems in diffusion processes: proceedings of the GAMM-SIAM Symposium. Proceedings in Applied Mathematics Series. Society for Industrial and Applied Mathematics, 1995.
[14] M. Fisher, J. Nocedal, Y. Trémolet, and S.J. Wright. Data assimilation in weather forecasting: A case study in pde-constrained optimization. Optimization and Engineering, 10:409-426, 2009.
[15] V. Fomin. Optimal Filtering. Kluwer Academic, Netherlands, 1999.
[16] A. Friedman. Foundations of Modern Analysis. Dover, New York, 1982.
[17] G. Fung and O. L. Mangasarian. Proximal support vector machine classifiers. In F. Provost and R. Srikant, editors, Proceedings KDD-2001: Knowledge Discovery and Data Mining, August 26-29, 2001, San Francisco, CA, pages 77-86, New York, 2001. Asscociation for Computing Machinery. ftp://ftp.cs.wisc.edu/pub/dmi/ tech-reports/01-02.ps.
[18] David W. Hawkins. Synthetic aperture imaging algorithms: with application to wide bandwidth sonar. PhD thesis, University of Canterbury, NZ, 1996.
[19] Eugenio Hernández and G.L. Weiss. A first course on wavelets. Studies in advanced mathematics. CRC Press, 1996.
[20] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategry as a semi-smooth newton method. SIAM Journal of Optimization, 13, 2002.
[21] V. Isakov. The inverse problem of option pricing. Inverse Problems, 13, 1997.
[22] V. Isakov. Inverse problems for partial differential equations. Number v. 127, pt. 2 in Applied mathematical sciences. Springer, 2006.
[23] K. Ito and Q. Huynh. Synthetic aperture sonar imaging via one-way wave equations. CRSC Technical Reports, CRSC-TR09-12, 2009.
[24] K. Ito, B. Jin, and T. Takeuchi. Multi-parameter tikhonov regularization. arXiv:1102.1173v1 [math.NA].
[25] K. Ito, B. Jin, and T. Takeuchi. A regularization parameter for nonsmooth tikhonov regularization. submitted.
[26] K. Ito, B. Jin, and J. Zou. A new choice rule for regularization parameters in tikhonov regularization. submitted.
[27] K. Ito and F. Kappel. Evolution Equations and Approximations. World Scientific, London, 2002.
[28] K. Ito and K. Kunisch. Bv-type regularization methods for convoluted objects with edge, flat and grey scales. Inverse Problems, 2000.
[29] K. Ito and K. Kunisch. Lagrange Multiplier Approach to Variational Problems and Applications. SIAM, Philadelphia, 2008.
[30] K. Ito, K. Ramdani, and M. Tucsnak. A time reversal based algorithm for solving initial data inverse problems. submitted.
[31] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. IEEE Trans. on Aut. Control, 2000.
[32] A.H. Jazwinski. Stochastic Processes and Filtering Theory. Academic Press, New York, 1970.
[33] B. Jin and J. Zou. Augmented tikhonov regularization. Inverse Problems, 25, 2009.
[34] F. Lewis and V. Syrmos. Optimal Control. Wiley-IEEE, New York, 1995.
[35] L. Ling, M. Yamamoto, Y.C. Hon, and T. Takeuchi. Identification of source locations in two-dimensional heat equations. Inverse Problems, 22:1289-1305, 2006.
[36] S. Mcallister and F. Neubrander. Stabilized approximations of strongly continuous semigroups. https://www.math.lsu.edu/~neubrand/McAllisterNeubrander_ Jan3_2007.pdf.
[37] R. McOwen. Partial Differential Equations. Prentice-Hall, New York, 2003.
[38] V.A. Morozov. On the solution of functional equations by the method of regularization. Soviet Math. Dokl., 7:414-417, 1966.
[39] B.K. Øksendal. Stochastic differential equations: an introduction with applications. Universitext (1979). Springer, 2003.
[40] Sheehan Olver. Numerical approximation of highly oscillatory integrals. PhD thesis, University of Cambridge, UK, 2008.
[41] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Machine Intell., 1990.
[42] A. Quarteroni, R. Sacco, and F. Saleri. Numerical Mathematics. Springer, New York, 2006.
[43] F. Riesz and B. Sz.-Nagy. Functional Analysis. Dover, New York, 1990.
[44] R.T. Rockafellar. Convex Analysis. Princeton University Press, Princeton, 1996.
[45] H.L. Royden. Real Analysis. Macmillan, New York, 1988.
[46] L. Rudin, S. Osher, and E. Fatemi. Total variation based noise removal algorithms. Physica D, 60:259-69, 1992.
[47] W. Rudin. Real and Complex Analysis. McGraw-Hill, New York, 1987.
[48] T. Takeuchi and M. Yamamoto. Tikhonov regularization by a reproducing kernel hilbert space for the cauchy problem for an elliptic equation. University of Tokyo Graduate School of Mathematical Sciences Preprint series, 2007.
[49] Andrey N. Tikhonov and Vasiliy Y. Arsenin. Solutions of Ill-Posed Problems. John Wiley \& Sons, New York, 1977.
[50] H. Wendland. Scattered data approximation. Cambridge monographs on applied and computational mathematics. Cambridge University Press, 2005.
[51] M. Yamamoto. Stability, reconstruction formula and regularization for an inverse source hyperbolic problem by a control method. Inverse Problems, 11:481-496, 1995.
[52] Y. Yavin. Numerical Studies in Nonlinear Filtering. Springer, New York, 1985.

