



Michigan Technological University
Create the Future Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's
Reports - Open

Dissertations, Master's Theses and Master's
Reports

2012

Numerical solutions of elliptic inverse problems via the equation error method

Mohammad F. Al-Jamal
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Mathematics Commons](#)

Copyright 2012 Mohammad F. Al-Jamal

Recommended Citation

Al-Jamal, Mohammad F., "Numerical solutions of elliptic inverse problems via the equation error method",
Dissertation, Michigan Technological University, 2012.
<https://doi.org/10.37099/mtu.dc.etds/209>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Mathematics Commons](#)

NUMERICAL SOLUTIONS OF ELLIPTIC INVERSE PROBLEMS VIA THE
EQUATION ERROR METHOD

by
Mohammad F. Al-Jamal

A DISSERTATION
Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
(Mathematical Sciences)

MICHIGAN TECHNOLOGICAL UNIVERSITY
2012

©2012 Mohammad F. Al-Jamal

This dissertation, “Numerical Solutions of Elliptic Inverse Problems Via the Equation Error Method,” is hereby approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY IN MATHEMATICAL SCIENCES.

Department of Mathematical Sciences

Signatures:

Dissertation Advisor _____
Mark S. Gockenbach

Committee Member _____
Allan A. Struthers

Committee Member _____
Tamara Olson

Committee Member _____
John A. Jaszczak

Department Chair _____
Mark S. Gockenbach

Date _____

***Dedicated** to my mother, Suha Solayman Al-Jamal, and to the memory of my father, Fawaz Abdullah Al-Jamal*

Table of Contents

List of Figures	vi
List of Tables	viii
Acknowledgements	ix
Abstract	x
1 Introduction	1
2 Background	5
2.1 Ill-posedness of inverse problems	5
2.1.1 Orientation	5
2.1.2 Tikhonov regularization of linear problems	10
2.1.3 Morozov's discrepancy principle	13
2.1.4 Numerical Example	14
2.1.5 Tikhonov regularization of nonlinear problems	17
2.2 Literature review	20
2.2.1 The output least-square method (OLS)	20
2.2.2 A variational method	24
2.2.3 Modified output-least squares (MOLS)	26
3 Preliminaries	29
3.1 Functional analysis	29
3.2 Sobolev spaces	33
3.3 Weak formulations of the forward problems	39
3.3.1 Abstract variational problems	40
3.3.2 The Neumann BVP	42

3.3.3	The Dirichlet BVP	46
3.3.4	The equations of isotropic elasticity	47
4	The equation error method	50
4.1	Motivation	50
4.2	Existence and uniqueness analysis	52
4.3	The discretized problems	56
4.4	Implementation	57
5	Stability and error estimates	60
5.1	The Neumann BVP	60
5.1.1	Numerical examples	69
5.2	Equations of isotropic elasticity	73
5.2.1	Numerical examples	79
5.3	The Dirichlet problem	81
5.3.1	Numerical examples	88
6	Heuristic results	90
6.1	Parameter choice methods	90
6.1.1	Numerical experiments	91
6.2	Data smoothing	97
6.2.1	Numerical experiments	101
7	Conclusion	109
	Bibliography	111
	Appendix A	115
A.1	Spectral theory of compact operators	115

List of Figures

2.1	Hanging Cable.	15
2.2	Exact (solid), computed (dots), $n = 20$, noise level 1%.	16
2.3	Exact (solid), computed (dots), $n = 40$, noise level 1%.	17
2.4	Exact (solid), regularized solution (dots).	17
5.1	Triangulation of Ω	69
5.2	Recovered parameter a_N in Example 5.1.1 with different noise and regularization parameters β	71
5.3	Recovered parameter in Example 5.1.2 with different noise and regularization parameters β	72
5.4	Recovered μ^* in Example 5.2.1 with different noise and regularization parameters α	81
5.5	Recovered λ^* in Example 5.2.1 with different noise and regularization parameters α	82
5.6	Recovered parameter in Example 5.3.1 with different noise and regularization parameters α	89
6.1	Recovered parameter in Example 6.1.1 using optimal β	92
6.2	L-curves for Example 6.1.1 with different noise levels.	94
6.3	L-curves for Example 6.1.2 with different noise levels.	95
6.4	Recovered parameter in Example 6.1.2 using optimal α	96
6.5	Exact data u (solid), noisy data z (dashed).	99
6.6	Exact data u (solid), smoothed data u_s by the SLO (dashed).	100
6.7	u' (solid), u'_s (dashed), and z' (dotted).	100
6.8	Smoothing the data of Example 6.2.2 with noise level $\delta = 0.5$	102
6.9	Recovered parameter in Example 6.2.3 before smoothing the data (a) and (b), and after smoothing the data using cubic smoothing spline (c) and (d).	104

6.10	Smoothing the data of Example 6.2.3 with noise level $\delta = 0.5$ using cubic smoothing spline.	105
6.11	Triangulation of Ω	107
6.12	Recovered coefficient by the OLS, noise level $\delta = 0.001$	108

List of Tables

5.1	Results of Example 5.1.1 with $\delta = 0$ and regularization parameter $\beta = h^2$	70
5.2	Results of Example 5.1.2 with $\delta = 0$ and regularization parameter $\beta = h^2$	71
5.3	Error results in estimating μ^* in Example 5.2.1, the noise level $\delta = 0$ and regularization parameter $\gamma = h^2$	80
5.4	Error results in estimating λ^* in Example 5.2.1, the noise level $\delta = 0$ and regularization parameter $\gamma = h^2$	80
5.5	Results of Example 5.3.1 with $\delta = 0$ and regularization parameter $\alpha = h^2$	88
6.1	Results of Example 6.1.1 for several noise levels using optimal choice of regularization parameter β	93
6.2	Results of Example 6.1.1 for several noise levels, regularization parameter β is computed using the L-curve method.	93
6.3	Results of Example 6.1.2 for several noise levels using optimal choice of regularization parameter α	95
6.4	Results of Example 6.1.2 for several noise levels. The regularization parameter α is chosen using the L-curve method.	96
6.5	Results of Example 6.2.2 using denoised by the SLO method.	101
6.6	Results of Example 6.2.3 using denoised data by cubic smoothing spline.	103
6.7	Comparing the times and number of iterations for the OLS with and without initialization from the equation error method. Asterisked numbers indicate failure to converge to the true parameter.	107

Acknowledgements

My sincere appreciation and gratitude to Dr. Mark S. Gockenbach for his supervision, guidance, and aid in overcoming various obstacles throughout this journey. Having such a mentor was and always will be an honor for me. I am grateful to the dissertation committee for their time, valuable comments, and advice.

Special thanks to my mother Suha, my father Fawaz (may he rest in peace), my brothers and sisters Wafa, Eyad, Ahmed, Neda, and Dema for their endless support and encouragement during my studies. A warm recognition extended to the staff of the Mathematical Sciences department, Jeanne, Margaret, Tori, my friend Jason, Naser, and all my professors and colleagues for their help and backup during hard times.

Abstract

To estimate a parameter in an elliptic boundary value problem, the method of equation error chooses the value that minimizes the error in the PDE and boundary condition (the solution of the BVP having been replaced by a measurement). The estimated parameter converges to the exact value as the measured data converge to the exact value, provided Tikhonov regularization is used to control the instability inherent in the problem. The error in the estimated solution can be bounded in an appropriate quotient norm; estimates can be derived for both the underlying (infinite-dimensional) problem and a finite-element discretization that can be implemented in a practical algorithm. Numerical experiments demonstrate the efficacy and limitations of the method.

Chapter 1

Introduction

Many problems in physics, engineering, and other disciplines can be well modeled by partial differential equations (PDEs). Most of these PDEs contain parameters that are determined by the physical system under study. While the direct (or forward) problem is concerned in solving the PDE for the unknown dependent variable, its related inverse problem is to estimate the parameters from observations of the solution to the direct problem. Such problems are known in the literature as parameter or coefficient identification problem. Due to the increasing number of applications, parameter identification problems now form an integral part of the theory of inverse and ill-posed problems.

In this work, we consider three inverse problems related to the following elliptic boundary value problems (BVPs):

- (i) Poisson's equation with Neumann boundary conditions:

$$\begin{aligned} -\nabla \cdot (a_N \nabla u) &= f_N \text{ in } \Omega, \\ a_N \frac{\partial u}{\partial n} &= g_N \text{ on } \partial\Omega, \end{aligned} \tag{1.1}$$

- (ii) Poisson's equation with Dirichlet boundary conditions:

$$\begin{aligned} -\nabla \cdot (a_D \nabla u) &= f_D \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{1.2}$$

(iii) Equations of isotropic linear elasticity:

$$\begin{aligned} -\nabla \cdot \sigma &= f_E \text{ in } \Omega, \\ \sigma n &= g_E \text{ on } \partial\Omega, \end{aligned} \tag{1.3}$$

where

$$\begin{aligned} \sigma &= \sigma(u) = 2\mu^* \epsilon + \lambda^* \text{tr}(\epsilon)I, \\ \epsilon &= \epsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T). \end{aligned}$$

Equation (1.1) can be viewed as the steady state of the groundwater flow model, where the vertical transmissivity is assumed to be small enough to consider the flow as only two-dimensional. In this case, u represents the piezometric head (groundwater level), a_N is the hydraulic conductivity or the transmissivity for a two-dimensional aquifer (it characterizes the ability of a geologic material, such as soil or rocks, to transmit water), f_N is the recharge (characterizing sources or sinks through Ω), and g_N describes the inflow and outflow through $\partial\Omega$. See, for example, [1–4] for a more detailed discussion.

For a heterogeneous, flat metal plate occupying the domain Ω , (1.2) models, for instance, the steady-state heat flow, where then u represents the temperature distribution, f_D is the heat source and/or sink, and a_D is the thermal conductivity.

In the study of material sciences, especially in the linear elasticity theory, the equations in (1.3) are known as the equations of isotropic elasticity, where u represents the displacement, f_E is a body force, σ is the stress tensor, and ϵ is the linearized strain tensor. The boundary condition in (1.3) indicates that the membrane is stretched by the edge traction g_E . The parameters μ^* and λ^* are called the Lamé moduli, and they describe the elasticity properties of the membrane.

Computing, for instance, the piezometric head u from a_N , f_N , and g_N will be called the *forward problem*; it is a matter of solving the BVP (1.1). Similarly for the other two BVPs. Here we are considering the following *inverse problems*:

- (1) Suppose that a_N and $u = u_N$ are related via the BVP (1.1), given f_N , g_N , and a measurement z_N of u_N , estimate a_N .
- (2) Suppose that a_D and $u = u_D$ are related via the BVP (1.2), given f_D and a measurement z_D of u_D , estimate a_D .

- (3) Suppose that μ^* , λ^* and $u = u_E$ are related via the BVP (1.3), given f_E , g_E , and a measurement z_E of u_E estimate μ^* and λ^* .

The forward problems above have been studied extensively throughout the last decades and their theoretical aspects are fairly complete. For a comprehensive treatment of the Poisson equation, we recommend the books Evans [5] and McOwen [6]. The text by Duvaut and Lions [7] is an excellent reference for the elasticity problem.

Accurate simulation of groundwater movement is a major concern in hydrology. It is vital in such simulations to have reliable values for the aquifer parameters such as the hydrolic conductivity. Besides the high cost of finding (or accurately estimating) these values from a large number of core samples, the values of these parameters can vary dramatically from one sample point to another, and consequently, such readings might give a very poor estimate to the true parameter. Instead, one can estimate the hydrolic conductivity from a measurement to the piezometric head by solving the inverse problems of the Poisson BVP, see [1–4, 8] for this direction. The inverse elasticity problem has been studied from theoretical standpoint, for example in [9–12]. Recently, interesting applications in elasticity imaging have emerged; see [13, 14] and the references therein. More specifically, using successive ultrasonic scans, it is possible to measure interior displacements in human tissue (for example, breast tissue). Since the elastic properties of abnormal tissue are significantly different from those of normal tissue, it might be possible to locate and discover tumors by solving an inverse problem for the Lamé moduli μ^* and λ^* .

A common feature of most inverse problems is the ill-posedness. As a result, finding solutions that are stable, both numerically and analytically, is very challenging and requires a fine blending of various branches of mathematics. This is reflected in the nature of the error bounds presented in Section 2.2; if the data cannot be measured sufficiently accurately, the convergence is not guaranteed. This is also shown by explicit examples by Kohn and Lowe [15] and Alessandrini [16]. Moreover, if ∇u vanishes on some open set, then (1.1) (similarly for the BVPs (1.2) and (1.3)) provides no information about the unknown parameter a_N on that set, and consequently, we have no uniqueness result. Constructing stable algorithms for solving inverse problems is usually done by *regularization*. Roughly speaking, this means replacing the original ill-posed problem by a nearby well-posed problem. We give a brief introduction to the theory of regularization in Section 2.1.

To solve the above inverse problems, we use the *method of equation error*. The method of equation error chooses the value of the parameter that minimizes the error in the PDE and boundary condition. We prove that the method converges provided Tikhonov regularization is used. We also derive error bounds on the computed solution. Numerical experiments demonstrate the efficacy and limitations of the method. The equation error approach applied to the above inverse problems has been used by many researchers; see [8, 12, 17]. However, our results have been obtained in a manner that is both more understandable and more amenable to generalization. The main results reside in Chapter 4 and 5. Further experiments and some heuristic results are given in Chapter 6.

Chapter 2

Background

2.1 Ill-posedness of inverse problems

The purpose of this section is to give a brief introduction to the theory of inverse and ill-posed problems, an active area of study due its numerous applications in science and engineering, including physical, biomedical, and geophysical applications. For in-depth theoretical treatment and a wide list of applications in the subject, we recommend the books [18–21].

2.1.1 Orientation

Probably there is no unified definition of what an inverse problem is. According to many references, inverse problems are concerned with determining causes for observed or desired effects. This definition is somewhat subjective and the precise meaning varies from application to another, so we shall not go over specific examples here.

However, it is almost universal that most of inverse problems encountered in science and engineering are *ill-posed*, and as a consequence, finding stable numerical and analytical methods for solving inverse problems is exceptionally difficult and very challenging. According to Jacques Hadamard [22], a problem is *well-posed* if and only if the following properties hold

- for each admissible data, at least one solution exists (existence);
- for each admissible data, at most one solution exists (uniqueness);

- the solution depends continuously on the data (stability).

If at least one of the above properties is violated, the problem is usually labeled *ill-posed* (in the sense of Hadamard). Hadamard believed that any mathematical model of a physical phenomenon should satisfy these properties.

Abstractly, let X and Y be vector spaces, and assume $T : X \rightarrow Y$ is given. An inverse problem can be posed as:

given $y \in Y$, find $x \in X$ satisfying $Tx = y$.

In other words, an inverse problem asks for a solution of the equation $Tx = y$. However, not every equation $Tx = y$ represents a ‘genuine’ inverse problem, but only those in which the operator T has certain properties. To simplify the exposition below, we will assume that X and Y are Hilbert spaces and $T : X \rightarrow Y$ is a bounded linear operator. Further we pose the above inverse problem more compactly as

$$Tx = y. \tag{2.1}$$

Mathematically, problem (2.1)¹ is said to be well-posed (in the sense of Hadamard) if and only if the following three conditions hold²:

- *Existence*: for each $y \in Y$, there exists at least one $x \in X$ such that $Tx = y$.
- *Uniqueness*: for each $y \in Y$, there exists at most one $x \in X$ such that $Tx = y$.
- *Stability*: the solution x of $Tx = y$ depends continuously on y ; that is, if $y \in Y$, $\{y_n\} \subset Y$, $Tx = y$, $Tx_n = y_n$ for all $n \in \mathbb{Z}^+$, then $\|y_n - y\|_Y \rightarrow 0$ implies $\|x_n - x\|_X \rightarrow 0$ as $n \rightarrow \infty$.

Since the first two conditions imply that T^{-1} exists, the third condition is equivalent to the condition that T^{-1} be continuous.

If T is not bijective, then, in hopes of getting a well-posed problem, one may instead consider the modified problem

$$\tilde{T}x = y, \tag{2.2}$$

¹Or, more precisely, the problem of solving equation (2.1).

²Here it is enough to assume X and Y are normed spaces.

where³ $\tilde{T} : \mathcal{N}(T)^\perp \rightarrow \mathcal{R}(T)$ and is defined by $\tilde{T}x = Tx$. Note that now \tilde{T} is bijective, and so the existence and uniqueness conditions above are satisfied. If further $\mathcal{R}(T)$ is closed (and hence it is a Banach space), then by the open mapping theorem it follows that \tilde{T}^{-1} is continuous, and thus problem (2.2) is well-posed. In this way we succeeded in fulfilling Hadamard's postulates by such reformulation of the original problem. Therefore the critical question is, what if $\mathcal{R}(T)$ is not closed? In this case, we show that \tilde{T}^{-1} cannot be continuous, and hence, neither the modified problem nor the original problem is well-posed.

Theorem 2.1.1. *\tilde{T}^{-1} is continuous if and only if $\mathcal{R}(T)$ is a closed subspace of Y .*

Proof. Assume \tilde{T}^{-1} is continuous, and let $y \in \overline{\mathcal{R}(T)}$ (the closure is taken in Y). Therefore, there exists a sequence $\{x_n\} \subset \mathcal{N}(T)^\perp$ such that $\tilde{T}x_n \rightarrow y$, and thus, $\{\tilde{T}x_n\}$ is Cauchy in Y . But $\{x_n\}$ is also Cauchy in the Banach space $\mathcal{N}(T)^\perp$ since

$$\|x_n - x_m\|_X \leq \|\tilde{T}^{-1}\| \|\tilde{T}x_n - \tilde{T}x_m\|_Y,$$

and so $x_n \rightarrow x \in \mathcal{N}(T)^\perp$. Since T is continuous, it follows that $y = Tx \in \mathcal{R}(T)$, which completes the proof. \square

From the above discussion we see that $Tx = y$ is a true inverse problem if $\mathcal{R}(T)$ fails to be closed. In this case, and the ill-posedness cannot be treated by any simple mathematical trickery as we tried to accomplish above (to obtain existence and uniqueness). Therefore, it might be wiser to leave our earlier approach and try to attack the problem from different perspectives.

In practice it is often the case that we don't know the exact y , but all what we know is a measurement $y^\delta \in Y$ of y , with an error, say, $\|y - y^\delta\|_Y \leq \delta$. But we might be still interested in obtaining an approximation to a solution of $Tx = y$. Since y^δ (or even y itself) might not be in $\mathcal{R}(T)$, we encounter another dilemma, not to mention the lack of uniqueness if T is not injective! All the above reasons motivates us to leave the search for only 'classical' solutions and attempt to generalize our concept of a solution to include 'generalized' solutions in a meaningful sense which we will describe below.

³Here, $\mathcal{N}(T)$ denotes the kernel (or the null space) of T , $\mathcal{N}(T)^\perp$ is the orthogonal complement of $\mathcal{N}(T)$, and $\mathcal{R}(T)$ denotes the range of T .

If the existence property fails for $Tx = y$, i.e. $\mathcal{R}(T)$ is not all of Y , then we may look for an element $x_l \in X$ such that Tx_l is ‘closest to’ y . Thus for $y \notin \mathcal{R}(T)$, we reformulate $Tx = y$ as: find $x_l \in X$ such that

$$\|Tx_l - y\|_Y = \min_{x \in X} \|Tx - y\|_Y.$$

This is called *the method of least-squares*, and such an x_l , if it exists, is called a *least-squares solution* of (2.1). Clearly for $y \in \mathcal{R}(T)$, equation (2.1) has a (classical) solution which is also a least-squares solution. However, for $y \notin \mathcal{R}(T)$, a least-squares solution is not guaranteed to exist. Actually, we have the following existence theorem, see [21] for proof.

Theorem 2.1.2. *Equation (2.1) has a least-squares solution if and only if $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$. Further, if x_l is a least-squares solution, then $Tx_l = Py$ where Py denotes the orthogonal projection of y onto $\overline{\mathcal{R}(T)}$.*

Thus if $\mathcal{R}(T)$ is closed, then $Y = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$, and so (2.1) has a least-squares solution for every $y \in Y$. Even with above pessimistic result, we have succeeded at least in extending space of admissible data, $\mathcal{R}(T)$, to the larger space, $\mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$, which is a dense subspace of Y .

Now if T is not injective, or equivalently the null space of T , $\mathcal{N}(T)$, is nontrivial, then a least-squares solution, if it exists, cannot be unique. For if x_l is a least-squares solution of (2.1) for a $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$, then

$$\|T(x_l + n) - y\|_Y = \|Tx_l - y\|_Y = \min_{x \in X} \|Tx - y\|_Y \quad \forall n \in \mathcal{N}(T),$$

and so a least-squares solution cannot be unique whenever T is not injective. On the other hand, if T is injective and x_1 and x_2 are any two least-squares solutions of (2.1), then by Theorem 2.1.2 we have the characterization

$$Tx_1 = Py = Tx_2,$$

and so $x_1 = x_2$ since T is injective, and thus, a least-squares solution is unique whenever T is injective.

When T is not injective, the usual remedy of the lack of uniqueness is to seek

the least-squares solution that is of smallest norm⁴. That is, we find an $\hat{x} \in X$ such that \hat{x} is a least-squares solution of (2.1) and

$$\|\hat{x}\|_X = \inf_{x \in X} \{\|x\|_X : x \text{ is a least-squares solution of (2.1)}\}.$$

In the literature, \hat{x} is called a *minimum-norm least-squares solution* of (2.1). It can be shown [18,19,21] that the minimum-norm least-squares solution of (2.1) is unique, and further it is characterized as the unique element in $\mathcal{N}(T)^\perp$ satisfying

$$T\hat{x} = Py,$$

provided, of course, $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$.