## APPENDICES

## Appendix A

## Semigroup Theory

In this section, the fundamental results of semigroup theory needed for the analysis of (2.1.1) are described. This very useful theory not only gives necessary and sufficient conditions for the existence of a unique solution to (2.1.1), but also gives practical results for the approximation of the solutions.

Definition A.0.1. Let $\left\{T_{t} ; t \geq 0\right\}$ be a one-parameter family of linear operators on a Banach space $X$ (i.e., $T_{t} \in L(X, X)$ for $t \geq 0$ ). If the operators $T_{t}$ satisfy the conditions

$$
\begin{align*}
& T_{t} T_{s}=T_{t+s} \text { for } t, s \geq 0  \tag{A.0.1}\\
& T_{0}=I  \tag{A.0.2}\\
& \lim _{t \rightarrow t_{0}} T_{t} x=T_{t_{0}} x \text { for each } t_{0} \geq 0 \text { and each } x \in X, \tag{A.0.3}
\end{align*}
$$

then $\left\{T_{t}\right\}$ is called a $C_{0}$-semigroup.
Definition A.0.2. Let $\left\{T_{t}\right\}$ be a $C_{0}$-semigroup on $X$. The linear operator $A$ defined by

$$
A x=\lim _{h \downarrow 0} \frac{1}{h}\left(T_{h} x-x\right), \quad x \in \operatorname{dom}(A)
$$

is called the infinitesimal generator of $\left\{T_{t}\right\}$, where the domain is the set $\operatorname{dom}(A)=\{x \in$ $X: \lim _{h \downarrow 0} \frac{1}{h}\left(T_{h} x-x\right)$ exists in $\left.X\right\}$.

## A. 1 Mild Solutions to the Abstract Cauchy Problem

Definition A.1.1. If $A$ is a closed operator on the space $X$ and $f \in L_{l o c}^{1}([0, \infty) ; X)$, then we define a mild solution of the non-homogeneous Cauchy problem

$$
\begin{equation*}
\frac{d x}{d t}=A x(t)+f(t) \tag{A.1.1}
\end{equation*}
$$

as a function $x \in C([0, \infty) ; X)$ such that

$$
\int_{0}^{t} x(s) d s \in \operatorname{dom} A, t \geq 0
$$

and

$$
x(t)=x_{0}+A \int_{0}^{t} x(s) d s+\int_{0}^{t} f(s) d s, t \geq 0
$$

The proof of the following theorem can be found in [27].
Theorem A.1.1. If $A$ is the infinitesimal generator of a $C_{0}$-semigroup $T_{t}$ and $f \in$ $L_{l o c}^{1}([0, \infty) ; X)$ then

$$
\begin{equation*}
x(t)=T_{t} x_{0}+\int_{0}^{t} T_{t-s} f(s) d s, t \geq 0 \tag{A.1.2}
\end{equation*}
$$

is the unique mild solution of (A.1.1).
Definition A.1.2. An operator $A$ is said to be dissipative if $\langle A x, x\rangle \leq 0$ for all $x \in$ $\operatorname{dom}(A)$.