Thus based on the above discussion, the general interpretation of $Tx = y$ is:

Find the minimum-norm least-squares solution of $Tx = y$.

This takes care of any lack of uniqueness and also a simple lack of existence, that is, $y \notin \mathcal{R}(T)$ but $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$.

Now let $T^\dagger : \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp \rightarrow X$ denote the mapping that assigns to each $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ the (unique) minimum-norm least-squares solution of $Tx = y$. Thus the equation $\hat{x} = T^\dagger y$ means \hat{x} is the minimum-norm least-squares solution of $Tx = y$. The operator T^\dagger is called the **(Moore-Penrose) generalized inverse** of T . For brevity, we will write $D(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ for the domain of T^\dagger . We summarize some of the important results regarding the operator T^\dagger ; see [18,19,21].

Theorem 2.1.3. *Let X and Y be Hilbert spaces and $T : X \rightarrow Y$ be a bounded linear operator. Then*

- (a) T^\dagger is linear,
- (b) $D(T^\dagger)$ is dense in Y ,
- (c) $\mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$,
- (d) $\mathcal{N}(T^\dagger) = \mathcal{R}(T)^\perp$,

⁴Another strategy, which is often suggested by the applications of (2.1), is to minimize $\|Lx\|_Z$, where $L : X \rightarrow Z$ is another operator, in place of $\|x\|_X$. For example, L might be a differentiation operator.

(e) T^\dagger is continuous if and only if $\mathcal{R}(T)$ is a closed subspace of Y .

Thus, from part (e) of the theorem above, we see again that (2.1) is a ‘true’ inverse problem when $\mathcal{R}(T)$ is not closed (even with the general interpretation of $Tx = y$). We will describe one way of dealing with the stability issue in the next section.

2.1.2 Tikhonov regularization of linear problems

In the last section, we have investigated the abstract inverse problem

$$Tx = y, \tag{2.3}$$

where X and Y are Hilbert spaces and $T : X \rightarrow Y$ is a bounded linear operator. We interpret the problem $Tx = y$ as asking for $x = T^\dagger y$, the minimum-norm least-squares solution of $Tx = y$. This takes care of any lack of uniqueness and also a simple lack of existence ($y \notin \mathcal{R}(T)$ but $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$). However, if $\mathcal{R}(T)$ is not closed, then T^\dagger is unbounded (discontinuous) and there exists a sequence $\{y_n\} \in Y$ with

$$y_n \rightarrow y \text{ and } T^\dagger y_n \not\rightarrow T^\dagger y = x$$

(or even worse, $T^\dagger y_n$ is undefined). In such situation, we would like to obtain stable approximations to $x = T^\dagger y$. Procedures that lead to stable approximations to an ill-posed problems are called *regularization methods*.

The idea of regularization is to approximate the operator T^\dagger by a family of operators $\{R_\alpha : Y \rightarrow X \mid \alpha > 0\}$ such that these approximations get better as $\alpha \rightarrow 0$. To put it in other words, the ill-posed problem $x = T^\dagger y$ is approximated by a family of ‘nearby’ well-posed problems $\{x_\alpha = R_\alpha y \mid \alpha > 0\}$ such that $x_\alpha \rightarrow x$ as $\alpha \rightarrow 0^+$. We will describe one of the most popular regularization methods, **Tikhonov regularization**.

In Tikhonov regularization, the regularization operator R_α is given by

$$R_\alpha = (T^*T + \alpha I)^{-1} T^*.$$

It can be shown [18, 19, 21] that $T^*T + \alpha I$ is invertible, and so R_α is well-defined. For $y \in Y$, $\alpha > 0$, let $x_{\alpha,y} = R_\alpha y$. Then $x_{\alpha,y}$ is characterized as the unique solution

of the optimization problem

$$\min_{x \in X} \|Tx - y\|_Y^2 + \alpha \|x\|_X^2.$$

The number $\alpha > 0$ is called a *regularization parameter*.

For $y \in D(T^\dagger)$, we write $x_{0,y} = T^\dagger y$. We have the following convergence result; see, for example, [18, 19, 21].

Theorem 2.1.4. *Let X and Y be Hilbert spaces and $T : X \rightarrow Y$ be a bounded linear operator. For all $y \in D(T^\dagger)$ we have*

$$x_{\alpha,y} \rightarrow x_{0,y} \text{ as } \alpha \rightarrow 0^+.$$

Although this result is important, it is not directly applicable to a practical problem since we don't know y , but only an estimate y^δ . We have the following result concerning this issue, see [18, 19, 21].

Theorem 2.1.5. *Let X and Y be Hilbert spaces and $T : X \rightarrow Y$ be a bounded linear operator. If $y, y^\delta \in Y$, $\|y^\delta - y\|_Y \leq \delta$, and $\alpha > 0$, then*

$$\|x_{\alpha,y^\delta} - x_{\alpha,y}\|_X \leq \frac{\delta}{2\sqrt{\alpha}}.$$

Now using the triangle inequality, we can decompose the total error as

$$\begin{aligned} \overbrace{\|x_{\alpha,y^\delta} - x_{0,y}\|_X}^{\text{total error}} &\leq \overbrace{\|x_{\alpha,y} - x_{0,y}\|_X}^{\text{regularization error}} + \overbrace{\|x_{\alpha,y^\delta} - x_{\alpha,y}\|_X}^{\text{perturbation error}} \\ &\leq \|x_{\alpha,y} - x_{0,y}\|_X + \frac{\delta}{2\sqrt{\alpha}}. \end{aligned}$$

Notice now, at one hand we would like to let $\alpha \rightarrow 0$ to force the regularization error to converge to zero, on the other hand we lose the grip over the perturbation error if α converges to zero at a faster rate than δ^2 does. This suggests that α cannot be chosen independent of δ . In particular, if we choose α as a function of δ ($\alpha = \alpha(\delta)$) such that

$$\alpha(\delta) \rightarrow 0 \text{ and } \frac{\delta}{\sqrt{\alpha(\delta)}} \rightarrow 0 \text{ as } \delta \rightarrow 0,$$

then it follows that

$$\|x_{0,y} - x_{\alpha,y^\delta}\|_X \rightarrow 0 \text{ as } \delta \rightarrow 0.$$

In the general regularization theory, this is to say that Tikhonov's method with the above choice of $\alpha = \alpha(\delta)$, leads to a *regular or convergent algorithm* for the solving problem (2.3).

Notice that this result does not give any convergence rates. However, under certain 'smoothness' assumptions⁵ on y , it can be shown [18, 19, 21] that:

Theorem 2.1.6. *Let X and Y be Hilbert spaces and $T : X \rightarrow Y$ be a bounded linear operator, and suppose that $y \in D(T^\dagger)$. If $x_{0,y} \in \mathcal{R}((T^*T)^\mu)$ for some $\mu \in (0, 1]$, then*

$$\|x_{\alpha,y} - x_{0,y}\|_X \leq C\alpha^\mu,$$

where $C > 0$ is independent of α .

Thus, provided $x_{0,y} \in \mathcal{R}((T^*T)^\mu)$ for some $\mu \in (0, 1]$, we see that

$$\|x_{\alpha,y^\delta} - x_{0,y}\|_X \leq C \max \left\{ \alpha^\mu, \frac{\delta}{\sqrt{\alpha}} \right\}$$

for some $C > 0$ independent of α and δ . In particular, if $\alpha = \alpha(\delta) = c_0\delta^{2/(2\mu+1)}$ for some $c_0 > 0$, then

$$\|x_{\alpha,y^\delta} - x_{0,y}\|_X = O\left(\delta^{2\mu/(2\mu+1)}\right) \text{ as } \delta \rightarrow 0.$$

Note that the best rate of convergence for $\|x_{\alpha,y^\delta} - x_{0,y}\|_X$ that one can obtain is $O(\delta^{2/3})$, which happens when $\mu = 1$.

One may ask if it is possible to improve the above order $O(\delta^{2/3})$ to say $o(\delta^{2/3})$, possibly by a stronger smoothness assumption on $x_{0,y}$, say if $\mu > 1$? The answer is negative and it can be shown this is the best (optimal) rate one can get from Tikhonov's method, regardless of how smooth $x_{0,y}$ is. Actually, it can be shown that if $x_{0,\alpha} = T^\dagger y \in \mathcal{R}(T^*T)$, and $\|x_{\alpha,y^\delta} - x_{0,y}\|_X = o(\delta^{2/3})$, and $\alpha = \alpha(\delta)$ is chosen such that $\alpha(\delta) \rightarrow 0$ as $\delta \rightarrow 0$, then $x_{0,\alpha}$ is trivial, that is, $x_{0,\alpha} = 0$.

⁵Some authors refer to these assumptions as 'source conditions' or 'abstract smoothness conditions'. The precise definition of the operator $(T^*T)^\mu$ (for general T) requires tools from the spectral theory which is beyond the scope of this presentation. In Appendix A, we outline the theory in the special case that T is compact.

2.1.3 Morozov's discrepancy principle

As we have seen in last section, any *a priori* choice $\alpha = \alpha(\delta)$ satisfying

$$\frac{\delta}{\sqrt{\alpha(\delta)}} \rightarrow 0 \text{ as } \alpha \rightarrow 0,$$

leads to a convergent algorithm for the solution of $Tx = y$. Further, with the particular choice of $\alpha(\delta) = c_0 \delta^{2/(2\mu+1)}$, one can obtain the optimal rate of convergence $O(\delta^{2\mu/(2\mu+1)})$, provided we know that $T^\dagger y \in \mathcal{R}((T^*T)^\mu)$. In practice, we either don't know μ , or even if we know μ , any positive c_0 gives an optimal asymptotic rate of convergence, but the choice of c_0 obviously has a huge impact for a given value of $\delta > 0$. For these reasons, it might be reasonable (and necessary) to take the actual data vector y^δ into account when choosing the regularization parameter α . A parameter choice method that incorporates both δ and y^δ is called an *a posteriori* parameter choice rule. We will describe one such rule, namely, the **Morozov's discrepancy principle** (MDP).

In the MDP, the parameter $\alpha = \alpha(\delta, y^\delta)$ is chosen such that

$$\|Tx_{\alpha, y^\delta} - y^\delta\|_X = \delta.$$

The following theorem shows that the MDP rule induces a convergent regularization method provided $y \in \mathcal{R}(T)$, see [18, 19, 21].

Theorem 2.1.7. *If $y \in \mathcal{R}(T)$, then if $\alpha(\delta, y^\delta)$ is chosen using the MDP, then*

$$x_{\alpha(\delta, y^\delta), y^\delta} \rightarrow T^\dagger y \text{ as } \delta \rightarrow 0.$$

For the rate of convergence, we have [18, 19, 21]:

Theorem 2.1.8. *Suppose $y \in \mathcal{R}(T)$ and $T^\dagger y \in \mathcal{R}(T^*)$. If $\alpha = \alpha(\delta, y^\delta)$ is chosen by the MDP, then*

$$\|x_{\alpha(\delta, y^\delta), y^\delta} - T^\dagger y\|_X = O(\delta^{1/2}).$$

Further, this rate is optimal no matter how smooth $T^\dagger y$ is, except in the case that T is finite-rank (i.e., $\mathcal{R}(T)$ is finite-dimensional).

In the next section, we present a numerical example demonstrating the efficiency of the MDP.

2.1.4 Numerical Example

Now we present an example of an inverse problem arising in statics. See [19] for the derivation of the model.

The Hanging Cable. Imagine a cable of variable density hanging between two horizontal supports, as in Figure 2.1. Assume that the tension T in the cable is constant and that the vertical deflection y of the cable at any point is small relative to the length of the cable. A mathematical model is then

$$y(s) = \int_0^1 k(s, t)x(t)dt, \quad 0 < s < 1,$$

where $x(s)$ is the weight density of the cable, and the kernel $k(s, t)$ is given by

$$k(s, t) = \begin{cases} t(1-s)/T, & 0 \leq t \leq s, \\ s(1-t)/T, & s \leq t \leq 1. \end{cases}$$

The **inverse problem** we wish to pose is: what distribution of the variable mass of the cable *causes* the *observed* deflection mode y ?

Mathematically, this inverse problem can be posed as:

$$Kx = y \tag{2.4}$$

where $K : L^2(0, 1) \rightarrow L^2(0, 1)$ is defined by

$$(Kx)(s) = \int_0^1 k(s, t)x(t)dt, \quad 0 < s < 1.$$

We point out [19, 21] that $K : L^2(0, 1) \rightarrow L^2(0, 1)$ is in fact continuous for any kernel $k \in L^2(0, 1) \times L^2(0, 1)$. First, we show that this inverse problem is ill-posed; stability will be our main concern. Using the Riemann-Lebesgue lemma [23], for any $f \in L^2(0, 1)$ we have

$$\lim_{n \rightarrow \infty} \int_0^1 f(t) \sin(n\pi t)dt = 0,$$

Consider any $x \in L^2(0, 1)$. For each $n \in \mathbb{Z}^+$, define $x_n(t) = x(t) + \sin(n\pi t)$. Then

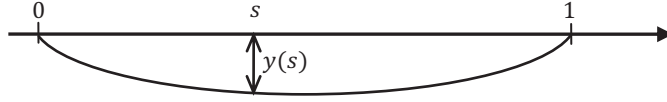


Figure 2.1: Hanging Cable.

by Lebesgue's dominated convergence theorem, we have

$$\lim_{n \rightarrow \infty} \|Kx_n - Kx\|_{L^2(0,1)}^2 = \int_0^1 \lim_{n \rightarrow \infty} \left(\int_0^1 k(s,t) \sin(n\pi t) dt \right)^2 ds = 0$$

for any $k \in L^2(0,1)^2$, while

$$\|x_n - x\|_{L^2(\Omega)}^2 = \int_0^1 \sin^2(n\pi t) dt = \frac{1}{2},$$

showing the instability of the problem. Since in practice the sag in the cable is measured at only finitely many points along the s -axis, the inverse problem must be discretized. To do so, we may choose to discretize the integral using the trapezoid rule on a uniform grid of size $h = 1/n$, and we sample y on the same grid, resulting in a system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{y},$$

where

$$A_{ij} = hk(s_i, t_j), \quad y_i = y(s_i), \quad i, j = 1, 2, \dots, n-1,$$

with

$$s_i = ih, \quad t_j = jh.$$

As a particular example⁶, we will take $y(s) = (s^3(s-2) + s)/12$ and $T = 1$, so that the exact solution is $x(t) = t(1-t)$. Since usually we have inexact data, we simulate the situation by adding $d\%$ (measured in the Euclidean norm) uniformly distributed random noise to the true \mathbf{y} ; thus the noisy data vector \mathbf{y}^δ satisfies

$$\|\mathbf{y}^\delta - \mathbf{y}\|/\|\mathbf{y}\| = d\%,$$

⁶Regardless of the physical plausibility of such choice.

where $\|\cdot\|$ is the Euclidean norm. Thus we are trying to solve the system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{y}^\delta \quad (2.5)$$

for \mathbf{x} . For the first test, we take $n = 20$ and $d = 1$. Figure 2.2 shows the computed solution (i.e. by solving (2.5)) along the exact solution. Obviously, the plot shows an unpleasant result; the computed solution is a poor approximation to the exact solution. Motivated by our numerical experience, we would expect to get better approximations as n increases. Thus, we repeated the experiment, with $n = 40$ at the same noise level, the result is shown in Figure 2.3. Now the computed solution has nothing to do with the exact solution! Such numerical instability is inherited from the original problem, which we know it is ill-posed. Since the finer the discretization, the closer the matrix \mathbf{A} approximates the operator K , the instability in the original problem (2.4) is inherited to the discretized problem (2.5). This explains why the unregularized solutions are getting worse as n increases.

Now we use Tikhonov regularization applied to the discrete problem (2.5) in hopes of getting better results. Thus, the solution is now computed using the formula

$$\mathbf{x} = (\mathbf{A}^*\mathbf{A} + \alpha\mathbf{I})^{-1} \mathbf{A}^*\mathbf{y}^\delta.$$

The regularization parameter α is chosen by the discrepancy principle. Plots for the exact solution versus the computed solution for various values of n and noise level $d\%$ are shown in Figure 2.4.

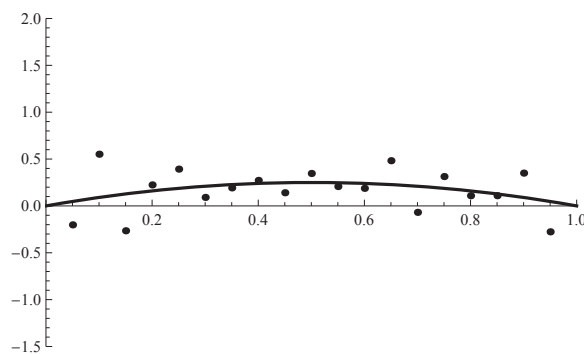


Figure 2.2: Exact (solid), computed (dots), $n = 20$, noise level 1%.

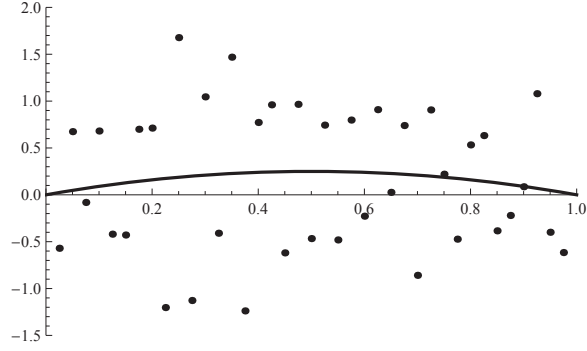


Figure 2.3: Exact (solid), computed (dots), $n = 40$, noise level 1%.

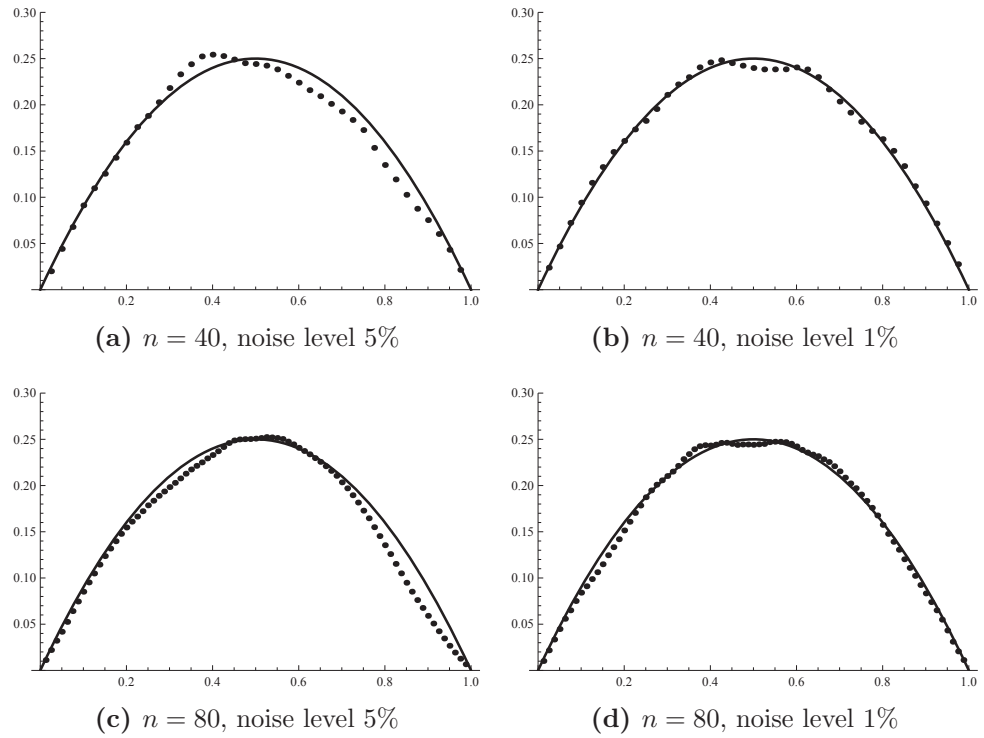


Figure 2.4: Exact (solid), regularized solution (dots).

2.1.5 Tikhonov regularization of nonlinear problems

In this section we will briefly go over the main results developed so far in theory of nonlinear ill-posed problems. All the results in this section can be found in [18].

We want to solve

$$F(x) = y \tag{2.6}$$

where $F : D(F) \subset X \rightarrow Y$ is a nonlinear operator between Hilbert spaces X and Y , and $D(F)$ denotes the domain of F . Throughout this section we assume that:

- (i) Equation (2.6) has an exact solution (but not necessarily unique), and the term *ill-posed nonlinear problem* will mean that the solutions do not depend continuously on the data;
- (ii) F is continuous;
- (iii) F is weakly (sequentially) closed: this means for any sequence $\{x_n\} \in D(F)$, weak convergence of x_n to x in X and weak convergence of $F(x_n)$ to y in Y imply that $x \in D(F)$ and $F(x) = y$.

For $x^* \in X$, we say x^\dagger is an x^* -minimum-norm solution of $F(x) = y$ if

$$F(x^\dagger) = y$$

and

$$\|x^\dagger - x^*\| = \min_{x \in D(F)} \{\|x - x^*\| \mid F(x) = y\}.$$

Usually x^* is an *a priori* guess of the exact solution, and allows one to select a particular solution in the case of multiple solutions. In general, an x^* -minimum-norm solution need not exist nor be unique.

Theorem 2.1.9. *Under the assumptions of this section, x^\dagger exists.*

If problem (2.6) is ill-posed, then it must be regularized. As for the case of linear problems, a widely used method is Tikhonov regularization. Thus one seeks a solution of the optimization problem

$$\min_{x \in D(F)} \|F(x) - y^\delta\|_Y^2 + \alpha \|x - x^*\|_X^2. \quad (2.7)$$

where $\alpha > 0$ is a regularization parameter, $y^\delta \in Y$ is an approximation of the exact right-hand side y of problem (2.6), and $x^* \in X$.

Theorem 2.1.10. *Under the assumptions of this section, problem (2.7) admits a solution. Since F is nonlinear, the solution will not be unique, in general.*

Any solution to (2.7) will be denoted by x_α^δ . The following theorem shows that the problem of solving (2.6) is stable in the sense of continuous dependence of the solutions on the data y^δ .

Theorem 2.1.11. *Let $\alpha > 0$ and let $\{y_k\}$ and $\{x_k\}$ be sequences where $y_k \rightarrow y^\delta$ and x_k is a minimizer of (2.7) with y^δ replaced by y_k . Then there exists a convergent subsequence of x_k and the limit of every convergent subsequence is a minimizer of (2.7).*

We have the following convergence result. Note the assumption on the choice of $\alpha = \alpha(\delta)$, which is the same as for the linear case.

Theorem 2.1.12. *Let $y^\delta \in Y$ with $\|y - y^\delta\| \leq \delta$ and let $\alpha(\delta)$ be such that $\alpha(\delta) \rightarrow 0$ and $\delta^2/\alpha(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Then every sequence $\{x_{\alpha_k}^{\delta_k}\}$ where $\delta_k \rightarrow 0$, $\alpha_k = \alpha(\delta_k)$, and $x_{\alpha_k}^{\delta_k}$ is a solution of (2.7), has a convergent subsequence. The limit of every convergent subsequence is an x^* -minimum-norm solution. If in addition, the x^* -minimum-norm solution x^\dagger is unique, then*

$$\lim_{\delta \rightarrow 0} x_{\alpha(\delta)}^\delta = x^\dagger.$$

We conclude by the following result which gives the convergence rate of the Tikhonov method for nonlinear problems.

Theorem 2.1.13. *Let $D(F)$ be convex, let $y^\delta \in Y$ with $\|y - y^\delta\| \leq \delta$ and let x^\dagger be an x^* -minimum-norm solution. Moreover, let the following conditions hold:*

- (i) F is (Fréchet) differentiable,
- (ii) there exists $\gamma > 0$ such that $\|F'(x^\dagger) - F'(x)\| \leq \gamma\|x^\dagger - x\|$ for all $x \in D(F)$ in a sufficiently large ball around x^\dagger ,
- (iii) there exists $w \in Y$ satisfying $x^\dagger - x^* = F'(x^\dagger)^*w$ and
- (iv) $\gamma\|w\| < 1$.

Then for the choice $\alpha \sim \delta$, we obtain

$$\|x_\alpha^\delta - x^\dagger\| = O(\sqrt{\delta}) \text{ and } \|F(x_\alpha^\delta) - y^\delta\| = O(\delta).$$

2.2 Literature review

Since most inverse problems are ill-posed, many popular techniques for solving inverse problems use some sort of numerical optimization. The unknown parameters are chosen to be those best agreeing with observed data according to some criterion. In this section, we will describe several optimization-based algorithms for solving the inverse problems considered in this work.

2.2.1 The output least-square method (OLS)

Assume that the observable data and the desired parameters are related by a mathematical model, such as a differential equation, and that the data can be simulated for any appropriate estimate of the parameters. The OLS is very natural: choose values for the parameters, simulate the data and compare it with the observed data, then measure the misfit (usually in some norm).

For example, in the context of BVP (1.1), this amounts to solving the optimization problem

$$\min_a \|u(a) - z\|,$$

where z is the observed data, and $u(a)$ represents the simulated data obtained by solving (1.1) with the exact parameter a_N replaced by a .

In practice, the solution of the BVP (1.1) is simulated by the finite element method. If we write $u_h(a_h)$ for the finite element solution of (1.1), then the unknown parameter a_N is estimated by a solution of the optimization problem

$$\min_{a_h \in K_h^{(r)}} \frac{1}{2} \|u_h(a_h) - z_N\|^2, \quad (2.8)$$

where h is the mesh size in the finite element discretization. To be precise, let $\{\mathcal{T}^h\}$ be a family⁷ of triangulations of the domain Ω , where h denotes the maximum

⁷For most of the results belows, it is also assumed to be regular and quasi-uniform, that is, there exists $\nu > 0$ such that

$$\nu h \leq \rho_T \leq h_T \leq h \text{ for all } T \in \mathcal{T}^h, h > 0,$$

where h_T is the diameter of T and ρ_T is the diameter of the largest ball contained in T .

diameter of any triangle in \mathcal{T}^h . Define

$$\begin{aligned} L_h^{(r)} &= \{w \in C(\bar{\Omega}) \mid w|_T \in P_r \text{ for all } T \in \mathcal{T}^h\}, \\ K_h^{(r)} &= \left\{a \in L_h^{(r)} \mid c_0 \leq a \leq c_1\right\}, \\ U_h^{(r)} &= L_h^{(r+1)} \times L_h^{(r+1)}, \end{aligned}$$

where P_r is the space of polynomials in two variables of degree at most r . The numbers c_0 and c_1 are given *a priori* bounds on the true parameter(s). Notice that c_0 must be positive in order for the simulation to be well-defined⁸.

In the case the norm in (2.8) is the L^2 norm, Falk [24] proved the following result.

Theorem 2.2.1. *Suppose that a_h is any solution of (2.8), there exists a constant $C > 0$ independent of h and $\|u_N - z_N\|_{L^2(\Omega)}$ such that*

$$\|a_h - a_N\|_{L^2(\Omega)} \leq C \left(h^r + \frac{\|u_N - z_N\|_{L^2(\Omega)}}{h^2} \right), \quad (2.9)$$

for all h sufficiently small.

In this result, it is assumed that $u_h \in L_h^{(r+1)}$, and that the true parameter a_N and the solution u_N are both smooth, namely, $a_N \in H^{r+1}(\Omega)$ and $u_N \in H^{r+2}(\Omega)$. Also, it is assumed that

$$\nabla u_N \cdot \vec{d} > 0 \text{ on } \Omega \quad (2.10)$$

for some constant unit vector $\vec{d} \in \mathbb{R}^2$. Assumption (2.10) can be perceived as a nondegeneracy condition on the experiment that resulted in the given data. In context of the groundwater flow model, this assumption means that there is always some flow in the direction \vec{d} . This allows PDE (1.1) to be considered as a first order hyperbolic PDE for a_N . We point out that under the nondegeneracy condition (2.10), Falk showed that there is at most one coefficient $a \in H^1(\Omega)$ satisfying (1.1).

The OLS method, applied to the problem of estimating the Lamé moduli in the system of linear, isotropic elasticity (1.3), was analyzed by Gockenbach [11]. For the sake of convenience, he expressed the inverse problem in terms of the shear modulus μ^* and the bulk modulus $\rho^* = \mu^* + \lambda^*$ instead of in terms of μ^* and λ^* . The main

⁸If $a_h < 0$, then the finite element solution $u_h(a_h)$ is not guaranteed to exist. This condition is also physical: transmissivity by definition is positive.

results obtained by Gockenbach are derived under the following assumption on the strain:

$$\min \{ |\epsilon(u_E)_{12}|, |\text{tr}(\epsilon(u_E))| \} \geq c > 0. \quad (2.11)$$

It is assumed that $u_E \in W^{1,r+3}(\Omega)^2$ and $m^* = (\mu^*, \rho^*) \in W^{1,r+1}(\Omega)^2$, for some positive integer r . Condition (2.11) can be considered as a nondegeneracy condition on the experiment: at each point in Ω , the strain is neither a pure shear nor a pure expansion, and so, it is possible to estimate both the shear modulus and the bulk modulus. Similar to (2.10), condition (2.11) allows the PDE in (1.3) to be viewed as a hyperbolic PDE for m^* , which was first proved by Cox and Gockenbach in [10]. The following two results were proved in [11].

Lemma 2.2.1. *There exists a constant C , depending only on u_E and the constant c in (2.11), such that*

$$\|m\|_{L^2(\Omega)} \leq C \|\sigma(m, u_E)\|_{L^2(\Omega)}. \quad (2.12)$$

Lemma 2.2.2. *Suppose $u \in H^3(\Omega)^2$ satisfies (2.11), and $m \in H^1(\Omega)^2$. Then there exists $a > 0$ such that $v = \sigma(m, u)q$, $q(x) = (e^{ax_1}, e^{ax_2})$, satisfies*

$$\|\sigma(m, u)\|_{L^2(\Omega)} \leq C \left\{ \|\sigma(m, u)n\|_{L^2(\partial\Omega)} \|\sigma(m, u)\|_{L^2(\partial\Omega)} + \left| \int_{\Omega} \sigma(m, u) \cdot \epsilon(v) \right| \right\}. \quad (2.13)$$

The constant C depends on u and a but is independent of m .

Using the previous two results, it was shown that the inverse problem has a unique solution, precisely:

Corollary 2.2.1. *Suppose $u^* \in H^3(\Omega)^2$, and $m_1 = (\mu_1, \rho_1), m_2 = (\mu_2, \rho_2) \in H^1(\Omega)^2$ each satisfies, together with $u = u^*$, the BVP (1.3), then $m_1 = m_2$.*

Let \tilde{m} be the L^2 -projection of the true coefficient m_E into the finite element space $L_h^{(r)} \times L_h^{(r)}$. Let us fix an $m_h \in K_h^{(r)} \times K_h^{(r)}$, then by the triangle inequality and Lemma 2.2.1, we have

$$\begin{aligned} \|m_h - m_E\|_{L^2(\Omega)} &\leq \|\hat{m}\|_{L^2(\Omega)} + \|\tilde{m} - m_E\|_{L^2(\Omega)} \\ &\leq \|\sigma(\hat{m}, u_E)\|_{L^2(\Omega)} + \|\tilde{m} - m_E\|_{L^2(\Omega)}, \end{aligned} \quad (2.14)$$

where $\hat{m} = m_h - \tilde{m}$. By a standard approximation result (see [11] and the references

therein) we have

$$\|\tilde{m} - m_E\|_{L^2(\Omega)} \leq Ch^{r+1}.$$

In view of (2.13), in order to bound the first term on the right-hand side of (2.14) one needs to have some control over the term

$$\|\sigma(m_h, u_h(m_h))n - g_E\|_{L^2(\partial\Omega)}^2,$$

which is not given by the fact that $u_h(m_h)$ solves the weak form of the BVP (1.3). Therefore, the OLS problem considered in [11] takes the form

$$\min_{m \in K_h^{(r)} \times K_h^{(r)}} J_h(m) = \|u_h(m) - z_E\|_{L^2(\Omega)}^2 + h^3 \|\sigma(m, u_h(m))n - g_E\|_{L^2(\partial\Omega)}^2.$$

We cite the main result obtained in [11].

Theorem 2.2.2. *There exists a constant C such that if m_h is a minimizer of J_h , then*

$$\|m_h - m_E\|_{L^2(\Omega)} \leq C \left(h^r + \frac{\|z_E - u_E\|_{L^2(\Omega)}}{h^2} \right). \quad (2.15)$$

The constant C depends on $c_0, c_1, \nu, \|m_E\|_{W^{r+1,\infty}(\Omega)}$ and $\|u_E\|_{W^{r+3,\infty}(\Omega)}$, but is independent of h .

In the result above, it is assumed that the boundary edge traction g_E is chosen in such a way that the nondegeneracy condition (2.11) is satisfied.

Note that Theorem 2.2.2 provides a convergence proof if the data z_E is accurate enough. For example, if z_E is the exact interpolant of u_E in U_h^r , then by a standard approximation result

$$\|z_E - u_E\|_{L^2(\Omega)} \leq Ch^{r+2},$$

and therefore, the error bound (2.15) reduces to

$$\|m_h - m_E\|_{L^2(\Omega)} \leq Ch^r,$$

showing the convergence of the proposed method. However, for less accurate data, the error bound (2.15) blows up as $h \rightarrow 0$, mirroring the instability of the inverse problem. Similar discussion extends to Falk's result for the scalar inverse problem.

2.2.2 A variational method

A variational method for identifying the coefficient a_N in the BVP (1.1) was introduced by Kohn and Lowe [15]. It is motivated by the simple observation that, for any positive weights γ_1 and γ_2 , the minimum of the functional

$$F(a, \sigma) = \|\sigma - a\nabla u\|_{L^2(\Omega)}^2 + \gamma_1 \|\nabla \cdot \sigma + f_N\|_{L^2(\Omega)}^2 + \gamma_2 \|\sigma \cdot n - g_N\|_{L^2(\partial\Omega)}^2$$

is achieved only when $\sigma = a\nabla u$ with $a(x)$ a solution of (1.1). Their method is based on minimizing F numerically over suitable finite-dimensional spaces with u replaced by a measurement u^m . The weights γ_1 and γ_2 are chosen so that each term of sum has the same order of magnitude. The resulting optimization problem takes the form

$$\min_{\substack{\sigma \in U_h^{(0)} \\ a \in K_h^{(1)}}} \|\sigma - a\nabla u^m\|_{L^2(\Omega)}^2 + h^2 \|\nabla \cdot \sigma + f_N^m\|_{L^2(\Omega)}^2 + h \|\sigma \cdot n - g_N^m\|_{L^2(\partial\Omega)}^2, \quad (2.16)$$

where f_N^m and g_N^m are measurements for f_N and g_N .

In the case f_N and g_N are known exactly, Kohn and Lowe's main result reads as:

Theorem 2.2.3. *Suppose that $u_N \in H^3(\Omega)$, $\Delta u_N \in C(\overline{\Omega})$, and $a_N \in H^2(\Omega)$ with $0 < c_0 \leq a_N \leq c_1$. If a_h solves (2.16) then*

$$\|a_h - a_N\|_{L^2(\Omega)} \leq C \left(h + \frac{\|u_N - u^m\|_{H^1(\Omega)}}{h} \right). \quad (2.17)$$

The error bound in (2.17) was derived under the following assumption

$$\inf_{\Omega} \max \{ |\nabla u_N|, \Delta u_N \} > 0, \quad (2.18)$$

which is less restrictive a condition than the nondegeneracy condition (2.10). Actually, Richter [25] proved that (1.1) is uniquely solvable for a_N provided that u_N satisfies (2.18) and a_N is prescribed along the inflow portion of $\partial\Omega$, that is, the portion for which $\nabla u_N \cdot n < 0$.

Notice that if ∇u_N vanishes on an open subset of Ω , then condition (2.18) is violated and so the error bound (2.17) is not valid in this case. Even more, when ∇u_N vanishes on a set of positive measure (say an open set contained in Ω), then

(1.1) provides no information about the parameter a_N on this set. Kohn and Lowe tackled this case by considering the following regularized version of (2.16)

$$\min_{\substack{\sigma \in U_h^{(0)} \\ a \in L_h^{(1)}}} \|\sigma - a \nabla u^m\|_{L^2(\Omega)}^2 + h^2 \|\nabla \cdot \sigma + f_N^m\|_{L^2(\Omega)}^2 + h \|\sigma \cdot n - g_N^m\|_{L^2(\partial\Omega)}^2 + \alpha \|\nabla a\|_{L^2(\Omega)}^2, \quad (2.19)$$

where $\alpha > 0$ is a regularization parameter. If $\|u_N - u^m\|_{H^1(\Omega)} \leq \epsilon$ and α is chosen such that $\alpha \sim (h^2 + \epsilon)^2$, Kohn and Lowe showed that⁹

$$\int_{\Omega} |a_{h,\alpha} - a_N| |\nabla u_N|^2 \leq C (h + \epsilon h^{-1}), \quad (2.20)$$

for any optimizer $a_{h,\alpha}$ of (2.19). The term $|\nabla u_N|^2$ in (2.20) indicates that no information about the quality of the recovered coefficient can be obtained on those parts¹⁰ of the domain where $\nabla u_N = 0$. However, if $|\nabla u_N|^2 \geq c > 0$, then the above estimate is valid for the whole domain, and actually the term $|\nabla u_N|^2$ can be removed totally from the estimate:

$$\int_{\Omega} |a_{h,\alpha} - a_N| \leq c^{-1} \int_{\Omega} |a_{h,\alpha} - a_N| |\nabla u_N|^2 \leq \tilde{C} (h + \epsilon h^{-1}).$$

Chen and Gockenbach [9] have extended the variational method of Kohn and Lowe to the problem of estimating the Lamé moduli in (1.3) in the case $f_E = 0$. They considered the following optimization problem

$$\min_{\substack{\mu, \rho \in K_h^{(1)} \\ \sigma_{ij} \in L_h^{(1)}, i,j=1,2}} J(\sigma, \rho, \mu), \quad (2.21)$$

where

$$\begin{aligned} J(\sigma, \rho, \mu) = & \|\sigma_{11} - (\epsilon_{11}^m + \epsilon_{22}^m) \rho - (\epsilon_{11}^m - \epsilon_{22}^m) \mu\|_{L^2(\Omega)}^2 + \|\sigma_{12} - 2\epsilon_{12}^m \mu\|_{L^2(\Omega)}^2 \\ & + \|\sigma_{22} - (\epsilon_{11}^m + \epsilon_{22}^m) \rho + (\epsilon_{11}^m - \epsilon_{22}^m) \mu\|_{L^2(\Omega)}^2 + h^2 \|\nabla \sigma\|_{L^2(\Omega)}^2 \\ & + h \|\sigma n - g^m\|_{L^2(\partial\Omega)}^2. \end{aligned} \quad (2.22)$$

⁹To be precise, this result is obtained by specializing Kohn and Lowe's result to the case in which f_N and g_N are known exactly.

¹⁰More precisely, those parts of positive measure.

Here ϵ^m is some measurement of $\epsilon(u_E)$. Suppose that $\mu^*, \rho^* \in H^2(\Omega)$, $u_E \in H^3(\Omega)^2$ and satisfies (2.11) and

$$\nabla((\epsilon(u)_{11} - \epsilon(u)_{22})/\epsilon(u)_{12}) \in L^\infty(\Omega)^{2 \times 2},$$

Chen and Gockenbach proved the following result.

Theorem 2.2.4. *Suppose that $\sigma_h \in (L_h^{(1)})^{2 \times 2}$ and $\mu_h, \rho_h \in K_h^{(1)}$ satisfy*

$$J(\sigma_h, \rho_h, \mu_h) = \min_{\substack{\mu, \rho \in K_h^{(1)} \\ \sigma_{ij} \in L_h^{(1)}, i,j=1,2}} J(\sigma, \rho, \mu),$$

and that $u^m \in L_h^{(k)} \times L_h^{(k)}$ for some fixed integer $k \geq 2$, $\|u_E - u^m\|_{H^1(\Omega)} \leq \eta$, and $\|g_E - g^m\|_{L^2(\partial\Omega)} \leq \delta$. Then

$$\|\rho^* - \rho_h\|_{L^2(\Omega)} + \|\mu^* - \mu_h\|_{L^2(\Omega)} \leq C(h + \eta h^{-1} + h^{-1/2}\delta),$$

where C is independent of h , η , and δ .

2.2.3 Modified output-least squares (MOLS)

The MOLS method can be viewed as an OLS method with a coefficient dependent energy norm. To the best of our knowledge, it first appeared in the work of Zou [26], and then was independently proposed by Knowles [3]. Recently, Hào and Quyen [27] investigated MOLS subjected to Tikhonov regularization. We will briefly comment on the above works.

In the context of the BVP (1.2), the MOLS functional takes the form

$$J_{z_D}(a) = \int_{\Omega} a |\nabla(u(a) - z_D)|^2$$

where z_D is a measurement of u_D , and $u(a)$ is the solution of the forward problem with the true coefficient a_D replaced by a .

In [26], Zou added a regularization term to the MOLS objective function to stabilize the method since the underlying inverse problem is known to be ill-posed.

In the finite element settings, he considered the following optimization problem

$$\min_{a \in K_h^{(1)}} \frac{1}{2} \int_{\Omega} a |\nabla u_h(a) - \nabla z_D|^2 + \gamma N_h(a), \quad (2.23)$$

where $\gamma > 0$ is a regularization parameter, and $N_h(a)$ is a regularization term defined by

$$N_h(a) = \int_{\Omega} |\nabla a|^2 \quad \text{or} \quad N_h(a) = \int_{\Omega} \sqrt{|\nabla a|^2 + \delta(h)},$$

where $\delta(h)$ is a positive function satisfying $\lim_{h \rightarrow 0} \delta(h) = 0$. Zou proved the existence of a minimizer of the optimization problem (2.23), and explained how to solve it (numerically) by Armijo-type of algorithms. However, he did not provide any stability or convergence results regarding the inverse problem.

Knowles [3] considered slightly different problem. He assumes that the true coefficient a_D is known on the boundary of Ω , and he defined the set

$$D_G = \{a \in L^\infty(\Omega) \mid a \geq \nu > 0, \text{ and } a = a_D \text{ on } \Gamma \subset \partial\Omega\}$$

assuming sufficient regularity on a for the trace on Γ makes sense. He examined the following continuous (non-discretized) optimization problem

$$\min_{a \in D_G} G(a) = \int_{\Omega} \{a (|\nabla u_D|^2 - |\nabla u(a)|^2) - 2(u_D - u(a))f_D\}.$$

In [3, Theorem 2.1] Knowles proved that $G(a) = J_{u_D}(a)$ and it is strictly convex on the convex set D_G , and further that a_D is the only zero for the gradient ∇G . Hence under the above assumptions, a_D is the unique global minimum for G . He analyzed a numerical implementation using a preconditioned conjugate gradient approach. Note that Knowles assumes exact observation, i.e. u_D is known exactly, which is little bit impractical, since usually only noisy data is given.