Example A.1.1 (1-D Heat equation). Consider the one-dimensional heat equation

$$
\frac{\partial u}{\partial t}=\frac{\partial^{2} u}{\partial x^{2}}
$$

so that the operator $A$ is defined by $A u=\frac{d^{2}}{d x^{2}} u$ and we take the domain

$$
\operatorname{dom}(A)=\left\{\frac{d u}{d x}, \frac{d^{2} u}{d x^{2}} \in L^{2}(0,1): u(0)=u(1)=0\right\}
$$

or

$$
\operatorname{dom}(A)=\left\{\frac{d u}{d x}, \frac{d^{2} u}{d x^{2}} \in L^{2}(0,1): u(0)=\frac{d}{d x} u(1)=0\right\} .
$$

For either choice of the domain we have

$$
\langle A u, u\rangle=\int_{0}^{1} \frac{d^{2} u}{d x^{2}} u d x=\left[\frac{d u}{d x} u\right]_{0}^{1}-\int_{0}^{1}\left(\frac{d u}{d x}\right)^{2} d x=-\int_{0}^{1}\left(\frac{d u}{d x}\right)^{2} d x \leq 0
$$

so that $A$ is dissipative. Now, consider the case when the domain of $A$ is

$$
\operatorname{dom}(A)=\left\{\frac{d u}{d x}, \frac{d^{2} u}{d x^{2}} \in L^{2}(0,1): u(0)=0, \frac{d}{d x} u(1)=-c u(1)\right\}
$$

In this case, we have

$$
\langle A u, u\rangle=\int_{0}^{1} \frac{d^{2} u}{d x^{2}} u d x=-c|u(1)|^{2}-\int_{0}^{1}\left|\frac{d u}{d x}\right|^{2} d x
$$

which we want to be less than or equal to zero. Since $u(0)=0$, we have

$$
u(1)=\int_{0}^{1} \frac{d u}{d x} d x
$$

so that

$$
|u(1)|^{2} \leq\left(\int_{0}^{1} 1^{2} d x\right)\left(\int_{0}^{1}\left|\frac{d u}{d x}\right|^{2} d x\right) \leq \int_{0}^{1}\left|\frac{d u}{d x}\right|^{2} d x
$$

Thus, in order for $A$ to be dissipative, we require that $-c-1 \leq 0$ or $c \geq-1$.
Example A.1.2 (Diffusion equation in $\mathbb{R}^{n}$ ). The diffusion equation on $\Omega \subset \mathbb{R}^{n}$ with Dirichlet boundary conditions is given by

$$
\begin{align*}
& \frac{\partial v}{\partial t}(x, t)=\Delta v(x, t)  \tag{A.1.3}\\
& v(x, t)=0 \quad x \in \partial \Omega \tag{A.1.4}
\end{align*}
$$

so that A coincides with the closure of the Laplace operator, defined by

$$
\Delta f(s):=\sum_{i=1}^{n} \frac{\partial^{2}}{\partial s_{i}^{2}} f\left(s_{1}, \ldots, s_{n}\right)
$$

for every $f$ in the Schwartz space

$$
\mathscr{S}\left(\mathbb{R}^{n}\right):=\left\{f \in C^{\infty}\left(\mathbb{R}^{n}\right): \lim _{|x| \rightarrow \infty}|x|^{k} D^{\alpha} f(x)=0 \text { for all } k \in \mathbb{N} \text { and } \alpha \in \mathbb{N}^{n}\right\} .
$$

In this case, the associated semigroup is defined by

$$
S_{t} f(x):=(4 \pi t)^{-n / 2} \int_{\mathbb{R}^{n}} e^{\frac{-|x-\xi|^{2}}{4 t}} f(\xi) d \xi .
$$

with the space $X=L^{p}\left(\mathbb{R}^{n}\right), 1 \leq p<\infty$. The domain of $A$ is taken as

$$
\operatorname{dom}(A)=H^{2}(\Omega) \cap H_{0}^{1}(\Omega)
$$

Example A.1.3 (Laplace operator is dissipative). Consider the Laplace operator on the space $X=H_{0}^{2}(\Omega)$. Then, by use of Green's identity we see that

$$
\langle\Delta u, u\rangle_{X}=\int_{\Omega} u \Delta u=\int_{\partial \Omega} u \frac{\partial u}{\partial n}-\int_{\Omega}|\nabla u|^{2}=-\|\nabla u\| \leq 0
$$

so that the Laplacian is dissipative for all $u \in X$.

## Appendix B

## Definitions and some Background Material

Definition B.0.3 (Minkowski's Inequality). Let $f, g \in L^{p}$ with $1 \leq p \leq \infty$, then

$$
\|f+g\|_{p} \leq\|f\|_{p}+\|g\|_{p}
$$

Definition B.0.4 (Hölder's Inequality). Let $f, g \in L^{p}$ with $1 \leq p \leq \infty$, then

$$
\left\{\int|f g|\right\} \leq\left\{\int|f|^{p}\right\}^{1 / p}\left\{\int|g|^{q}\right\}^{1 / q}
$$

where $\frac{1}{p}+\frac{1}{q}=1$.
Lemma B.0.1. Let $M$ be a closed convex set in a Hilbert space H. For every point $x \in H$ there exists a unique $y^{*} \in M$ such that

$$
\left\|x-y^{*}\right\|=\inf _{y \in M}\|x-y\|
$$

Theorem B.0.2 (Projection Theorem). Let $M$ be a closed linear subspace of a Hilbert space $H$. For any $x \in H$ there exist unique elements $y \in M$ and $z \in M^{\perp}$ such that $x=y+z$.

Theorem B.0.3 (Riesz Representation Theorem). Let $H$ be a Hilbert space and let $x^{*} \in H^{*}$. Then there exists a unique $z \in H$ such that $x^{*}(x)=\langle x, z\rangle_{H}$ for all $x \in H$, and $\left\|x^{*}\right\|=\|z\|$.

Definition B.0.5. A bilinear functional $\sigma$ is called coercive if there exists a constant $M$ such that

$$
\sigma(x, x) \geq M\|x\|^{2} \text { for all } x \in A .
$$

Theorem B.0.4 (Lax-Milgram Theorem). Let $\sigma$ be a bounded, coercive, bilinear functional on a Hilbert space $H$. Then for every $f \in H^{*}$ there exists a unique $z \in H$ such that

$$
f(x)=\sigma(x, z) \text { for all } x \in H .
$$

Further, the following estimate holds:

$$
\|z\|_{H} \leq \frac{1}{M}\|f\|_{H^{*}}
$$

Definition B.0.6. A sequence $\left\{x_{n}\right\} \in X$ is said to be weakly convergent if there exists $x \in X$ such that $x^{*}\left(x_{n}\right) \rightarrow x^{*}(x)$ for any $x^{*} \in X^{*}$. Suppose $X$ is a Hilbert space. By the Riesz Representation Theorem, weak convergence is equivalent to $\left(x_{n}, z\right) \rightarrow(x, z)$ for any $z \in X$. We us the notation $x_{n} \rightharpoonup x$ to denote weak convergence.

Theorem B.0.5 (Bessel's Inequality). Let $\left\{u_{n}\right\}$ be an orthonormal sequence in an inner product space $X$. For every $x \in X$ we have

$$
\left\|x-\sum_{i=1}^{n}\left\langle x, x_{i}\right\rangle x_{i}\right\|^{2}=\|x\|^{2}-\sum_{i=1}^{n}\left|\left\langle x, x_{i}\right\rangle\right|^{2}
$$

and

$$
\sum_{i=1}^{n}\left|\left\langle x, x_{i}\right\rangle\right|^{2} \leq\|x\|^{2}
$$

Definition B.0.7. An orthonormal sequence is said to be complete if for every $x \in X$

$$
x=\sum_{n=1}^{\infty}\left\langle x, x_{n}\right\rangle x_{n} .
$$

Theorem B.0.6 (Parseval's Formula). An orthonormal sequence $\left\{u_{n}\right\}$ in a Hilbert space $H$ is complete if and only if

$$
\|x\|^{2}=\sum_{n=1}^{\infty}\left|\left\langle x, x_{n}\right\rangle\right|^{2} .
$$

Definition B.0.8. The conjugate(dual) $f^{*}: X^{*} \rightarrow[-\infty, \infty]$ of a functional $f$ is defined by

$$
f^{*}\left(x^{*}\right)=\sup _{x \in X}\left\{\left\langle x^{*}, x\right\rangle-f(x)\right\} .
$$

## Appendix C

## Sparsity Regularization for a Neural Classification Problem

## C. 1 The Support Vector Machine

In this section, we give a basic outline of the Support Vector Machine (SVM) algorithms. We are given training data $\mathcal{D}$, a set of $n$ points of the form

$$
\mathcal{D}=\left\{\left(x_{i}, d_{i}\right) \mid x_{i} \in \mathbb{R}^{m}, d_{i} \in\{-1,1\}\right\}_{i=1}^{n}
$$

where the $d_{i}$ is either 1 or -1 . We want to find the maximum-margin hyperplane that divides the points having $d_{i}=1$ from those having $d_{i}=-1$. Any hyperplane can be written as the set of points $x$ satisfying