Similar to Zou's work, Hào and Quyen [27] applied Tikhonov regularization to stabilize the MOLS, considering the optimization problem

$$\min_{a \in A} J_{z^\delta}(a) \int_{\Omega} a |\nabla(u(a) - z^\delta)|^2 + \rho \|a - a^*\|_{L^2(\Omega)}^2, \quad (2.24)$$

where $\rho > 0$ is the regularization parameter, a^* is an *a priori* estimate for the true

coefficient a_D , and z^δ is a measurement for u_D . Here the set A is defined by

$$A = \{a \in L^\infty(\Omega) \mid 0 < c_0 \leq a(x) \leq c_1 \text{ a.e. on } \Omega\},$$

and it is assumed that $a_D \in A$. They proved that the functional $J_{z^\delta}(a)$ is convex on the convex set A , and that there exists a unique solution a_ρ^δ of the problem (2.24). Further, they established the following continuity result for the solution a_ρ^δ with respect to the data z^δ .

Theorem 2.2.5. *For a fixed $\rho > 0$, let $z_n \rightarrow z^\delta$ in $H_0^1(\Omega)$ and $\{a_n\}$ be minimizers of $J_{z_n}(a)$. Then $a_n \rightarrow a_\rho^\delta$ in $L^2(\Omega)$.*

Hào and Quyen defined the set

$$\Pi = \{a \in A \mid u(a) = u_D\},$$

and they showed that Π is nonempty, bounded, and closed in the $L^2(\Omega)$ -norm, and that there is a unique solution a^\dagger of the problem

$$\min_{a \in \Pi} \|a - a^*\|_{L^2(\Omega)}^2$$

which is called the a^* -minimum norm solution of the identification problem. Hào and Quyen proved the following stability result.

Theorem 2.2.6.

$$\|a_\rho^\delta - a^\dagger\|_{L^2(\Omega)} = O(\sqrt{\delta}) \text{ and } \|u(a_\rho^\delta) - z^\delta\|_{L^2(\Omega)} = O(\delta)$$

as $\rho \rightarrow 0$ and $\rho \sim \delta$.

Chapter 3

Preliminaries

In this chapter, we recall some definitions and results from functional analysis and Sobolev space theory. We study the forward problems and prove results regarding existence, uniqueness, and stability.

3.1 Functional analysis

In this section we review some definitions and results from functional analysis which we shall need throughout this work. The reader may refer to [28–31] for more detailed discussions.

Let X and Y be normed spaces and $L : X \rightarrow Y$ be a linear operator. We say L is *bounded* if there exists a real number C such that

$$\|Lx\|_Y \leq C\|x\| \quad \forall x \in X.$$

If L is bounded, the smallest such C is denoted by $\|L\|$; thus

$$\|L\| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\|Lx\|_Y}{\|x\|_X}.$$

A standard functional analysis result states that L is bounded if and only if it is continuous. Let $\mathcal{L}(X; Y)$ be the set of all bounded linear operators from X into Y . $\mathcal{L}(X; Y)$ becomes a vector space with usual way operators are added and scaled, which also can be made into a normed space with the norm defined above.

A scalar-valued function on a vector space X is called *functional*. Let H be a Hilbert space (that is, a complete inner product space). The *dual space* of H , denoted by H^* , is the space of all bounded linear functionals defined on H . Sometimes, the evaluation of an $\ell \in H^*$ at a $v \in H$ will be written as $\langle \ell, v \rangle$. In some literature, the mapping $\langle \cdot, \cdot \rangle$ is called the *duality pairing* on $H^* \times H$ (or, the duality pairing between H^* and H).

Let us also recall the following fundamental results in Hilbert space theory.

Theorem 3.1.1 (The Riesz Representation Theorem). *Every bounded linear functional ℓ on a Hilbert space H can be represented uniquely as*

$$\ell(v) = (u, v)_H \quad \forall v \in H$$

where $u \in H$ depends on ℓ with $\|u\|_H = \|\ell\|_{H^*}$.

Conversely, for $u \in H$, the functional ℓ defined by

$$\ell(v) = (u, v)_H \quad \forall v \in H$$

belongs to H^* with $\|\ell\|_{H^*} = \|u\|_H$.

Theorem 3.1.2 (The Projection Theorem). *Let H be a Hilbert space, and U be a closed subspace of H . Then for any $v \in H$, there is a unique $u \in U$ such that*

$$\|u - v\| = \inf_{w \in U} \|w - v\|.$$

We call u the *projection of v onto U* ; for short we write $u = \text{proj}_U v$. Further, u is characterized by

$$(u - v, w)_H = 0 \quad \forall w \in U.$$

If X is an inner product space, the *Cauchy-Schwarz inequality* states that

$$|(u, v)_X| \leq \|u\|_X \|v\|_X \quad \forall u, v \in X,$$

where $\|u\|_X = \sqrt{(u, u)_X}$. Using the Riesz representation theorem and the Cauchy-Schwarz inequality, one can conclude the following theorem, which we will be needed in later chapters.

Theorem 3.1.3. *Let H be a real Hilbert space, and H^* be its dual space. Then the mapping $P : H \rightarrow H^*$ given by*

$$\langle Pv, w \rangle = (v, w)_H \quad \forall v, w \in H,$$

is an isometric isomorphism from H onto H^ . Moreover, we have*

$$\|\psi\|_{H^*}^2 = \langle \psi, P^{-1}\psi \rangle = \|P^{-1}\psi\|_H^2 \quad \forall \psi \in H^*. \quad (3.1)$$

We also recall the definition of compact (linear) operators and some of their properties.

Definition 3.1.1. *A linear operator K from a normed space X into a normed space Y is called **compact** if and only if it maps bounded sets in X to precompact sets in Y , i.e., if $M \subset X$ is bounded, then $\overline{K(M)}$ is compact in Y . Equivalently, K is compact if and only if for every bounded sequence $\{x_n\} \subset X$ the sequence $\{Kx_n\}$ has a convergent subsequence in Y .*

Remark 3.1.1. *It can be shown that every compact operator is continuous (bounded). The converse need not be true.*

The proof of the following theorem can be found in Kreyszig [29, Theorem 8.1-7].

Theorem 3.1.4. *Let X and Y be normed spaces and $K : X \rightarrow Y$ a compact linear operator. Suppose that $\{x_n\}$ in X is weakly convergent, say, $x_n \rightharpoonup x$. Then $\{Kx_n\}$ is strongly convergent in Y and has the limit $y = Kx$.*

We shall also need the following definitions and results from convex analysis; see for example [28] and references therein.

Definition 3.1.2. *Let V be a normed space, $K \subset V$. A function $f : K \rightarrow \mathbb{R}$ is called **lower semicontinuous** if $\{v_n\} \subset K$ and $v_n \rightarrow v \in K$ imply*

$$f(v) \leq \liminf_{n \rightarrow \infty} f(v_n).$$

*The function f is called **weakly lower semicontinuous** if the above inequality is valid for any sequence $\{v_n\} \subset K$ with $v_n \rightharpoonup v \in K$. Here the notation $v_n \rightharpoonup v$ means*

v_n converges to v weakly; that is

$$\ell(v_n) \rightarrow \ell(v) \quad \forall \ell \in \mathcal{L}(V; \mathbb{R}).$$

Theorem 3.1.5. A norm $\|\cdot\|$ on a normed space is weakly lower semicontinuous.

Definition 3.1.3. Let V be a real or complex linear space, $K \subset V$.

- The set K is said to be **convex** if

$$u, v \in K \implies \lambda u + (1 - \lambda)v \in K \quad \forall \lambda \in [0, 1].$$

Clearly, if K is a subspace of V , then K is convex.

- Assume K is convex. A function $f : K \rightarrow \mathbb{R}$ is said to be **convex** if

$$f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v) \quad \forall u, v \in K, \forall \lambda \in [0, 1].$$

The function f is **strictly convex** if the above inequality is strict for $u \neq v$ and $\lambda \in (0, 1)$.

- Assume V is a normed space. A real-valued function f on V is said to be **coercive** over K if

$$f(v) \rightarrow \infty \text{ as } \|v\| \rightarrow \infty, \quad v \in K.$$

Theorem 3.1.6. (a) If V is an inner product space, then the function $f(v) = \|v\|_V^2 = (v, v)_V$ is strictly convex.

(b) If f_c is convex on V and f_s is strictly convex on V , then the function $f = f_c + f_s$ is strictly convex on V .

Theorem 3.1.7. Assume V is a normed space, $K \subset V$ is a convex and closed finite-dimensional subset, and $f : K \rightarrow \mathbb{R}$ is convex and lower semi-continuous. If either

(a) K is bounded

or

(b) f is coercive on K ,
then the minimization problem

$$\inf_{v \in K} f(v)$$

has a solution. Furthermore, if f is strictly convex, then a solution to this problem is unique.

3.2 Sobolev spaces

In this section we recall some definitions and results from the theory of Sobolev spaces. The reader may refer to [5, 28, 32–37] for more detailed discussions. Throughout this section, Ω is a nonempty open subset of \mathbb{R}^N .

For any extended real number $1 \leq p \leq \infty$, the space $L^p(\Omega)$ consists of equivalence classes of measurable functions $f : \Omega \rightarrow \mathbb{R}$ such that

$$\begin{aligned} \int_{\Omega} |f|^p &< \infty, \quad \text{if } 1 \leq p < \infty, \\ \text{ess sup}_{x \in \Omega} |f(x)| &< \infty, \quad \text{if } p = \infty, \end{aligned} \tag{3.2}$$

where two measurable functions are equivalent if they are equal almost everywhere (a.e.) in Ω . Here the integral is to be understood in the Lebesgue sense, and¹

$$\text{ess sup}_{x \in \Omega} |f(x)| = \inf \{a \in \mathbb{R} \mid \mu(x : |f(x)| > a) = 0\}.$$

For convenience, it is customary not to make any distinction between a function and its equivalence class (except when the precise pointwise values of a representative function are significant). Thus, for a measurable function f , we will write $f \in L^p(\Omega)$ if f satisfies (3.2), and $f = 0$ in $L^p(\Omega)$ if $f(x) = 0$ a.e. in Ω .

The space $L^p(\Omega)$ is a Banach space with the norm defined by

$$\begin{aligned} \|f\|_{L^p(\Omega)} &= \left(\int_{\Omega} |f|^p \right)^{1/p}, \quad \text{when } 1 \leq p < \infty, \\ \|f\|_{L^\infty(\Omega)} &= \text{ess sup}_{x \in \Omega} |f(x)| < \infty, \quad \text{when } p = \infty. \end{aligned}$$

¹Here, $\mu(\cdot)$ denotes the Lebesgue measure in \mathbb{R}^N ; occasionally the notation $|\cdot|$ will also be used.

In particular, the space $L^2(\Omega)$ is a Hilbert space with inner product given by

$$(u, v)_{L^2(\Omega)} = \int_{\Omega} u(x)v(x)dx.$$

The definitions of the spaces $L^p(\Omega)^m$ and $L^p(\Omega)^{m \times n}$ are very similar. For example,

$$L^p(\Omega)^m = \{f \mid f_i \in L^p(\Omega), i = 1, 2, \dots, m\},$$

and the norm becomes

$$\begin{aligned} \|f\|_{L^p(\Omega)} &= \left(\sum_{i=1}^m \|f_i\|_{L^p(\Omega)}^p \right)^{1/p}, \quad 1 \leq p < \infty, \\ \|f\|_{L^\infty(\Omega)} &= \max \{ \|f_1\|_{L^\infty(\Omega)}, \|f_2\|_{L^\infty(\Omega)}, \dots, \|f_n\|_{L^\infty(\Omega)} \}. \end{aligned}$$

We shall need the following definition to introduce the concept of (weak) derivatives of ‘functions’ in the space $L^p(\Omega)$.

Definition 3.2.1 (Locally integrable). *A function $f : \Omega \rightarrow \mathbb{R}$ is said to be **locally integrable** on Ω if $f \in L^1(U)$ for every compact $U \subset \Omega$. In this case we write $f \in L^1_{\text{loc}}(\Omega)$.*

For a multi-index $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N) \in \mathbb{N}_0^N$ and a function $u : \Omega \rightarrow \mathbb{R}$, the notation $\partial^\alpha u$ will be used to denote the α^{th} partial derivative of u , that is,

$$\partial^\alpha u = \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_N^{\alpha_N}},$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_N$.

Let Ω_0 be a nonempty subset of \mathbb{R}^N , and let ϕ be any function defined on Ω_0 . The *support* of ϕ is defined to be the set

$$\text{supp}(\phi) = \overline{\{x \in \Omega_0 \mid \phi(x) \neq 0\}},$$

where the closure is taken in \mathbb{R}^N . If ϕ is also defined on Ω and $\text{supp}(\phi)$ is a compact subset of \mathbb{R}^N (that is, closed and bounded subset of \mathbb{R}^N) with $\text{supp}(\phi) \subset \Omega$, then ϕ is said to be *compactly supported* in Ω . The space $C_0^\infty(\Omega)$ is defined to be the set of all functions that are infinitely differentiable on Ω and compactly supported in Ω .

Now we are ready to introduce the concept of a weak derivative, one of the fundamental building blocks of Sobolev spaces.

Definition 3.2.2 (Weak derivative). *A function $u \in L^1_{\text{loc}}(\Omega)$ is said to have a **weak derivative** of an order α if there exists a function $v \in L^1_{\text{loc}}(\Omega)$ satisfying*

$$\int_{\Omega} v \phi = (-1)^{|\alpha|} \int_{\Omega} u \partial^{\alpha} \phi \quad \forall \phi \in C_0^{\infty}(\Omega).$$

In this case we write $v = \partial^{\alpha} u$.

It can be shown that weak derivatives are unique (at least, up to a set of measure zero). Moreover, if $u \in C^m(\Omega)$, then for all $|\alpha| \leq m$, the classical partial derivative $\partial^{\alpha} u$ coincides with α^{th} weak derivative of u . Of course, $\partial^{\alpha} u$ may exist in the weak sense without existing in the classical sense. See [28] for proofs and examples. Unless otherwise stated, from now on, all derivatives should be understood in the weak sense.

The Sobolev space $W^{m,p}(\Omega)$ is defined by

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) : \partial^{\alpha} u \in L^p(\Omega) \forall |\alpha| \leq m\},$$

This is a Banach space under the norm

$$\|u\|_{W^{m,p}(\Omega)} = \left(\sum_{|\alpha| \leq m} \|\partial^{\alpha} u\|_{L^p(\Omega)}^p \right)^{1/p},$$

with the appropriate modification for $p = \infty$. When $p = 2$, $W^{m,2}(\Omega)$ is usually denoted by $H^m(\Omega)$, which is a Hilbert space with inner product

$$(u, v)_{H^m(\Omega)} = \sum_{|\alpha| \leq m} \int_{\Omega} \partial^{\alpha} u \partial^{\alpha} v.$$

Also we mention following *semi-norm* on $H^m(\Omega)$

$$|u|_{H^m(\Omega)} = \sqrt{\sum_{|\alpha|=m} \|\partial^{\alpha} u\|_{L^2(\Omega)}^2}.$$

The space $H_0^1(\Omega)$ is defined to be the closure of $C_0^{\infty}(\Omega)$ in $H^1(\Omega)$. Hence $H_0^1(\Omega)$

is a closed subspace of $H^1(\Omega)$, and so it is also a Hilbert space under the H^1 -inner product. We denote by $H^{-1}(\Omega)$ the dual space of $H_0^1(\Omega)$. The norm on $H^{-1}(\Omega)$ is given by

$$\|\ell\|_{H^{-1}(\Omega)} = \sup_{v \in H_0^1(\Omega)} \frac{\langle \ell, v \rangle_{-1}}{\|v\|_{H_0^1(\Omega)}},$$

where $\langle \cdot, \cdot \rangle_{-1}$ denotes the duality pairing on $H^{-1}(\Omega) \times H_0^1(\Omega)$.

We denote by $H^1(\Omega)^*$ the dual space of $H^1(\Omega)$; the norm on $H^1(\Omega)^*$ is

$$\|\ell\|_{H^1(\Omega)^*} = \sup_{v \in H^1(\Omega)} \frac{\langle \ell, v \rangle_*}{\|v\|_{H^1(\Omega)}}$$

where $\langle \cdot, \cdot \rangle_*$ denotes the duality pairing on $H^1(\Omega)^* \times H^1(\Omega)$.

The following theorem shows how to interpret boundary values for functions from $W^{1,p}(\Omega)$; see for example [5, 28, 35–37].

Theorem 3.2.1 (Trace Theorem). *Let Ω be an open, bounded, connected subset of \mathbb{R}^d with Lipschitz-continuous boundary $\partial\Omega$, and $1 < p < \infty$. Then there exists a unique continuous linear operator $\gamma : W^{1,p}(\Omega) \rightarrow L^p(\partial\Omega)$ such that $\gamma v = v|_{\partial\Omega}$ for all $v \in C^\infty(\bar{\Omega})$. The operator γ is called the **trace operator**.*

We point out that γ is neither injective nor surjective. In the case $p = 2$, $\mathcal{R}(\gamma)$ is denoted by $H^{1/2}(\partial\Omega)$, which is a Hilbert space with the norm defined by

$$\|u\|_{H^{1/2}(\partial\Omega)} = \inf_{v \in H^1(\Omega)} \left\{ \|v\|_{H^1(\Omega)} \mid \gamma v = u \right\}.$$

From this definition one can conclude the inequality (better-known as *the trace inequality*)

$$\|\gamma u\|_{H^{1/2}(\partial\Omega)} \leq \|u\|_{H^1(\Omega)} \quad \forall u \in H^1(\Omega).$$

We denote by $H^{-1/2}(\partial\Omega)$ the dual space of $H^{1/2}(\partial\Omega)$, and we will write $\langle \cdot, \cdot \rangle_{\partial\Omega}$ to denote the duality pairing between $H^{-1/2}(\partial\Omega)$ and $H^{1/2}(\partial\Omega)$. The norm on $H^{-1/2}(\partial\Omega)$ is given by

$$\|g\|_{H^{-1/2}(\partial\Omega)} = \sup_{v \in H^{1/2}(\partial\Omega)} \frac{\langle g, v \rangle_{\partial\Omega}}{\|v\|_{H^{1/2}(\partial\Omega)}}.$$

Using the concept of traces, the space $H_0^1(\Omega)$ can be characterized as

$$H_0^1(\Omega) = \ker(\gamma) = \{u \in H^1(\Omega) \mid \gamma u = 0\}.$$

For $u \in H^1(\Omega)$, it is accustomed to write γu simply as u , which we will follow this convention throughout.

The divergence of a 2-tensor σ and the gradient of a vector-valued function v are defined respectively as

$$\nabla \cdot \sigma = \begin{bmatrix} \frac{\partial \sigma_{11}}{\partial x} + \frac{\partial \sigma_{12}}{\partial y} \\ \frac{\partial \sigma_{21}}{\partial x} + \frac{\partial \sigma_{22}}{\partial y} \end{bmatrix}, \quad \nabla v = \begin{bmatrix} \frac{\partial v_1}{\partial x} & \frac{\partial v_1}{\partial y} \\ \frac{\partial v_2}{\partial x} & \frac{\partial v_2}{\partial y} \end{bmatrix}.$$

The dot product of two 2-tensors σ and ϵ is defined by

$$\sigma \cdot \epsilon = \sigma_{11}\epsilon_{11} + \sigma_{12}\epsilon_{12} + \sigma_{21}\epsilon_{21} + \sigma_{22}\epsilon_{22}.$$

In the special case that σ is symmetric and $\epsilon = \frac{1}{2}(\nabla v + \nabla v^T)$, for some vector-valued function v , we have the identity

$$\sigma \cdot \nabla v = \sigma \cdot \nabla v^T = \sigma \cdot \epsilon. \quad (3.3)$$

We shall need the following Green's formulas (the multidimensional analogue of integration by parts); see for example [36, 38].

Theorem 3.2.2 (Green's Formulas). *Let Ω be a bounded open subset of \mathbb{R}^2 with Lipschitz-continuous boundary $\partial\Omega$, and $n = (n_1, n_2)$ denote the outer unit normal to $\partial\Omega$.*

(a) *For $u, v \in H^1(\Omega)$ and for $i = 1, 2$, we have*

$$\int_{\Omega} u \partial_{x_i} v = - \int_{\Omega} v \partial_{x_i} u + \int_{\partial\Omega} u v n_i. \quad (3.4)$$

(b) *For $u \in H^2(\Omega)$ and $v \in H^1(\Omega)$, we have*

$$- \int_{\Omega} \nabla \cdot (a \nabla u) v = \int_{\Omega} a \nabla u \cdot \nabla v - \int_{\partial\Omega} a \frac{\partial u}{\partial n} v, \quad (3.5)$$

provided the function a is smooth enough, e.g., in the space $C^1(\bar{\Omega})$, so that the function $a\nabla u$ belongs to the space $H^1(\Omega)^2$.

(c) If $\sigma \in H^1(\Omega)^{2 \times 2}$ is a symmetric tensor and $v \in H^1(\Omega)^2$, we have

$$-\int_{\Omega} (\nabla \cdot \sigma) \cdot v = \int_{\Omega} \sigma \cdot \epsilon_v - \int_{\partial\Omega} v \cdot (\sigma n), \quad (3.6)$$

where $\epsilon_v = \frac{1}{2} (\nabla v + \nabla v^T)$.