$$
x_{i} \cdot w-\gamma=0
$$

To this end the linear SVM determines the hyperplane $(w, \gamma)^{t}$ by the constrained minimization;

$$
\begin{equation*}
\min _{(w, \gamma)} \nu \sum_{i=1}^{m} y_{i}+\frac{1}{2}\left(|w|^{2}+\gamma^{2}\right) \tag{C.1.1}
\end{equation*}
$$

$$
\text { subject to } d_{i}\left(x_{i} \cdot w-\gamma\right) \geq 1-y_{i}, \quad y_{i} \geq 0
$$

where $y_{i}$ measures the degree of misclassification and $\nu>0$ is a chosen parameter. That is, the SVM algorithm classifies data into two categories, $\Omega^{-}$and $\Omega^{+}$, geometrically
separated by the plane $\{x: x \cdot w=\gamma\}$, and clustered around the two planes

$$
\begin{align*}
& \Omega^{-}=\left\{x \in \mathbb{R}^{m}: x \cdot w-\gamma \leq-1\right\}  \tag{C.1.2}\\
& \Omega^{+}=\left\{x \in \mathbb{R}^{m}: x \cdot w-\gamma \geq+1\right\} .
\end{align*}
$$

The authors of [17] formulate the inequality $y \geq 0$ in terms of a penalty,

$$
\begin{equation*}
\min _{w, \gamma, y} \frac{\nu}{2}|y|^{2}+\frac{1}{2}\left(|w|^{2}+\gamma^{2}\right) \quad \text { subject to } D(A w-\gamma e)+y \geq e, \tag{C.1.3}
\end{equation*}
$$

where $D=\operatorname{diag}\left(d_{i}\right)$ and $A \in \mathbb{R}^{n \times m}$ with rows $A_{i}=x_{i}, 1 \leq i \leq n$. Furthermore, the PSVM algorithm in [17] replaces the inequality as the equality constraint and formulate the unconstrained minimization

$$
\begin{equation*}
\min \quad \frac{\nu}{2}|y|^{2}+\frac{1}{2}\left(|w|^{2}+\gamma^{2}\right) \quad \text { subject to } D(A w-\gamma e)+y=e . \tag{C.1.4}
\end{equation*}
$$

Our formulation is also an unconstrained minimization of the form

$$
\begin{equation*}
\min \quad \frac{\nu}{2}\left|y^{+}\right|^{2}+\frac{1}{2}\left(|w|^{2}+\gamma^{2}\right) \quad \text { subject to } D(A w-\gamma e)+y=e, \tag{C.1.5}
\end{equation*}
$$

where $y_{i}^{+}=\max \left(0, y_{i}\right)$. Our motivation for choosing $y^{+}$in this manner can be understood by the following simple argument. Note that if $y_{i} \geq 0$ in (C.1.4) then

$$
d_{i}\left(x_{i} \cdot w-\gamma\right)=1-y_{i}
$$

and thus $y_{i}$ is the degree of misclassification. However, if $y_{i} \leq 0$

$$
d_{i}\left(x_{i} \cdot w-\gamma\right)=1-y_{i} \geq 1
$$

and thus the case is allowed. We are motivated by this fact to only penalize $y_{i}^{+}=$ $\max \left(0, y_{i}\right)$ in the formulation (C.1.5). In this sense the formulation (C.1.5) is penalizing the inequality constraint of (C.1.1). It should improve the separability of the classes based on the least squares formulation (C.1.4).

But, the advantage of (C.1.4) is that it has the closed form solution $(w \gamma)^{t}$. That is,
(C.1.4) is equivalent to

$$
\frac{\nu}{2}\left|H(w \gamma)^{t}-e\right|^{2}+\frac{1}{2}\left(|w|^{2}+\gamma^{2}\right)
$$

where $H=D[A-e]$ (i.e. $\left.y=e-H(w \gamma)^{t}\right)$ and thus

$$
u=(w \gamma)^{t}=\left(I+\nu H^{t} H\right)^{-1} H^{t} e
$$

However, it will be shown in Section C. 2 that the formulation (C.1.5) has an efficient implementation as well.

An important consideration for classification problems of this form is the possibility of ill-posedness. If $H$ is very ill-conditioned, i.e. the singular values of $H$ decrease very rapidly to zero, then the solution is very sensitive to the selection of $\nu>0$. The second term $\frac{1}{2}\left(|w|^{2}+\gamma^{2}\right)$ in (C.1.4) represents the 2-norm of $u=(w \gamma)^{t}$. It is more reasonable to use some other norms to obtain a desirable classifier. One of our requirements is that fewer nonzero components of $w$ are in the final solution. To satisfy this requirement, we use the $\ell^{p}$ norm with $0<p \leq 1$ for our formulation in Section C. 2 to obtain the sparse solution. The nonzero components of $w$ represent the essential and critical neurons for classifying the specified movement. In this way we can obtain the neural network information of the Braingate technology.