For scalar-valued function $f \in H^1(\Omega)$ and vector valued function $\vec{f} \in H^1(\Omega)^2$ the operators curl and rot are defined by

$$\text{curl } f = (\partial_y f, -\partial_x f), \quad \text{rot } \vec{f} = \partial_x f_2 - \partial_y f_1.$$

Notice then using Green's formula (3.4), for all $\vec{f} \in H^1(\Omega)^2$ and $v \in H_0^1(\Omega)$ we have

$$\begin{aligned} (\text{rot } \vec{f}, v)_{L^2(\Omega)} &= \int_{\Omega} (\partial_x f_2 - \partial_y f_1) v \\ &= \int_{\Omega} \vec{f} \cdot (\partial_y v, -\partial_x v) \\ &= (\vec{f}, \text{curl } v)_{L^2(\Omega)}. \end{aligned} \quad (3.7)$$

The right-hand side of (3.7) makes sense even if \vec{f} is only in $L^2(\Omega)^2$. Thus for $\vec{f} \in L^2(\Omega)^2$ we regard $\text{rot } \vec{f}$ as an element of $H^{-1}(\Omega)$ via the duality

$$\langle \text{rot } \vec{f}, v \rangle_{H^{-1}(\Omega) \times H_0^1(\Omega)} = (\vec{f}, \text{curl } v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega).$$

Notice that $\text{rot } \vec{f} \in H^{-1}(\Omega)$ since

$$\|\text{rot } \vec{f}\|_{H^{-1}(\Omega)} = \sup_{v \in H_0^1(\Omega)} \frac{(\vec{f}, \text{curl } v)_{L^2(\Omega)}}{\|v\|_{H^1(\Omega)}} \leq \|\vec{f}\|_{L^2(\Omega)},$$

which follows from the Cauchy-Schwarz inequality and the fact that, in \mathbb{R}^2 ,

$$\|\text{curl } v\|_{L^2(\Omega)} = \|\nabla v\|_{L^2(\Omega)} \leq \|v\|_{H^1(\Omega)}.$$

Moreover, it can easily be shown that the operator $\text{rot} : L^2(\Omega)^2 \rightarrow H^{-1}(\Omega)$ is linear, and it is bounded since

$$\|\text{rot}\|_{\mathcal{L}(H^{-1}(\Omega); H_0^1(\Omega))} = \sup_{\vec{f} \in L^2(\Omega)^2} \frac{\|\text{rot } \vec{f}\|_{H^{-1}(\Omega)}}{\|\vec{f}\|_{L^2(\Omega)}} \leq 1.$$

We shall also need the following fundamental theorem, which is a special case of the so-called Sobolev embedding theorems; see for example [5, 32, 37].

Theorem 3.2.3 (Rellich's). *Let Ω be a bounded open subset of \mathbb{R}^2 with Lipschitz-continuous boundary. Then $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$, that is, the identity operator $I : H^1(\Omega) \rightarrow L^2(\Omega)$ is compact.*

Let $\{u_n\}$ be any bounded sequence in $H^1(\Omega)$. From functional analysis, every bounded sequence in a Hilbert space has a weakly convergent subsequence. Thus, there exists a subsequence $\{u_{n'}\}$ of $\{u_n\}$ and $u \in H^1(\Omega)$ such that $u_{n'} \rightharpoonup u$ in $H^1(\Omega)$. By Rellich's theorem, there exists a subsequence $\{u_{n'_k}\}$ of $\{u_{n'}\}$ and a vector $w \in L^2(\Omega)$ such that $u_{n'_k} \rightarrow w$ in $L^2(\Omega)$. Since $L^2(\Omega)^* \subset H^1(\Omega)^*$, it follows that $\ell(u_{n'_k}) \rightarrow \ell(u)$ for all $\ell \in L^2(\Omega)$, and so, $u_{n'_k} \rightarrow u$ in $L^2(\Omega)$. Since the weak limit is unique, it follows that $w = u \in H^1(\Omega)$. We summarize this and another result in the following corollary.

Corollary 3.2.1. *Let Ω be a bounded open subset of \mathbb{R}^2 with Lipschitz-continuous boundary, and let $\{u_n\}$ be any bounded sequence in $H^1(\Omega)$.*

- (a) *There exists a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ and a vector $u \in H^1(\Omega)$ such that $u_{n_k} \rightharpoonup u$ in $H^1(\Omega)$ and $u_{n_k} \rightarrow u$ in $L^2(\Omega)$.*
- (b) *If further $u_n \rightharpoonup u$ in $H^1(\Omega)$, then by Theorem 3.1.4 we have $u_n \rightarrow u$ in $L^2(\Omega)$.*

3.3 Weak formulations of the forward problems

One popular approach to the study of partial differential equations is to transform the original problem, which is typically given in a strong form, into another form called the weak or variational form of the PDE. It turns out that most of the questions related to well-posedness can then be answered more satisfactorily, and to some

extent, in a more unified framework. For introductory texts in the subject, we recommend the books [28, 39]. A more comprehensive treatment can be found in the monographs [36, 37].

In this section, we will derive and prove the well-posedness of the weak formulations corresponding to the boundary value problems (1.1), (1.2), and (1.3).

From now on, we assume Ω is a bounded, simply-connected domain in \mathbb{R}^2 with Lipschitz-continuous boundary $\partial\Omega$, and n denotes the outer unit normal to $\partial\Omega$. All the smoothness assumptions regarding the functions involved will be given in the appropriate places.

3.3.1 Abstract variational problems

We will develop a functional analysis framework for studying abstract variational problems in Hilbert space settings. We start with some definitions and results.

Definition 3.3.1. A **bilinear form** $a(\cdot, \cdot)$ on a linear vector space V is a mapping $a : V \times V \rightarrow \mathbb{R}$ that satisfies the properties

1. $a(\alpha u + \beta v, w) = \alpha a(u, w) + \beta a(v, w)$,
2. $a(w, \alpha u + \beta v) = \alpha a(w, u) + \beta a(w, v)$,

for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$. If in addition, $a(u, v) = a(v, u)$ for all $u, v \in V$, then $a(\cdot, \cdot)$ is called **symmetric**.

Clearly any real inner product defines a symmetric bilinear form.

Definition 3.3.2. Assume V is a normed space. A bilinear form $a(\cdot, \cdot)$ on V is called

- (a) ***V-elliptic***, or for short ***elliptic*** or ***coercive***, if there exists a constant $\alpha > 0$ such that

$$a(v, v) \geq \alpha \|v\|_V^2 \quad \forall v \in V.$$

- (b) ***bounded***, if there exists a constant $\beta > 0$ such that

$$a(u, v) \leq \beta \|u\|_U \|v\|_U \quad \forall u, v \in U.$$

Let H be a Hilbert space, and suppose that $a(\cdot, \cdot)$ is a symmetric bilinear form on H which is bounded and H -elliptic. Let $\ell \in H^*$, and consider the abstract problem

$$u \in H, \quad a(u, v) = \ell(v) \quad \forall v \in H. \quad (3.8)$$

We will now explain how the Riesz representation theorem can be used to answer questions about existence, uniqueness, and stability of the abstract problem (3.8).

Since the bilinear form $a(\cdot, \cdot)$ is symmetric and H -elliptic, it is easy to show that it defines an alternate inner product on H . Let $\|\cdot\|_a$ denotes the norm induced by this new inner product, that is

$$\|v\|_a = \sqrt{a(v, v)} \quad \forall v \in H.$$

In some contexts, the norm $\|\cdot\|_a$ is called the *energy norm*. From the boundedness and ellipticity of $a(\cdot, \cdot)$, it follows that $\|\cdot\|_H$ is a norm equivalent to the norm $\|\cdot\|_a$. In fact

$$\sqrt{\alpha}\|v\|_H \leq \|v\|_a \leq \sqrt{\beta}\|v\|_H \quad \forall v \in H, \quad (3.9)$$

where α and β are the constants appearing in the definitions of H -ellipticity and boundedness of $a(\cdot, \cdot)$, respectively. Thus, H is also complete under the energy norm, and hence it is a Hilbert space with respect to the inner product $a(\cdot, \cdot)$.

Since by assumption ℓ is bounded with respect to the original norm on H , we have

$$|\ell(v)| \leq \|\ell\|_{H^*} \|v\|_H \leq \alpha^{-1/2} \|\ell\|_{H^*} \|v\|_a \quad \forall v \in H,$$

and so ℓ is also bounded with respect to the energy norm with

$$\|\ell\|_{a^*} \leq \alpha^{-1/2} \|\ell\|_{H^*}, \quad (3.10)$$

where $\|\ell\|_{a^*}$ denotes the norm of ℓ with respect to the energy norm $\|\cdot\|_a$.

Thus, by Riesz representation theorem, there exists a unique vector $u \in H$ satisfying the abstract problem (3.8). Further, $\|u\|_a = \|\ell\|_{a^*}$, and so in view of (3.9) and (3.10) we see that

$$\|u\|_H \leq \alpha^{-1} \|\ell\|_{H^*},$$

which expresses the *stability* or *continuous dependence* of the solution u with respect

to the data, ℓ .

We summarize what we have established in the following lemma.

Lemma 3.3.1. *Let H be a Hilbert space. Suppose that $a(\cdot, \cdot)$ is a symmetric bilinear form on H that is bounded and H -elliptic, and that $\ell \in H^*$. Then there exists a unique $u \in H$ such that*

$$a(u, v) = \ell(v) \quad \forall v \in H.$$

Furthermore, the solution u depends continuously on the data ℓ in the sense that

$$\|u\|_H \leq C \|\ell\|_{H^*}$$

for some constant C independent of u and ℓ .

3.3.2 The Neumann BVP

To derive the weak formulation corresponding to BVP (1.1), we temporarily assume that $u \in C^2(\Omega) \cap C^1(\bar{\Omega})$ is a classical solution of the BVP (1.1). Further, let us also assume that $a_N \in C^1(\bar{\Omega})$, $f_N \in C(\bar{\Omega})$, and $g_N \in C(\partial\Omega)$, so that the following computations are valid. Multiplying both sides of the PDE in (1.1) by a function $v \in C^1(\bar{\Omega})$ (the so-called test function) then integrating the resulted equation over Ω yields

$$\int_{\Omega} -\nabla \cdot (a_N \nabla u) v = \int_{\Omega} f_N v.$$

Next apply Green's identity (3.5) to left-hand side of this relation and use the fact that $a_N \frac{\partial u}{\partial n} = g_N$ on $\partial\Omega$ to obtain

$$\int_{\Omega} a_N \nabla u \cdot \nabla v = \int_{\Omega} f_N v + \int_{\partial\Omega} g_N v. \quad (3.11)$$

By taking $v = 1$ in equation (3.11), we see that

$$\int_{\Omega} f_N + \int_{\partial\Omega} g_N = 0,$$

which gives a compatibility condition on the data f_N and g_N ; no solution exists unless f_N and g_N satisfy this condition. Moreover, if $c \in \mathbb{R}$, then $u + c$ is another solution, and hence, the BVP (1.1) has no unique solution.

Equation (3.11) was derived under the assumption that $u \in C^2(\Omega) \cap C^1(\bar{\Omega})$, provided that $a_N \in C^1(\bar{\Omega})$, $f_N \in C(\bar{\Omega})$, and $g_N \in C(\partial\Omega)$. However, in order for equation (3.11) to make sense, we only need to assume $u, v \in H^1(\Omega)$, $a_N \in L^\infty(\Omega)$, $f_N \in L^2(\Omega)$, and $g_N \in L^2(\partial\Omega)$. Actually, we can still weaken the smoothness assumption on g_N . It is sufficient to assume $g_N \in H^{-1/2}(\partial\Omega)$, as long as we interpret the integral $\int_{\partial\Omega} g_N v$ as the duality pairing $\langle g_N, v \rangle_{\partial\Omega}$ between $H^{-1/2}(\partial\Omega)$ and $H^{1/2}(\partial\Omega)$. Therefore the weak formulation of the BVP (1.1) can be posed as

$$u \in H^1(\Omega), \quad (a_N \nabla u, \nabla v)_{L^2(\Omega)} = (f_N, v)_{L^2(\Omega)} + \langle g_N, v \rangle_{\partial\Omega} \quad \forall v \in H^1(\Omega). \quad (3.12)$$

Notice the weak formulation has the same properties as the strong formulation (1.1). If u solves (3.12), then adding a constant to u produces another solution, and hence (3.12) has no unique solution. Further, by taking $v = 1$ in (3.12), we get the following compatibility condition on the data f_N and g_N

$$(f_N, 1)_{L^2(\Omega)} + \langle g_N, 1 \rangle_{\partial\Omega} = 0. \quad (3.13)$$

Clearly this condition is necessary for a solution to exist. Indeed, later we will show that this condition is also sufficient.

We start with some auxiliary results. Let us define the subspace $V \subset H^1(\Omega)$ by

$$V = \{v \in H^1(\Omega) \mid (v, 1)_{L^2(\Omega)} = 0\}.$$

We will show that V is closed. Define the operator $L : H^1(\Omega) \rightarrow \mathbb{R}$ by

$$L(v) = (v, 1)_{L^2(\Omega)} \quad \forall v \in H^1(\Omega).$$

Clearly L is linear, and further it is bounded since by the Cauchy-Schwarz inequality

$$|L(v)| \leq \|1\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \leq |\Omega|^{1/2} \|v\|_{H^1(\Omega)} \quad \forall v \in H^1(\Omega),$$

where $|\Omega|$ denotes the (Lebesgue) measure of Ω in \mathbb{R}^2 . Since $V = \mathcal{N}(L)$, and L is linear and bounded, it follows from a standard result that V is a closed subspace of $H^1(\Omega)$, and hence it is also a Hilbert space. Now consider the variational problem

$$u \in V, \quad a(u, v) = \ell(v) \quad \forall v \in V, \quad (3.14)$$

where

$$\begin{aligned} a(u, v) &= (a_N \nabla u, \nabla v)_{L^2(\Omega)} \quad \forall u, v \in V, \\ \ell(v) &= (f_N, v)_{L^2(\Omega)} + \langle g_N, v \rangle_{\partial\Omega} \quad \forall v \in V. \end{aligned}$$

It can be easily seen that $a(\cdot, \cdot)$ defines a symmetric bilinear form on V . It is also bounded since

$$\begin{aligned} a(u, v) &\leq \int_{\Omega} |a_N \nabla u \cdot \nabla v| \\ &\leq \|a_N\|_{L^\infty(\Omega)} \int_{\Omega} |\nabla u \cdot \nabla v| \quad (\text{H\"older's inequality}) \\ &\leq \|a_N\|_{L^\infty(\Omega)} \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \quad (\text{Cauchy-Schwarz inequality}) \\ &\leq \|a_N\|_{L^\infty(\Omega)} \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \quad \forall u, v \in V. \end{aligned}$$

To show that $a(\cdot, \cdot)$ is V -elliptic, we need the following result which can be concluded from [28, Theorem 7.3.12] (see also, [40, Theorem 2.6]).

Lemma 3.3.2. *Let Ω be an open, bounded, connected subset of \mathbb{R}^d with Lipschitz-continuous boundary. The function $\|\cdot\| : H^1(\Omega) \rightarrow \mathbb{R}$ given by*

$$\|v\| = \|\nabla v\|_{L^2(\Omega)} + \|(v, 1)_{L^2(\Omega)}\|$$

defines a norm on $H^1(\Omega)$, which is equivalent to the norm $\|v\|_{H^1(\Omega)}$. More precisely, there exist constants $M \geq m > 0$ depending only on Ω such that

$$m\|v\| \leq \|v\|_{H^1(\Omega)} \leq M\|v\| \quad \forall v \in H^1(\Omega).$$

As a corollary, we cite²:

Corollary 3.3.1. *Over the space V , the semi-norm $|\cdot|_{H^1(\Omega)}$ is a norm equivalent to the norm $\|\cdot\|_{H^1(\Omega)}$.*

Thus for all $u \in V$ we have

$$a(u, u) \geq k \|\nabla u\|_{L^2(\Omega)}^2 = k |u|_{H^1(\Omega)}^2 \geq \alpha \|u\|_{H^1(\Omega)}^2,$$

²We point out that the same result can be concluded from *Friedrich's inequality* (see for example, [33, 41]), but we need different (and stronger) assumptions on Ω .

where $\alpha = kM^{-2}$, and so $a(\cdot, \cdot)$ is V -elliptic. Notice, however, that $a(\cdot, \cdot)$ is not $H^1(\Omega)$ -elliptic (take $v = 1$; then $a(v, v) = 0$ but $\|v\|_{H^1(\Omega)} \neq 0$). This is one of the reasons of introducing the space V .

Clearly ℓ is linear. Moreover, using the Cauchy-Schwarz and trace inequalities

$$\begin{aligned} |\ell(v)| &\leq \|f_N\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + \|g_N\|_{H^{-1/2}(\partial\Omega)} \|v\|_{H^{1/2}(\partial\Omega)} \\ &\leq (\|f_N\|_{L^2(\Omega)} + \|g_N\|_{H^{-1/2}(\partial\Omega)}) \|v\|_{H^1(\Omega)} \quad \forall v \in V, \end{aligned}$$

showing that ℓ is bounded with

$$\|\ell\|_{V^*} \leq \|f_N\|_{L^2(\Omega)} + \|g_N\|_{H^{-1/2}(\partial\Omega)}.$$

Now we can apply Lemma 3.3.1, to conclude that the variational problem (3.14) has a unique solution $u \in V$, provided that $a_N \in L^\infty(\Omega)$ and is strictly positive over Ω . Moreover, we have the stability result

$$\|u\|_{H^1(\Omega)} \leq \alpha^{-1} (\|f_N\|_{L^2(\Omega)} + \|g_N\|_{H^{-1/2}(\partial\Omega)}). \quad (3.15)$$

Finally, we show that if u is the solution to the variational problem (3.14), then u also solves the variational problem (3.12) provided the compatibility condition (3.13) is satisfied. Write any $v \in H^1(\Omega)$ as $v = v_{||} + v_{\perp}$ where

$$\begin{aligned} v_{||} &= v - |\Omega|^{-1}(v, 1)_{L^2(\Omega)}, \\ v_{\perp} &= |\Omega|^{-1}(v, 1)_{L^2(\Omega)}. \end{aligned}$$

Obviously $v_{\perp} \in \mathbb{R}$, it can be easily shown that $v_{||} \in V$ and (and that $(v_{||}, v_{\perp})_{H^1(\Omega)} = 0$, and so we have the orthogonal decomposition $H^1(\Omega) = V \oplus \mathbb{R}$). Consequently,

$$\begin{aligned} (a_N \nabla u, \nabla v)_{L^2(\Omega)} &= (a_N \nabla u, \nabla v_{||})_{L^2(\Omega)} = a(u, v_{||}) \\ &= \ell(v_{||}) = (f_N, v_{||})_{L^2(\Omega)} + \langle g_N, v_{||} \rangle_{\partial\Omega} \\ &= (f_N, v)_{L^2(\Omega)} + \langle g_N, v \rangle_{\partial\Omega} + v_{\perp} \left((f_N, 1)_{L^2(\Omega)} + \langle g_N, 1 \rangle_{\partial\Omega} \right) \\ &= (f_N, v)_{L^2(\Omega)} + \langle g_N, v \rangle_{\partial\Omega} \quad \forall v \in H^1(\Omega) \end{aligned}$$

showing that u is a solution to the variational problem (3.12), and further, it is the only solution to (3.12) that lies in the space V . The continuous dependence of u on

the data is shown in (3.15).

3.3.3 The Dirichlet BVP

Now we derive the weak formulation of the BVP (1.2). Let us assume that u satisfies (1.2). Multiply both sides of the PDE in (1.2) by a test function $v \in H_0^1(\Omega)$, then integrate over Ω to get

$$-\int_{\Omega} \nabla(a_D \nabla u_D) v = \int_{\Omega} f_D v.$$

Now apply Green's identity (3.5) to left-hand side then rearrange to get

$$\int_{\Omega} a_D \nabla u_D \cdot \nabla v = \int_{\Omega} f_D v + \int_{\partial\Omega} a_D \frac{\partial u_D}{\partial n} v. \quad (3.16)$$

Since $v \in H_0^1(\Omega)$, v vanishes on $\partial\Omega$ and so the second integral in the right-hand side of (3.16) is zero. Thus the weak formulation of the BVP (1.2) can be posed as

$$u \in H_0^1(\Omega), \quad \int_{\Omega} a_D \nabla u \cdot \nabla v = \int_{\Omega} f_D v \quad \forall v \in H_0^1(\Omega). \quad (3.17)$$

Notice how the homogeneous Dirichlet boundary conditions are explicitly imposed in the weak form. For this reason, Dirichlet conditions are often called *essential* boundary conditions.

Now we show that (3.17) is a well-posed variational problem. We shall assume $a_D \in L^\infty(\Omega)$ and is strictly positive over Ω , say $a_D \geq k > 0$, and that $f_D \in L^2(\Omega)$. Define the auxiliary functions $a(\cdot, \cdot)$ and $\ell(\cdot)$ by

$$\begin{aligned} a(u, v) &= \int_{\Omega} a_D \nabla u \cdot \nabla v \quad \forall u, v \in H_0^1(\Omega), \\ \ell(v) &= \int_{\Omega} f_D v \quad \forall v \in H_0^1(\Omega). \end{aligned}$$

So the weak formulation (3.17) can be written more concisely as

$$u \in H_0^1(\Omega), \quad a(u, v) = \ell(v) \quad \forall v \in H_0^1(\Omega).$$

Clearly $a(\cdot, \cdot)$ defines a symmetric bilinear form on $H_0^1(\Omega)$. We show it is also $H_0^1(\Omega)$ -elliptic. We need the following result, see for example [33, 41].

Theorem 3.3.1 (Poincaré Inequality). *Suppose Ω is a bounded set in \mathbb{R}^N . Then there exists a positive constant C , depending only on the domain Ω , such that*

$$\|v\|_{H^1(\Omega)} \leq C |v|_{H^1(\Omega)} \quad \forall v \in H_0^1(\Omega).$$

Using the Poincaré inequality and the fact that $a_D \geq k > 0$ a.e. on Ω , we see that

$$a(v, v) \geq k |v|_{H^1(\Omega)}^2 \geq \alpha \|v\|_{H^1(\Omega)}^2 \quad \forall v \in H_0^1(\Omega),$$

where $\alpha = k C^{-2}$, and thus, $a(\cdot, \cdot)$ is $H_0^1(\Omega)$ -elliptic. It is a straightforward argument to show that $a(\cdot, \cdot)$ and ℓ are both bounded, with

$$\|\ell\|_{H^{-1}(\Omega)} \leq \|f\|_{L^2(\Omega)}.$$

By Lemma 3.3.1, it follows that the variational problem (3.17) has unique solution $u \in H_0^1(\Omega)$, which also depends continuously on the data f .

3.3.4 The equations of isotropic elasticity

Assume u satisfies the equations of linear, isotropic elasticity (1.3). Multiplying both sides of the PDE in (1.3) by a smooth test function v , integrating the relation over Ω , then applying Green's formula (3.6), yields

$$\int_{\Omega} \sigma(m^*, u) \cdot \epsilon(v) = \int_{\Omega} f_E \cdot v + \int_{\partial\Omega} g_E \cdot v. \quad (3.18)$$

In order for every term in (3.18) to make sense, we only need to assume $u, v \in H^1(\Omega)^2$, provided $m^* = (\lambda^*, \rho^*) \in L^\infty(\Omega)^2$, $f_E \in H^1(\Omega)^2$, and³ $g_E \in H^{-1/2}(\partial\Omega)^2$. Thus the weak formulation of (1.3) can be posed as

$$u \in H^1(\Omega), \quad \int_{\Omega} \sigma(m^*, u) \cdot \epsilon(v) = \int_{\Omega} f_E \cdot v + \int_{\partial\Omega} g_E \cdot v \quad \forall v \in H^1(\Omega)^2. \quad (3.19)$$

Notice that for a solution to exist, the load f_E and the traction g_E must satisfy compatibility condition

$$\int_{\Omega} f_E + \int_{\partial\Omega} g_E = 0. \quad (3.20)$$

³Provided the proper interpretation of the integral.

Furthermore, since $\epsilon(v) = 0$ if and only if $v \in N$, where

$$N = \{(c_1 y + c_2, -c_1 x + c_3) \mid c_1, c_2, c_3 \in \mathbb{R}\},$$

we see that the BVP cannot have a unique solution. If we define the space U by

$$U = \left\{ v \in H^1(\Omega)^2 \mid \int_{\Omega} v_1 = \int_{\Omega} v_2 = 0, \int_{\Omega} \operatorname{rot} v = 0 \right\},$$

then we have the (not necessarily orthogonal) decomposition $H^1(\Omega)^2 = U \oplus N$.

Consider the variational problem:

$$u \in U, \quad a(u, v) = \int_{\Omega} \sigma(m^*, u) \cdot \epsilon(v) = \int_{\Omega} f_E \cdot v + \int_{\partial\Omega} g_E \cdot v = \ell(v) \quad \forall v \in U.$$

By [7, Chapter 3], one can conclude the following version of what is known as Korn's inequality

$$\|\epsilon(v)\|_{L^2(\Omega)}^2 \geq c \|v\|_{H^1(\Omega)}^2 \quad \forall v \in U,$$

where $c > 0$ depends only on Ω . If $k = 2 \min_{\Omega} \{\mu^*, \rho^*\}$, then a straightforward calculation shows that

$$(2\mu^* \epsilon + \lambda^* \operatorname{tr}(\epsilon) I) \cdot \epsilon \geq k \epsilon \cdot \epsilon \quad \forall \epsilon.$$

Assuming $k > 0$, which is reasonable since μ^* is the shear modulus of the elastic material and $\rho^* = \mu^* + \lambda^*$ is the bulk modulus, we have

$$a(v, v) \geq k \int_{\Omega} \epsilon(v) \cdot \epsilon(v) \geq ck \|v\|_{H^1(\Omega)}^2 \quad \forall v \in U.$$

Moreover, it is straightforward to verify that there exists $c_1, c_2 > 0$ such that

$$\|\sigma(m, w)\|_{L^2(\Omega)} \leq c_1 \|m\|_{L^2(\Omega)} \|w\|_{W^{1,\infty}(\Omega)} \quad \forall m \in L^2(\Omega)^2, w \in W^{1,\infty}(\Omega)^2, \quad (3.21)$$

$$\|\sigma(m, w)\|_{L^2(\Omega)} \leq c_2 \|m\|_{L^\infty(\Omega)} \|w\|_{H^1(\Omega)} \quad \forall m \in L^\infty(\Omega)^2, w \in H^1(\Omega)^2. \quad (3.22)$$

Thus, provided $m^* \in L^\infty(\Omega)^2$, there exists $\beta > 0$ such that

$$|a(u, v)| \leq \beta \|m^*\|_{L^\infty(\Omega)} \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \quad \forall m^* \in L^\infty(\Omega)^2, u, v \in H^1(\Omega)^2.$$

Thus for any $m^* = (\mu^*, \rho^*) \in L^\infty(\Omega)^2$ satisfying $\min_\Omega \{\mu^*, \rho^*\} > 0$ in Ω , the symmetric bilinear form $a(\cdot, \cdot)$ is bounded and U -elliptic. Further, it is easy to see that ℓ is bounded. Hence, by Lemma 3.3.1 it follows that

$$u \in U, \quad a(u, v) = \ell(v) \quad \forall v \in U$$

is a well-posed variational problem. As in the case of the Neumann BVP, it can be established that (3.19) has a unique solution that belongs to U .

Chapter 4

The equation error method

In this chapter, we give the precise formulation of the equation error method. We prove existence and uniqueness results. We then describe how to implement the equation error method in a practical algorithm.

Throughout the rest of this work we assume:

- $z_N \in W^{1,\infty}(\Omega)$, $u_N \in H^1(\Omega)$, and $f_N \in L^2(\Omega)$, $g_N \in H^{-1/2}(\partial\Omega)$ satisfy the compatibility condition (3.13);
- $z_D \in W^{1,\infty}(\Omega)$, $u_D \in H_0^1(\Omega)$, and $f_D \in L^2(\Omega)$;
- $z_E \in W^{1,\infty}(\Omega)^2$, $u_E \in H^1(\Omega)^2$, and $f_E \in L^2(\Omega)$, $g_E \in H^{-1/2}(\partial\Omega)^2$ satisfy the compatibility condition (3.20).

4.1 Motivation

First we will explain the equation error method in context of the BVP (1.1). Since the exact values $a = a_N$ and $u = u_N$ make

$$\nabla \cdot (a \nabla u) + f_N = 0 \text{ in } \Omega,$$

we wish to choose a to make

$$\nabla \cdot (a \nabla z_N) + f_N$$

as small as possible, i.e. minimize the error in the equation. If we include the boundary condition, we want to make both

$$\nabla \cdot (a \nabla z_N) + f_N \text{ and } a \frac{\partial z_N}{\partial n} - g_N$$

small. In this strong form it is not clear how to combine these two terms in a single objective function to minimize. However, the weak form (3.12) combines the PDE and the boundary condition in a single equation. Notice that the left-hand side of (3.12) makes sense for $a_N \nabla u_N \in L^2(\Omega)^2$ (for instance, if $a_N \in L^2(\Omega)$ and $u_N \in W^{1,\infty}(\Omega)$ or if $a_N \in L^\infty(\Omega)$ and $u_N \in H^1(\Omega)$). Define the functional ℓ^N by

$$\langle \ell^N, v \rangle_* = (f_N, v)_{L^2(\Omega)} + \langle g_N, v \rangle_{\partial\Omega} \quad \forall v \in H^1(\Omega)$$

and, for functions a and u such that $a \nabla u \in L^2(\Omega)^2$, define the functional $\ell_a^N(u)$ by

$$\langle \ell_a^N(u), v \rangle_* = (a \nabla u, \nabla v)_{L^2(\Omega)} \quad \forall v \in H^1(\Omega).$$

Clearly $\ell_a^N(u)$ and ℓ^N are linear. Further, it follows then from the Cauchy-Schwarz and trace inequalities that

$$|\langle \ell^N, v \rangle_*| \leq (\|f_N\|_{L^2(\Omega)} + \|g_N\|_{\partial\Omega}) \|v\|_{H^1(\Omega)} \quad \forall v \in H^1(\Omega),$$

and

$$|\langle \ell_a^N(u), v \rangle_*| \leq \|a \nabla u\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)} \quad \forall v \in H^1(\Omega),$$

and so $\ell_a^N(u), \ell^N \in H^1(\Omega)^*$. If we set $a = a_N$ and $u = u_N$, then from (3.12) we see that

$$\langle \ell_{a_N}^N(u_N), v \rangle_* = \langle \ell^N, v \rangle_* \quad \forall v \in H^1(\Omega),$$

and hence $\|\ell_{a_N}^N(u_N) - \ell^N\|_{H^1(\Omega)^*} = 0$. This motivates us to find an a such that the residual $\|\ell_a^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2$ is minimum. However, since the inverse problem is known to be ill-posed, we instead consider the regularized problem

$$\min_{a \in H^1(\Omega)} J_N(a) = \frac{1}{2} \|\ell_a^N(z) - \ell^N\|_{H^1(\Omega)^*}^2 + \frac{\beta}{2} \|a\|_{H^1(\Omega)}^2, \quad (4.1)$$

where $\beta > 0$ is a regularization parameter.

We proceed in a similar fashion for the other two inverse problems: let us define the operators ℓ^D and ℓ^E by

$$\begin{aligned}\langle \ell^D, v \rangle_{-1} &= (f_D, v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega), \\ \langle \ell^E, v \rangle_* &= (f_E, v)_{L^2(\Omega)} + \langle g_E, v \rangle_{\partial\Omega} \quad \forall v \in H^1(\Omega)^2,\end{aligned}$$

and notice that $\ell^D \in H^{-1}(\Omega)$, and that $\ell^E \in (H^1(\Omega)^2)^*$. Further, define the operators $\ell_a^D(u)$ and $\ell_m^E(w)$ by

$$\begin{aligned}\langle \ell_a^D(u), v \rangle_{-1} &= (a \nabla u, \nabla v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega), \\ \langle \ell_m^E(w), v \rangle_* &= (\sigma(m, w), \epsilon(v))_{L^2(\Omega)} \quad \forall v \in H^1(\Omega)^2,\end{aligned}$$

for sufficiently smooth functions a , u , m , and w , so that $\ell_a^D(u) \in H^{-1}(\Omega)$ and $\ell_m^E(w) \in (H^1(\Omega)^2)^*$. From (3.17) we see that

$$\langle \ell_{a_D}^D(u_D), v \rangle_{-1} = \langle \ell^D, v \rangle_{-1} \quad \forall v \in H_0^1(\Omega),$$

and from (3.19)

$$\langle \ell_{m_E}^E(u_E), v \rangle_* = \langle \ell^E, v \rangle_* \quad \forall v \in H^1(\Omega)^2,$$

and so $\|\ell_{m_D}^D(u_D) - \ell^D\|_{H^{-1}(\Omega)} = 0$, and $\|\ell_{m_E}^E(u_E) - \ell^E\|_{(H^1(\Omega)^2)^*} = 0$. Therefore, the equation error approach corresponding to the inverse problems related to the BVP (1.2) and (1.3) can be posed as

$$\min_{a \in H^1(\Omega)} J_D(a) = \frac{1}{2} \|\ell_a^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \frac{\alpha}{2} \|a\|_{H^1(\Omega)}^2, \quad (4.2)$$

and

$$\min_{m \in H^1(\Omega)^2} J_E(m) = \frac{1}{2} \|\ell_m^E(z_E) - \ell^E\|_{(H^1(\Omega)^2)^*}^2 + \frac{\gamma}{2} \|m\|_{H^1(\Omega)}^2, \quad (4.3)$$

respectively, where $\alpha, \gamma > 0$ are regularization parameters.

4.2 Existence and uniqueness analysis

In this section, we prove existence and uniqueness results concerning the optimization problems (4.1), (4.2), and (4.3).

First we shall need the following auxiliary results which we cite here. For all functions a and u such that $a\nabla u \in L^2(\Omega)^2$ we have

$$\|\ell_a^N(u)\|_{H^1(\Omega)^*} = \sup_{u \in H^1(\Omega)} \frac{(a\nabla u, \nabla v)_{L^2(\Omega)}}{\|v\|_{H^1(\Omega)}} \leq \|a\nabla u\|_{L^2(\Omega)},$$

from which we conclude that¹

$$\|\ell_a^N(u)\|_{H^1(\Omega)^*} \leq \|a\|_{L^2(\Omega)} \|\nabla u\|_{L^\infty(\Omega)} \quad \forall a \in L^2(\Omega), u \in W^{1,\infty}(\Omega) \quad (4.4)$$

$$\|\ell_a^D(u)\|_{H^{-1}(\Omega)} \leq \|a\|_{L^2(\Omega)} \|\nabla u\|_{L^\infty(\Omega)} \quad \forall a \in L^2(\Omega), u \in W^{1,\infty}(\Omega), \quad (4.5)$$

and

$$\|\ell_a^N(u)\|_{H^1(\Omega)^*} \leq \|a\|_{L^\infty(\Omega)} \|\nabla u\|_{L^2(\Omega)} \quad \forall a \in L^\infty(\Omega), u \in H^1(\Omega) \quad (4.6)$$

$$\|\ell_a^D(u)\|_{H^{-1}(\Omega)} \leq \|a\|_{L^\infty(\Omega)} \|\nabla u\|_{L^2(\Omega)} \quad \forall a \in L^\infty(\Omega), u \in H^1(\Omega). \quad (4.7)$$

Further, in view of (3.21) and (3.22), we have

$$\|\ell_m^E(u)\|_{(H^1(\Omega)^2)^*} \leq c_1 \|m\|_{L^2(\Omega)} \|u\|_{W^{1,\infty}(\Omega)}, \quad \forall m \in L^2(\Omega)^2, u \in W^{1,\infty}(\Omega)^2, \quad (4.8)$$

$$\|\ell_m^E(u)\|_{(H^1(\Omega)^2)^*} \leq c_2 \|m\|_{L^\infty(\Omega)} \|u\|_{H^1(\Omega)}, \quad \forall m \in L^\infty(\Omega)^2, u \in H^1(\Omega)^2, \quad (4.9)$$

where $c_1, c_2 > 0$ and independent of m and u .

We shall also need the following theorems.

Theorem 4.2.1. *Suppose that $\{a_n\} \subset L^2(\Omega)$ with $a_n \rightarrow a$ in $L^2(\Omega)$. Then for any $u \in W^{1,\infty}(\Omega)$ we have*

$$\ell_{a_n}^N(u) \rightarrow \ell_a^N(u), \quad (4.10)$$

$$\ell_{a_n}^D(u) \rightarrow \ell_a^D(u). \quad (4.11)$$

Further, if $\{m_n\} \subset L^2(\Omega)^2$ with $m_n \rightarrow m$ in $L^2(\Omega)^2$ and $u \in W^{1,\infty}(\Omega)^2$, then

$$\ell_{m_n}^E(u) \rightarrow \ell_m^E(u). \quad (4.12)$$

¹Since $H_0^1(\Omega) \subset H^1(\Omega)$ we also have $\|\ell_a^D(u)\|_{H^{-1}(\Omega)} \leq \|\ell_a^N(u)\|_{H^1(\Omega)^*}$.

Proof. Notice that $\ell_{(\cdot)}^N(w)$ is linear, and so by (4.4) we have

$$\|\ell_{a_n}^N(u) - \ell_a^N(u)\|_{H^1(\Omega)^*} = \|\ell_{(a_n-a)}^N(u)\|_{H^1(\Omega)^*} \leq \|a_n - a\|_{L^2(\Omega)} \|\nabla u\|_{W^{1,\infty}(\Omega)},$$

which proves (4.10). With a similar argument, we conclude (4.11) and (4.12). \square

Theorem 4.2.2. *Assume $z_N, z_D \in W^{1,\infty}(\Omega)$ and that $z_E \in W^{1,\infty}(\Omega)^2$. Then the functionals $J_N(\cdot)$, $J_D(\cdot)$, and $J_E(\cdot)$ are strictly convex and weakly lower semicontinuous.*

Proof. We prove the theorem for J_N , the other two functionals can be treated in the same way.

First, we will show that $J_N(\cdot)$ is strictly convex. In view of Theorem 3.1.6, it suffices to show that the functional $J_c(a) = \|\ell_a^N(u) - \ell^N\|_{H^1(\Omega)^*}^2$ is convex. For all $\lambda \in [0, 1]$ and $a, b \in H^1(\Omega)$ we have

$$\begin{aligned} J_c(\lambda a + (1 - \lambda)b) &= \|\lambda \ell_a^N(z_N) + (1 - \lambda) \ell_b^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 \\ &= \|\lambda(\ell_a^N(z_N) - \ell^N) + (1 - \lambda)(\ell_b^N(z_N) - \ell^N)\|_{H^1(\Omega)^*}^2 \\ &\leq (\lambda \|\ell_a^N(z_N) - \ell^N\|_{H^1(\Omega)^*} + (1 - \lambda) \|\ell_b^N(z_N) - \ell^N\|_{H^1(\Omega)^*})^2 \\ &\leq \lambda J_c(a) + (1 - \lambda) J_c(b) \quad (\text{since } |\cdot|^2 \text{ is convex}) \end{aligned}$$

showing J_c is convex, and thus, J_N is strictly convex.

Next we show J_N is weakly lower semicontinuous. Let $a \in H^1(\Omega)$, and let $\{a_n\}$ be any sequence in $H^1(\Omega)$ with $a_n \rightharpoonup a$ in $H^1(\Omega)$. As a consequence of Corollary 3.2.1, it follows that $a_n \rightarrow a$ in $L^2(\Omega)$. Therefore, in view of (4.10) and the fact that $\|\cdot\|$ is weakly lower semicontinuous, we have

$$\begin{aligned} J_N(a) &= \frac{1}{2} \|\ell_a^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \frac{\beta}{2} \|a\|_{H^1(\Omega)}^2 \\ &\leq \varliminf_{n \rightarrow \infty} \frac{1}{2} \|\ell_{a_n}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \varliminf_{n \rightarrow \infty} \frac{\beta}{2} \|a_n\|_{H^1(\Omega)}^2 \\ &\leq \varliminf_{n \rightarrow \infty} \left(\frac{1}{2} \|\ell_{a_n}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \frac{\beta}{2} \|a_n\|_{H^1(\Omega)}^2 \right) \\ &= \varliminf_{n \rightarrow \infty} J_N(a_n). \end{aligned}$$

Thus J_N is weakly lower semicontinuous. \square

Now we are ready to prove the existence and uniqueness results for the optimization problems given in (4.1), (4.2), and (4.3).

Theorem 4.2.3. (a) Assuming $z_N \in W^{1,\infty}(\Omega)$, then the optimization problem (4.1) has a unique solution $a_\beta \in H^1(\Omega)$.

(b) Assuming $z_D \in W^{1,\infty}(\Omega)$, then the optimization problem (4.2) has a unique solution $a_\alpha \in H^1(\Omega)$.

(c) Assuming $z_E \in W^{1,\infty}(\Omega)^2$, then the optimization problem (4.3) has a unique solution $m_\gamma \in H^1(\Omega)^2$.

Proof. The proof of (a) can be found in [8, Theorem 3.1]. We prove (b); part (c) can be proved by a similar argument.

Let

$$\varepsilon = \inf_{a \in H^1(\Omega)} \|\ell_a^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \frac{\alpha}{2} \|a\|_{H^1(\Omega)}^2.$$

Clearly $\varepsilon > 0$. By the definition of infimum, there exists a sequence $\{a_n\} \subset H^1(\Omega)$ (usually called a *minimizing sequence*) such that

$$J_D(a_n) \rightarrow \varepsilon.$$

Since $\|a_n\|_{H^1(\Omega)}^2 \leq 2\alpha^{-1} J_D(a_n)$, $\{a_n\}$ is bounded. From Corollary 3.2.1, there exists a subsequence of $\{a_n\}$, which we still denote by $\{a_n\}$, and an $\tilde{a} \in H^1(\Omega)$ such that $a_n \rightharpoonup \tilde{a}$ in $H^1(\Omega)$ and $a_n \rightarrow \tilde{a}$ in $L^2(\Omega)$. Since by assumption $z_D \in W^{1,\infty}(\Omega)$, it follows from Theorem 4.2.2 that J_D is weakly lower semicontinuous, and so

$$J_D(\tilde{a}) \leq \liminf_{n \rightarrow \infty} J_D(a_n) = \varepsilon.$$

Therefore, $J_D(\tilde{a}) = \varepsilon$, showing \tilde{a} is a solution of the optimization problem (4.1).

Next we show \tilde{a} is unique. Assume, by the way of contradiction, that $\bar{a} \neq \tilde{a}$ is another minimizer of J_D . But then $a = (\bar{a} + \tilde{a})/2 \in H^1(\Omega)$, and so the strict convexity of J_D implies

$$J_D(a) < \frac{1}{2} (J_D(\bar{a}) + J_D(\tilde{a})) = \varepsilon,$$

which contradicts the assumption that \tilde{a} is a minimizer. Thus \tilde{a} is unique, and we re-label \tilde{a} by a_α . \square

4.3 The discretized problems

In the actual computations we compute the minimizer from a finite-dimensional subspace of $H^1(\Omega)$; when analyzing the results, we will assume that Ω has a polygonal boundary. For this purpose, we shall recall a few definitions from finite element theory. Let $\{T_h\}$ be a family of shape regular triangulations of $\bar{\Omega}$, where h denotes the maximum diameter for any triangle in T_h . This means there exists a number $\kappa > 0$ such that, for all $h > 0$, every T in T_h contains a circle of radius ρ_T with

$$\rho_T \geq \frac{h_T}{\kappa},$$

where h_T is diameter of T . Define

$$V_h^k = \{v \in C^0(\bar{\Omega}) \mid v|_T \in P_k \ \forall T \in T_h\},$$

where P_k is the space of polynomials of degree at most k . It is known [38] that V_h^k is a finite-dimensional subspace of $H^1(\Omega)$, and hence it is convex and closed. From now on, let $\{\varphi_1, \dots, \varphi_n\}$ be a basis for V_h^k .