## C. 2 New Approach

## C.2.1 Improving separability via inequality constraint

We now present an algorithm which indirectly utilizes the inequality $D(A w-e \gamma)+$ $y \geq e$ for the optimization of $(w, \gamma)^{t}$. For this approach, we consider the constrained minimization

$$
\begin{array}{cl}
\min _{(w, \gamma)} & \phi(y)+\beta \psi(w, \gamma) \\
\text { subject to } & D(A w-e \gamma)+y=e \tag{C.2.2}
\end{array}
$$

where we design $\phi$ to incorporate the weighting for regions where $(A w-e \gamma)+y<e$ and $(A w-e \gamma)+y>e$. In this case, the functional $\phi$ is taken to be

$$
\begin{equation*}
\phi(y)=\frac{1}{2} \min (0, H u-e)^{2} . \tag{C.2.3}
\end{equation*}
$$

Here, we are attempting to create more of a division between the two classes of data, so that the data is classified more distinctly. The choice (C.2.3) works well since if $y \geq 0$ we will have

$$
(x \cdot w-\gamma)=1-y_{i}
$$

which means that $y_{i}$ is the degree of misclassification, however, if $y_{i} \leq 0$ we have

$$
\left(x_{i} \cdot w-\gamma\right)=1-y_{i} \geq 1
$$

which is desirable since the data is pushed farther away from 1.

## C.2.2 Weights for reducing bias

Due to the large amount of data for periods of rest, the standard classification algorithm could have a bias towards identifying the "rest" state, hence reducing the identification of an imagined movement. The rest state corresponds to data for which $d_{i}=-1$. Hence, it is reasonable to consider incorporating different weights for the two cases $d_{i}=-1$ and $d_{i}=1$. To reduce the bias towards identifying coefficients for which $d_{i}=-1$, we choose a parameter $\alpha \leq 1$ for the weight corresponding to this data. In that way, for $d_{i}=-1$, we weight the data by selecting a parameter $\alpha$ and we define the norm

$$
\phi(y)=\frac{1}{2}\left\|S^{1 / 2}(H u-e)\right\|^{2}
$$

where the matrix $S$ is defined by

$$
\begin{cases}S_{i i}=1 & \text { if } d_{i}=1 \\ S_{i i}=\alpha & \text { if } d_{i}=-1\end{cases}
$$

in order to account for the bias.

## C.2.3 Algorithm

We now provide the details of the numerical implementation of the methods discussed in Section C. 2 which contains all the desired properties discussed. For all algorithms presented in this section, the Tikhonov regularization parameter is $\beta$ and we define the matrix

$$
\begin{equation*}
H=D[A-e] . \tag{C.2.4}
\end{equation*}
$$

We develop an iterative algorithm which incorporates all aspects given in this section, based on the iterative method (5.0.8). In summary, the iterative method for computing the classifier $(w, \gamma)^{t}$ is given by

$$
\begin{align*}
& \left(H^{t} S \Gamma^{k} H+\beta\left(\begin{array}{cc}
T^{k} & 0 \\
0 & 1
\end{array}\right)\right)\binom{w^{k+1}}{\gamma^{k+1}}=H^{t} S \Gamma^{k} e  \tag{C.2.5}\\
& \Gamma_{i i}^{k}=\max \left(0,1-d_{k}\left(x_{k}-\gamma\right)\right), \quad T_{j j}^{k}=\frac{p}{\max \left(\epsilon^{2-p},\left|w_{j}^{k}\right|^{2-p}\right)}
\end{align*}
$$

for some small $\epsilon>0$. Each step involves forming the diagonal matrices $\Gamma, T$ and solving the linear equation for $\left(w^{k+1}, \gamma^{k+1}\right)^{t}$. For the application considered in this paper, convergence is achieved with a relatively low number of iterations(3 or 4 is reasonable), so that the complexity of the proposed algorithm nearly equals the complexity of the PSVM. Thus, the advantages of this approach are realized with only a small increase in computational cost for our application. Note that (C.2.5) is equivalent to the PSVM formulation if we set $S, T, \Gamma=I \in \mathbb{R}^{n \times m}$.