Since J_N , J_D , and J_N are weakly lower semicontinuous, strictly convex, and coercive, it follows from Theorem 3.1.7 that the optimization problems (4.1), (4.2), and (4.3) admit unique solutions over V_h^k . Hence, in what follows let $a_{h,\beta}$, $a_{h,\alpha}$, and $m_{h,\gamma}$ denote the (unique) solutions of the optimization problems

$$\min_{a \in V_h^k} \frac{1}{2} \|\ell_a^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \frac{\beta}{2} \|a\|_{H^1(\Omega)}^2, \quad (4.13)$$

$$\min_{a \in V_h^k} \frac{1}{2} \|\ell_a^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \frac{\alpha}{2} \|a\|_{H^1(\Omega)}^2, \quad (4.14)$$

and

$$\min_{m \in V_h^k \times V_h^k} \frac{1}{2} \|\ell_m^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 + \frac{\gamma}{2} \|m\|_{H^1(\Omega)}^2, \quad (4.15)$$

respectively.

In Chapter 5, we will develop stability results and error estimates for both the continuous problems (4.1), (4.2), and (4.3), and the discretized problems (4.13), (4.14), and (4.15).

4.4 Implementation

In this section we will explain how one can implement the equation error method in a practical algorithm. We shall use (4.1) as a model problem, problems (4.2) and (4.3) can be treated in the same fashion.

In practice, the optimization problem posed in (4.1) is replaced by the finite dimensional optimization problem (4.13), which in view of Theorem 3.1 can be written as

$$\min_{a \in V_h} \frac{1}{2} \langle \ell_a^N(z_N) - \ell^N, P^{-1}(\ell_a^N(z_N) - \ell^N) \rangle + \frac{\beta}{2} \|a\|_{H^1(\Omega)}^2,$$

where V_h is some finite-dimensional subspace of $H^1(\Omega)$, and the mapping $P : H^1(\Omega) \rightarrow H^1(\Omega)^*$ is given by

$$\langle Pu, v \rangle = \int_{\Omega} \{uv + \nabla u \cdot \nabla v\} \quad \forall v \in H^1(\Omega).$$

Computing the objective function above is problematic because we cannot compute the action of P^{-1} , that is, we cannot solve $Pu = \phi$ for u given $\phi \in H^1(\Omega)^*$. We thus must discretize the operator P as well. In place of $P : H^1(\Omega) \rightarrow H^1(\Omega)^*$, we will use $P_h : V_h \rightarrow V_h^*$. The definition of P_h is the same as that of P :

$$\langle P_h u, v \rangle = \int_{\Omega} \{uv + \nabla u \cdot \nabla v\} \quad \forall v \in V_h.$$

Note that elements of V_h^* are simply elements of $H^1(\Omega)^*$ restricted to V_h . That is, given $\phi \in H^1(\Omega)^*$, we can define $\phi_h \in V_h^*$ by

$$\langle \phi_h, v \rangle = \langle \phi, v \rangle \quad \forall v \in V_h.$$

We write $\langle \cdot, \cdot \rangle$ for the pairing between V_h and V_h^* as well as for the pairing between $H^1(\Omega)$ and $H^1(\Omega)^*$. Any $\phi \in H^1(\Omega)^*$ defines, by restriction, an element of V_h^* , which we also denote by ϕ (thus we don't distinguish between $\ell_a^N(z_N)$ and its restriction

to V_h , and similarly for ℓ^N).

We now pose the discretized optimization problem as

$$\min_{a \in V_h} J_h(a) = \frac{1}{2} \langle \ell_a^N(z_N) - \ell^N, P_h^{-1}(\ell_a^N(z_N) - \ell^N) \rangle + \frac{\beta}{2} \|a\|_{H^1(\Omega)}^2.$$

To compute the discretized objective function above, we must compute

$$u = P_h^{-1}(\ell_a^N(z_N) - \ell^N).$$

By definition, $u \in V_h$ satisfies

$$\int_{\Omega} \{u\varphi_i + \nabla u \cdot \nabla \varphi_i\} = \int_{\Omega} a \nabla z \cdot \nabla \varphi_i - \int_{\Omega} f \varphi_i - \int_{\partial\Omega} g \varphi_i \quad \forall i = 1, 2, \dots, n, \quad (4.16)$$

where $\{\varphi_i\}$ is a basis for V_h . Now write a and u as²

$$a = \sum_{j=1}^n A_j \varphi_j, \quad u = \sum_{j=1}^n U_j \varphi_j,$$

then (4.16) reduces to a system of equations $PU = KA - F$, where the matrices $P, K \in \mathbb{R}^{n \times n}$ and the vector $F \in \mathbb{R}^n$ are given by

$$(P)_{i,j} = \int_{\Omega} \{\varphi_j \varphi_i + \nabla \varphi_j \cdot \nabla \varphi_i\}$$

$$(K)_{i,j} = \int_{\Omega} \varphi_j \nabla z \cdot \nabla \varphi_i$$

$$(F)_i = \int_{\Omega} f \varphi_i + \int_{\partial\Omega} g \varphi_i.$$

Finally, since $u = P_h^{-1}(\ell_a(z) - \ell)$ and $U = P^{-1}(KA - F)$, we have

$$\begin{aligned} J_h(a) &= \frac{1}{2} \langle P_h u, u \rangle + \frac{\beta}{2} \langle P_h a, a \rangle \\ &= \frac{1}{2} U^T P U + \frac{1}{2} A^T P A \\ &= \frac{1}{2} (KA - F)^T P^{-1} (KA - F) + \frac{\beta}{2} A^T P A. \end{aligned}$$

²For the Dirichlet problem, one uses different bases to represent a and u .

Since J_h is convex, the optimality condition then implies

$$\nabla J_h(a) = K^T P^{-1}(KA - F) + \beta PA = 0,$$

and hence, A is obtained by solving the linear system of equations

$$(K^T P^{-1} K + \beta P)A = K^T P^{-1} F.$$

Numerical examples will be presented in Chapter 5 and 6.

Chapter 5

Stability and error estimates

In this chapter we prove stability and convergence results regarding the equation error approach. Numerical examples are also presented.

5.1 The Neumann BVP

For convenience, let S_N be the set of all $a \in H^1(\Omega)$ satisfying (3.12), that is

$$S_N = \{a \in H^1(\Omega) \mid \ell_a^N(u_N) = \ell^N \text{ in } H^1(\Omega)^*\}.$$

We shall make the assumption that S_N is nonempty.

The results of this section are proved in [42]. We start with the following result regarding the stability of the equation error method applied to BVP (1.1).

Theorem 5.1.1. *Suppose that $u_N \in W^{1,\infty}(\Omega)$. Let $\{z_n\} \subset W^{1,\infty}(\Omega)$ be a sequence of observations of u_N , and let $\{\epsilon_n\}$ and $\{\beta_n\}$ be two sequences of real numbers such that*

1. $\epsilon_n^2 \leq \beta_n \leq \epsilon_n \ \forall n \in \mathbb{N}$, and $\epsilon_n^2/\beta_n \rightarrow 0$ as $n \rightarrow \infty$,
2. $\|u_N - z_n\|_{W^{1,\infty}(\Omega)} \leq \epsilon_n \ \forall n \in \mathbb{N}$,
3. $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

For each $n \in \mathbb{N}$, let $a_n \in H^1(\Omega)$ be the unique solution (guaranteed by Theorem 4.2.3) of the optimization problem

$$\min_{b \in H^1(\Omega)} \frac{1}{2} \|\ell_b^N(z_n) - \ell^N\|_{H^1(\Omega)^*}^2 + \frac{\beta_n}{2} \|b\|_{H^1(\Omega)}^2.$$

Then there is an $\tilde{a} \in S_N$ such that $\|a_n - \tilde{a}\|_{H^1(\Omega)} \rightarrow 0$. Further, $\|\tilde{a}\|_{H^1(\Omega)} \leq \|a^*\|_{H^1(\Omega)}$ for all $a^* \in S_N$.

Proof. Using (4.4) and the fact that $\ell_{a^*}^N(u_N) = \ell^N$ in $H^1(\Omega)^*$ for all $a^* \in S_N$, we have

$$\begin{aligned}
\beta_n \|a_n\|_{H^1(\Omega)}^2 &\leq \|\ell_{a^*}^N(z_n) - \ell^N\|_{H^1(\Omega)^*}^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2 \\
&= \|\ell_{a^*}^N(z_n - u_N)\|_{H^1(\Omega)^*}^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2 \\
&\leq \|a^*\|_{L^2(\Omega)}^2 \|\nabla(z_n - u_N)\|_{L^\infty(\Omega)}^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2 \\
&\leq \|a^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2 \\
&\leq \beta_n \|a^*\|_{L^2(\Omega)}^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2,
\end{aligned} \tag{5.1}$$

and consequently

$$\|a_n\|_{H^1(\Omega)}^2 \leq \|a^*\|_{L^2(\Omega)}^2 + \|a^*\|_{H^1(\Omega)}^2 \quad \forall n \in \mathbb{N}, a^* \in S_N.$$

Therefore $\{a_n\}$ is bounded in $H^1(\Omega)$, and by Corollary 3.2.1, there exists a subsequence of $\{a_n\}$, which we still denote by $\{a_n\}$, and a vector $\tilde{a} \in H^1(\Omega)$ such that $a_n \rightharpoonup \tilde{a}$ in $H^1(\Omega)$ and $a_n \rightarrow \tilde{a}$ in $L^2(\Omega)$. Now from (4.4) we have

$$\begin{aligned}
\|\ell_{a_n}^N(u_N) - \ell_{a_n}^N(z_n)\|_{H^1(\Omega)^*}^2 &= \|\ell_{a_n}^N(u_N - z_n)\|_{H^1(\Omega)^*}^2 \\
&\leq \|a_n\|_{L^2(\Omega)}^2 \|\nabla(u_N - z_n)\|_{L^\infty(\Omega)}^2 \\
&\leq \|a_n\|_{L^2(\Omega)}^2 \epsilon_n^2,
\end{aligned} \tag{5.2}$$

and by the definition of a_n we also have

$$\begin{aligned}
\|\ell_{a_n}^N(z_n) - \ell^N\|_{H^1(\Omega)^*}^2 &\leq \|\ell_{a^*}^N(z_n) - \ell^N\|_{H^1(\Omega)^*}^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2 \\
&= \|\ell_{a^*}^N(z_n - u_N)\|_{H^1(\Omega)^*}^2 + \beta_n \|a^*\|_{H^1(\Omega)}^2 \\
&\leq \|a^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \epsilon_n \|a^*\|_{H^1(\Omega)}^2.
\end{aligned} \tag{5.3}$$

Combine (5.2), (5.3) and the fact that a convergent sequence is necessarily bounded to see that

$$\begin{aligned}
\|\ell_{a_n}^N(u_N) - \ell^N\|_{H^1(\Omega)^*}^2 &\leq 2\|\ell_{a_n}^N(u_N) - \ell_{a_n}^N(z_n)\|_{H^1(\Omega)^*}^2 + 2\|\ell_{a_n}^N(z_n) - \ell^N\|_{H^1(\Omega)^*}^2 \\
&\leq 2\|a_n\|_{L^2(\Omega)}^2 \epsilon_n^2 + 2\left(\|a^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \epsilon_n \|a^*\|_{H^1(\Omega)}^2\right) \\
&\leq C(\epsilon_n^2 + \epsilon_n).
\end{aligned}$$

Consequently, in view of (4.10), we conclude that

$$\|\ell_{\tilde{a}}^N(u_N) - \ell^N\|_{H^1(\Omega)^*} = \lim_{n \rightarrow \infty} \|\ell_{a_n}^N(u_N) - \ell^N\|_{H^1(\Omega)^*} = 0,$$

showing $\tilde{a} \in S_N$. Since $a_n \rightharpoonup \tilde{a}$ in $H^1(\Omega)$, by Theorem 3.1.5 we have

$$\|\tilde{a}\|_{H^1(\Omega)} \leq \varliminf_{n \rightarrow \infty} \|a_n\|_{H^1(\Omega)},$$

and by the fourth inequality in (5.1) and the fact that $\tilde{a} \in S_N$ we have

$$\overline{\lim}_{n \rightarrow \infty} \|a_n\|_{H^1(\Omega)}^2 \leq \lim_{n \rightarrow \infty} \left(\frac{\epsilon_n^2}{\beta_n} \|\tilde{a}\|_{L^2(\Omega)}^2 + \|\tilde{a}\|_{H^1(\Omega)}^2 \right) = \|\tilde{a}\|_{H^1(\Omega)}^2.$$

Hence $\|a_n\|_{H^1(\Omega)} \rightarrow \|\tilde{a}\|_{H^1(\Omega)}$, and by a standard functional analysis result¹, we conclude that $a_n \rightarrow \tilde{a}$ in $H^1(\Omega)$. Further, in view of (5.1), we conclude

$$\|\tilde{a}\|_{H^1(\Omega)}^2 = \lim_{n \rightarrow \infty} \|a_n\|_{H^1(\Omega)}^2 \leq \lim_{n \rightarrow \infty} \left(\frac{\epsilon_n^2}{\beta_n} \|a^*\|_{L^2(\Omega)}^2 + \|a^*\|_{H^1(\Omega)}^2 \right) = \|a^*\|_{H^1(\Omega)}^2,$$

for all $a^* \in S_N$, which completes the proof. \square

Remark 5.1.1. *Acar [17] has also proved a convergence result for the equation error method applied to (1.1). However, he assumes that a_N is found in $H^2(\Omega)$ and his regularization term uses the H^2 norm, instead of the H^1 norm, as in our analysis. He assumes that the error in the data goes to zero in H^1 , and shows that the error in the estimated parameter goes to zero in L^∞ ; in this regard, his result cannot be directly compared to ours.*

We believe that few workers in this field are interested in using H^2 regularization because it forces too much smoothness on the solution; the trend is towards H^1 or even BV regularization. Our result above succeeds in proving the convergence of the equation error method under H^1 regularization, and ours is (we believe) the only such proof.

Remark 5.1.2. *There are several results in the literature concerning the uniqueness of the unknown coefficient a_N in the BVP (1.1). For instance, under the non-degeneracy condition (2.10), Falk [24] proved that there is at most one coefficient*

¹Precisely: in an inner product space, $v_n \rightarrow v$ if and only if $v_n \rightharpoonup v$ and $\|v_n\| \rightarrow \|v\|$. See for example [28, 30].

$a_N \in H^1(\Omega)$ satisfying (1.1). Richter [25] also proved the uniqueness but under the more general and less restrictive nondegeneracy condition (2.18). See Section 2.2 for more details.

In our derivations below, we will not assume any nondegeneracy condition comparable to (2.10) and (2.18). Rather, we will include ∇z_N in the expression for the error, which implies that the error estimate provides no information about the error in any region in which ∇z_N is zero. In what follows we let a_N be any function satisfying (3.12).

To derive an upper bound on the error, we measure the error in a certain quotient norm. To this end, we need the following result (see [35, Theorem 3.2] for the proof).

Theorem 5.1.2. *Assume Ω is a bounded, simply-connected domain in \mathbb{R}^2 with Lipschitz-continuous boundary $\partial\Omega$. Every function \vec{v} of $L^2(\Omega)^2$ has the following orthogonal decomposition:*

$$\vec{v} = \nabla q + \text{curl} \phi,$$

where $q \in V$ is the only solution of

$$(\nabla q, \nabla \mu)_{L^2(\Omega)} = (\vec{v}, \nabla \mu)_{L^2(\Omega)} \quad \forall \mu \in H^1(\Omega), \quad (5.4)$$

and $\phi \in H_0^1(\Omega)$ is the only solution of

$$(\text{curl} \phi, \text{curl} \chi)_{L^2(\Omega)} = (\vec{v}, \text{curl} \chi)_{L^2(\Omega)} \quad \forall \chi \in H_0^1(\Omega).$$

Recall that:

$$V = \left\{ v \in H^1(\Omega) \mid \int_{\Omega} v = 0 \right\}.$$

We also need the following definition and results concerning quotient spaces, see for example [28, 31].

Definition 5.1.1. *Let M be a subspace of a vector space X . For each $x \in X$, let $\dot{x} = x + M = \{x + m : m \in M\}$, and define the set X/M by*

$$X/M = \{\dot{x} : x \in X\}.$$

Theorem 5.1.3. *Let M and X be defined as in Definition 5.1.1.*

(a) The set X/M is a vector space with the vector space operations

$$(x + M) + (y + M) = (x + y) + M$$

and

$$\alpha(x + M) = (\alpha x) + M.$$

Equipped with these two operations, X/M is called a **quotient space**.

(b) If X is a normed space and M is a closed subspace of X , then the function $\|\cdot\|_{X/M}$ given by

$$\|\dot{x}\|_{X/M} = \inf_{m \in M} \|x - m\|_X$$

is a norm on X/M called the **quotient norm**. If, in addition, X is a Banach space, then X/M is also a Banach space.

Now we can define the spaces and norms we use to give the stability results for the equation error approach. To this end, observe that the space

$$R = \ker(\text{rot}) = \{\vec{v} \in L^2(\Omega)^2 : \text{rot } \vec{v} = 0\}$$

is a closed subspace of $L^2(\Omega)^2$ since it is the kernel of the bounded linear operator $\text{rot} : L^2(\Omega)^2 \rightarrow H^{-1}(\Omega)$, and hence we have the orthogonal decomposition

$$L^2(\Omega)^2 = R \oplus R^\perp.$$

Further, in view of Theorem 5.1.3, the following quotient space and norm are well defined:

$$Q = L^2(\Omega)^2 / R^\perp = \{\dot{v} = \vec{v} + R^\perp : \vec{v} \in L^2(\Omega)^2\},$$

$$\|\dot{v}\|_Q = \inf_{\vec{p} \in R^\perp} \|\vec{v} - \vec{p}\|_{L^2(\Omega)^2}.$$

Note that by the projection theorem, we have the alternative formula

$$\|\dot{v}\|_Q = \|\vec{v} - \text{proj}_{R^\perp} \vec{v}\|_{L^2(\Omega)^2}.$$

Finally, before stating the first error estimate, we shall also need the following theorem.

Theorem 5.1.4. *If $q \in H^1(\Omega)$, then $\nabla q \in \ker(\text{rot})$, i.e., $\text{rot}\nabla q = 0$ in $H^{-1}(\Omega)$.*

Proof. By Green's formula (3.4), for any $\phi \in C_0^\infty(\Omega)$ we have

$$\begin{aligned} \langle \text{rot}\nabla q, \phi \rangle_{-1} &= \int_{\Omega} \nabla q \cdot \text{curl } \phi \\ &= \int_{\Omega} \left\{ \frac{\partial q}{\partial x} \frac{\partial \phi}{\partial y} - \frac{\partial q}{\partial y} \frac{\partial \phi}{\partial x} \right\} \\ &= \int_{\Omega} \left\{ q \frac{\partial^2 \phi}{\partial x \partial y} - q \frac{\partial^2 \phi}{\partial y \partial x} \right\} \\ &= 0. \end{aligned}$$

We finish the proof by a density argument. Let $v \in H_0^1(\Omega)$. Since $C_0^\infty(\Omega)$ is dense in $H_0^1(\Omega)$, there exists a sequence $\{\psi_n\} \subset C_0^\infty(\Omega)$ such that $\psi_n \rightarrow v$ in the H^1 norm. But since $\text{rot}\nabla q \in H^{-1}(\Omega)$, it is continuous, and so

$$\langle \text{rot}\nabla q, v \rangle_{-1} = \lim_{n \rightarrow \infty} \langle \text{rot}\nabla q, \psi_n \rangle_{-1} = 0.$$

Since $v \in H_0^1(\Omega)$ was arbitrary, the result follows. \square

Now we are ready to state our first error estimate for the equation error approach.

Theorem 5.1.5. *Assume that $a_N \in H^1(\Omega)$ and $u_N, z_N \in W^{1,\infty}(\Omega)$, and let a_β be the solution of the optimization problem (4.1). If $\|u_N - z_N\|_{W^{1,\infty}(\Omega)} \leq \epsilon$, then for the error vector $\vec{e} = (a_\beta - a_N)\nabla z_N$ we have*

$$\|\dot{e}\|_Q \leq C(\epsilon + \sqrt{\beta}),$$

where C depends on $\|a_N\|_{H^1(\Omega)}$ but is independent of β and ϵ .

Proof. Obviously $\vec{e} \in L^2(\Omega)^2$ and so by Theorem 5.1.2 we have the decomposition

$$\vec{e} = \nabla q + \text{curl } \phi.$$

By Theorem 5.1.4, $\nabla q \in R$, and so $\text{proj}_{R^\perp} \vec{e} = \text{curl } \phi$. Consequently it follows that

$$\|\dot{e}\|_Q = \|\vec{e} - \text{curl } \phi\|_{L^2(\Omega)} = \|\nabla q\|_{L^2(\Omega)}.$$

Now by (5.4) and the fact that $\ell_{a_N}^N(u_N) = \ell^N$, we see that

$$\begin{aligned}
(\nabla q, \nabla \mu)_{L^2(\Omega)} &= (\vec{e}, \nabla \mu)_{L^2(\Omega)} \\
&= (a_\beta \nabla z_N, \nabla \mu)_{L^2(\Omega)} - (a_N \nabla z_N, \nabla \mu)_{L^2(\Omega)} \\
&= (a_\beta \nabla z_N, \nabla \mu)_{L^2(\Omega)} - (a_N \nabla u_N, \nabla \mu)_{L^2(\Omega)} \\
&\quad + (a_N \nabla u_N, \nabla \mu)_{L^2(\Omega)} - (a_N \nabla z_N, \nabla \mu)_{L^2(\Omega)} \\
&= \left\langle \ell_{a_\beta}^N(z_N) - \ell^N, \mu \right\rangle_* + (a_N \nabla(u_N - z_N), \nabla \mu)_{L^2(\Omega)} \quad \forall \mu \in H^1(\Omega).
\end{aligned}$$

Since $q \in V$, by Corollary 3.3.1, we have $\|q\|_{H^1(\Omega)} \leq C_1 \|\nabla q\|_{L^2(\Omega)}$ for some constant C_1 independent of q . Thus, taking $\mu = q$ in the inequality above, we obtain

$$\begin{aligned}
\|\nabla q\|_{L^2(\Omega)}^2 &\leq \|\ell_{a_\beta}^N(z_N) - \ell^N\|_{H^1(\Omega)^*} \|q\|_{H^1(\Omega)} + \|a_N\|_{L^2(\Omega)} \|\nabla(u_N - z_N)\|_{L^\infty(\Omega)} \|\nabla q\|_{L^2(\Omega)} \\
&\leq C \left(\|\ell_{a_\beta}^N(z_N) - \ell^N\|_{H^1(\Omega)^*} \|\nabla q\|_{L^2(\Omega)} + \epsilon \|\nabla q\|_{L^2(\Omega)} \right).
\end{aligned}$$

Consequently, we have

$$\|\nabla q\|_{L^2(\Omega)}^2 \leq C \left(\|\ell_{a_\beta}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \epsilon^2 \right). \quad (5.5)$$

Using (4.4) and noting that $\ell_{a_N}^N(u_N) = \ell^N$ in $H^1(\Omega)^*$, we obtain

$$\|\ell_{a_N}^N(z_N) - \ell^N\|_{H^1(\Omega)^*} = \|\ell_{a_N}^N(z_N - u_N)\|_{H^1(\Omega)^*} \leq \epsilon \|a_N\|_{L^2(\Omega)}. \quad (5.6)$$

Finally, combine (5.5), (5.6) and the fact that a_β solves (4.1) to see that

$$\|\nabla q\|_{L^2(\Omega)}^2 \leq C \left(\|\ell_{a_N}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \beta \|a_N\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \leq C (\epsilon^2 + \beta),$$

which completes the proof. \square

Using (4.6) instead of (4.4) in the proof above, it is easy to prove the following result, which allows the error in the data to be measured in the weaker H^1 norm instead of the $W^{1,\infty}$ norm, and eliminates the requirement that $u \in W^{1,\infty}(\Omega)$.

Theorem 5.1.6. *Assume that $a_N \in L^\infty(\Omega) \cap H^1(\Omega)$ and $z_N \in W^{1,\infty}(\Omega)$, and let a_β be the solution of the optimization problem (4.1). If $\|u_N - z_N\|_{H^1(\Omega)} \leq \epsilon$, then for*

the error vector $\vec{e} = (a_\beta - a_N)\nabla z_N$ we have

$$\|\dot{e}\|_Q \leq C(\epsilon + \sqrt{\beta}),$$

where C depends on $\|a_N\|_{L^\infty(\Omega)}$ and $\|a_N\|_{H^1(\Omega)}$ but is independent of β and ϵ .

Remark 5.1.3. The same conclusion of Theorem 5.1.5 is valid if the error vector is defined as $\vec{e} = (a_\beta - a_N)\nabla u_N$ provided $\beta \geq C\epsilon^2$; the analysis changes little.

Now we give an error estimate for the finite element solution of the equation error method.

Theorem 5.1.7. Assume that $a_N \in H^r(\Omega)$ for some $1 \leq r \leq k+1$, and that $u_N, z_N \in W^{1,\infty}(\Omega)$. Let $a_{h,\beta}$ be the solution of the optimization problem (4.13). If $\|u_N - z_N\|_{W^{1,\infty}(\Omega)} \leq \epsilon$, then for the error vector $\vec{e}_h = (a_{h,\beta} - a_N)\nabla z_N$ we have

$$\|\dot{e}_h\|_Q \leq C \left(h^r + \epsilon + \sqrt{\beta} \right),$$

where C depends on $\|a_N\|_{H^r(\Omega)}$ and $\|z_N\|_{W^{1,\infty}(\Omega)}$ but is independent of β , ϵ , and h .

Proof. Let \tilde{a}_h be the H^1 -projection of a_N into V_h^k (which exists and is unique since V_h^k is a closed subspace of $H^1(\Omega)$), that is

$$(\tilde{a}_h, b_h)_{H^1(\Omega)} = (a_N, b_h)_{H^1(\Omega)} \quad \forall b_h \in V_h^k.$$

Notice that $\|\tilde{a}_h\|_{H^1(\Omega)}^2 \leq \|a_N\|_{H^1(\Omega)}^2$. Using the same method as in the proof of Theorem 5.1.5, we obtain

$$\begin{aligned} \|\dot{e}_h\|_Q^2 &\leq C \left(\|\ell_{a_{h,\beta}}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{\tilde{a}_h}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \beta \|\tilde{a}_h\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{(\tilde{a}_h - a_N)}^N(z_N)\|_{H^1(\Omega)^*}^2 + \|\ell_{a_N}^N(z_N) - \ell^N\|_{H^1(\Omega)^*}^2 + \beta \|a_N\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{(\tilde{a}_h - a_N)}^N(z_N)\|_{H^1(\Omega)^*}^2 + \epsilon^2 + \beta \right). \end{aligned}$$

We need to bound the first term on the right hand side of the above inequality. From [41] (Corollary 7.8 and its proof) one can conclude the inequality

$$\|\tilde{a}_h - a_N\|_{L^2(\Omega)} \leq Ch \|\tilde{a}_h - a_N\|_{H^1(\Omega)} \leq Ch \|a_N\|_{H^1(\Omega)}. \quad (5.7)$$

Thus, in the case that $r = 1$, it follows directly that

$$\|\tilde{a}_h - a_N\|_{L^2(\Omega)} \leq Ch \|a_N\|_{H^1(\Omega)}. \quad (5.8)$$

Now for $r \geq 2$, let a_I^{r-1} be the piecewise interpolant of degree $(r-1)$ of a_N from V_h^{r-1} . Then it follows from the definition of \tilde{a}_h and a standard approximation result [33, Theorem 4.4.20] that

$$\begin{aligned} \|\tilde{a}_h - a_N\|_{H^1(\Omega)} &\leq \inf_{b_h \in V_h^k} \|a_N - b_h\|_{H^1(\Omega)} \\ &\leq \inf_{b_h \in V_h^{r-1}} \|a_N - b_h\|_{H^1(\Omega)} \\ &\leq \|a_N - a_I^{r-1}\|_{H^1(\Omega)} \\ &\leq Ch^{r-1} \|a_N\|_{H^r(\Omega)}. \end{aligned}$$

Combining the last inequality with (5.7) and (5.8) we conclude that

$$\|\tilde{a}_h - a_N\|_{L^2(\Omega)} \leq Ch^r, \quad \text{for } 1 \leq r \leq k+1. \quad (5.9)$$

Finally, using (4.4) we have

$$\|\ell_{(\tilde{a}_h - a_N)}^N(z_N)\|_{H^1(\Omega)^*} \leq \|\nabla z_N\|_{L^\infty(\Omega)} \|\tilde{a}_h - a_N\|_{L^2(\Omega)} \leq Ch^r,$$

which completes the proof. \square

With a minor modification of the proof above, one can easily conclude the following result.

Theorem 5.1.8. *Assume that $a_N \in L^\infty(\Omega) \cap H^r(\Omega)$ for some $1 \leq r \leq k+1$, and that $z_N \in W^{1,\infty}(\Omega)$. Let $a_{h,\beta}$ be the solution of the optimization problem (4.13). If $\|u_N - z_N\|_{H^1(\Omega)} \leq \epsilon$, then for the error vector $\vec{e}_h = (a_{h,\beta} - a_N)\nabla z_N$ we have*

$$\|\vec{e}_h\|_Q \leq C \left(h^r + \epsilon + \sqrt{\beta} \right),$$

where C depends on $\|a_N\|_{L^\infty(\Omega)}$, $\|a_N\|_{H^r(\Omega)}$, and $\|z_N\|_{W^{1,\infty}(\Omega)}$ but is independent of β , ϵ , and h .

Remark 5.1.4. *If z_N is the exact piecewise quadratic interpolant of u_N , then from a*

standard approximation result we have $\|z_N - u_N\|_{L^2(\Omega)} = O(h^3)$ and $\|z_N - u_N\|_{H^1(\Omega)} = O(h^2)$. Consequently if $\beta \sim h^2$ and $r = 1$, we see that the error bounds given in (2.9), (2.17), and Theorem 5.1.8 are of $O(h)$. Notice however, if z_N is the exact piecewise linear interpolant of u_N , then $\|z_N - u_N\|_{L^2(\Omega)} = O(h^2)$ and $\|z_N - u_N\|_{H^1(\Omega)} = O(h)$, and so the error estimates given in (2.9) and (2.17) provide no information about the convergence rate in this case, but Theorem 5.1.8 still yields $O(h)$ convergence.

Remark 5.1.5. Kärkkäinen [8] introduced the use of the quotient norm $\|\cdot\|_Q$, and proved a result equivalent to Theorem 5.1.7 (the bound for the error in the discretized problem). He treated a slightly more general problem (allowing a lower-order term in the PDE); on the other hand, he did not prove any result comparable to Theorem 5.1.5, treating the original infinite-dimensional problem. Our analysis is considerably simpler than his, as we avoid integration by parts and the consequent (difficult) estimates of the error in the boundary terms. This allows us to extend the analysis to the other two inverse problems, which seems difficult or impossible using his approach.

5.1.1 Numerical examples

We consider the finite element solution using the space of continuous piecewise linear polynomials over the unit square $[0, 1]^2$ with a uniform triangulation similar to the one shown in Figure 5.1.

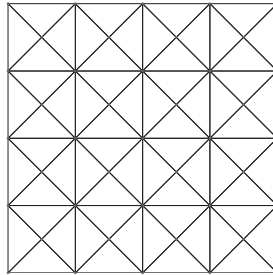


Figure 5.1: Triangulation of Ω .

At every node x_i in the triangulation, the synthetic data z_N is computed according to the formula

$$z_N(x_i) = u_N(x_i) + \hat{\delta}\eta_i, \quad (5.10)$$

where η_i is a uniformly distributed random number in $[-1, 1]$ and $\hat{\delta}$ determines the noise level. So when $\hat{\delta} = 0$, z_N is the interpolant of u_N from V_h^1 defined by the nodal values given in (5.10). In what follows, let δ denote the relative error in the data in the H^1 norm, that is,

$$\delta = \frac{\|z_N - u_N\|_{H^1(\Omega)}}{\|u_N\|_{H^1(\Omega)}}.$$

Example 5.1.1. *We consider the problem with $u_N(x, y) = 2 \cos(\pi x) \cos(\pi y)$, where f_N and g_N are chosen so that the exact parameter is*

$$a_N(x, y) = \exp[-10(x - 0.5)^2 - 10(y - 0.5)^2].$$

First we consider the case in which $\delta = 0$, so that z_N represents the interpolant of u_N . The computed errors are presented in Table 5.1. Next, we plot the exact and the recovered parameter with the presence of noise, results for various noise levels are shown in Figure 5.2.

Example 5.1.2. *We consider the problem with $u_N(x, y) = 8x^2 + e^y$, where f_N and g_N are chosen so that the exact parameter is*

$$a_N(x, y) = 2 + x + y.$$

First we consider the case in which $\delta = 0$. The computed errors are presented in Table 5.2. Next, we plot the exact and the recovered parameter with the presence of noise, results for various noise levels are shown in Figure 5.3.

Table 5.1

Results of Example 5.1.1 with $\delta = 0$ and regularization parameter $\beta = h^2$.

h	$\ a_{h,\beta} - a_N\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
1/4	$1.624 \cdot 10^{-1}$	$4.104 \cdot 10^{-1}$	$2.204 \cdot 10^{-1}$
1/8	$7.980 \cdot 10^{-2}$	$2.016 \cdot 10^{-1}$	$9.140 \cdot 10^{-2}$
1/16	$3.780 \cdot 10^{-2}$	$9.550 \cdot 10^{-2}$	$3.190 \cdot 10^{-2}$
1/32	$1.770 \cdot 10^{-2}$	$4.470 \cdot 10^{-2}$	$1.060 \cdot 10^{-2}$
1/64	$8.300 \cdot 10^{-3}$	$2.090 \cdot 10^{-2}$	$3.500 \cdot 10^{-3}$

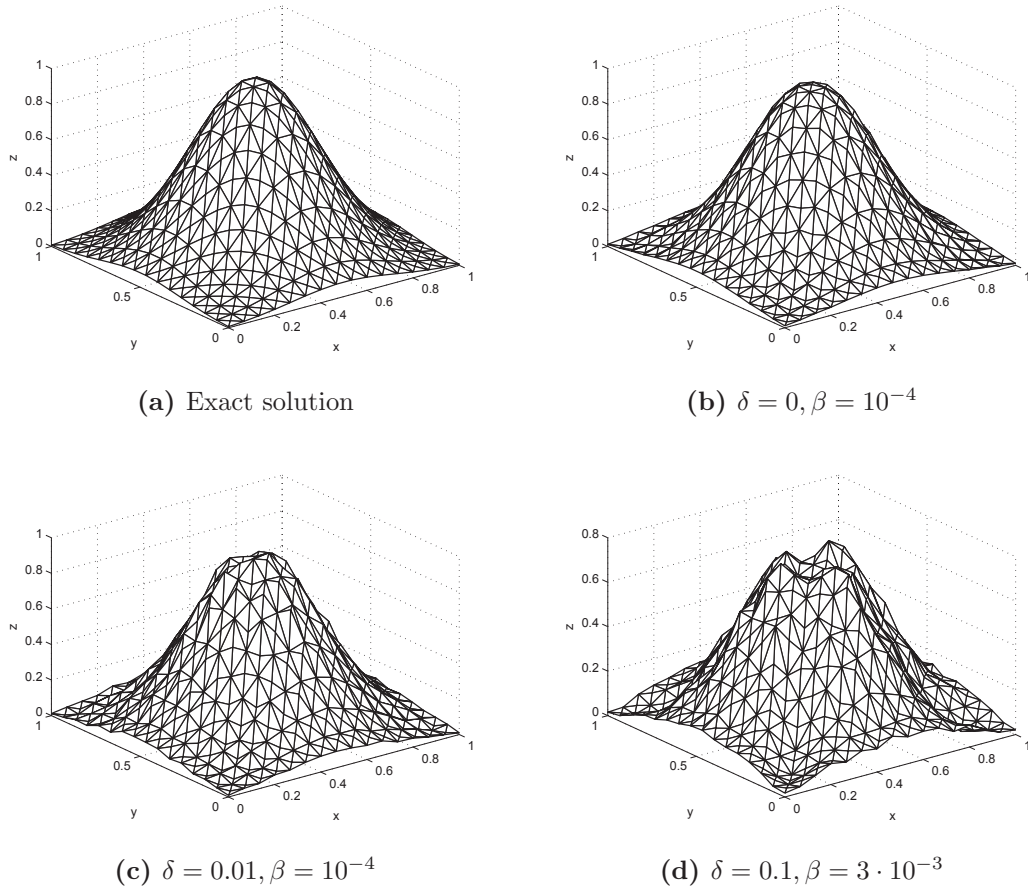
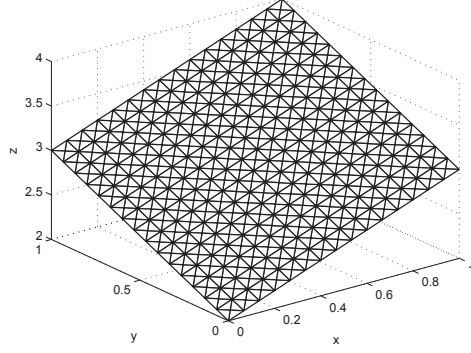


Figure 5.2: Recovered parameter a_N in Example 5.1.1 with different noise and regularization parameters β .

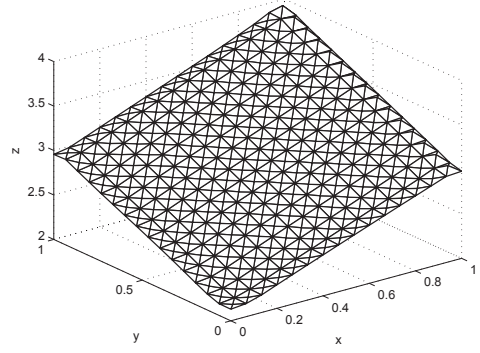
Table 5.2

Results of Example 5.1.2 with $\delta = 0$ and regularization parameter $\beta = h^2$.

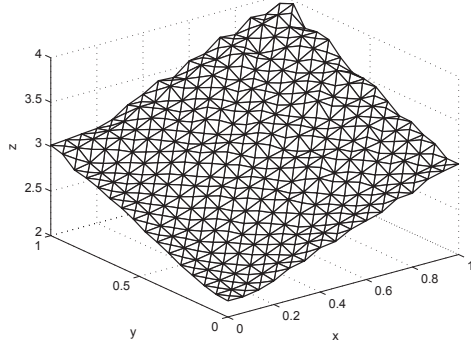
h	$\ a_{h,\beta} - a_N\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
1/4	$8.820 \cdot 10^{-2}$	$2.910 \cdot 10^{-2}$	$4.229 \cdot 10^{-1}$
1/8	$3.570 \cdot 10^{-2}$	$1.180 \cdot 10^{-2}$	$1.458 \cdot 10^{-1}$
1/16	$1.470 \cdot 10^{-2}$	$4.800 \cdot 10^{-3}$	$5.060 \cdot 10^{-2}$
1/32	$6.500 \cdot 10^{-3}$	$2.200 \cdot 10^{-3}$	$1.780 \cdot 10^{-2}$
1/64	$3.200 \cdot 10^{-3}$	$1.100 \cdot 10^{-3}$	$6.300 \cdot 10^{-3}$



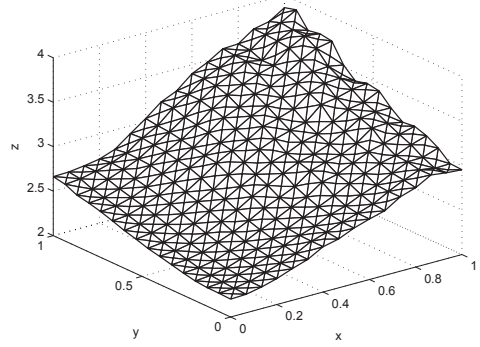
(a) Exact solution



(b) $\delta = 0, \beta = 5.5 \cdot 10^{-3}$



(c) $\delta = 0.01, \beta = 2 \cdot 10^{-2}$



(d) $\delta = 0.1, \beta = 5 \cdot 10^{-1}$

Figure 5.3: Recovered parameter in Example 5.1.2 with different noise and regularization parameters β .

When $\delta = 0$, the above examples suggest a convergence rate of $O(h)$ in the L^2 norm, and of $O(h^{1.5})$ in the quotient norm. Since the given data z_N is the exact continuous piecewise linear interpolant of u_N , we know that $\|z_N - u_N\|_{H^1(\Omega)} = O(h)$, and hence the theoretic order given in Theorem 5.1.8 would be of $O(h)$, which is less than the actual order as the above examples reveal. Notice further, the error estimates (2.9) and (2.17) provide no information regarding the convergence rate in this case.

The effect of noise is very apparent as one can see from the plots in Figures 5.2 and 5.3. So the quality of the recovered coefficient relies heavily on the quality of the observation, z_N . The choice of the regularization parameter β was chosen to be optimal, using the knowledge of the exact solution, but obviously without

any a priori information about the exact coefficient, a good choice of regularization parameter β is vital in this type of analysis. Moreover, the need to differentiate the data makes the equation error approach very sensitive especially if the data is noisy. Therefore one should consider some sort of data-smoothing techniques to stabilize the method. Automatic data smoothing and robust methods for choosing the regularization parameter β will be the subject of Chapter 6.

5.2 Equations of isotropic elasticity

Following [11], we will express the inverse problem in terms of the shear modulus μ^* and the bulk modulus $\rho^* = \mu^* + \lambda^*$ instead of in terms of μ^* and λ^* . Using this convention, we can rewrite the stress tensor $\sigma = \sigma(m, u) = 2\mu\epsilon(u) + \lambda\text{tr}(\epsilon(u))I$ as

$$\sigma(m, u) = \begin{bmatrix} m_2\text{tr}(\epsilon(u)) + m_1(\epsilon(u)_{11} - \epsilon(u)_{22}) & 2m_1\epsilon(u)_{12} \\ 2m_1\epsilon(u)_{12} & m_2\text{tr}(\epsilon(u)) - m_1(\epsilon(u)_{11} - \epsilon(u)_{22}) \end{bmatrix}$$

where $m = (m_1, m_2) = (\mu, \rho)$, $\rho = \lambda + \mu$, and $\epsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$.

Let S_E denote the set of all $m = (\mu, \rho) \in H^1(\Omega)^2$ satisfying the weak form (3.19), that is

$$S_E = \{m = (\mu, \rho) \in H^1(\Omega)^2 \mid \ell_m^E(u_E) = \ell^E \text{ in } (H^1(\Omega)^2)^*\}.$$

We shall assume S_E is nonempty. Further, let $m^* = (\mu^*, \rho^*)$ be any element in S_E .

We start with the following stability result regarding the equation error method applied to the elasticity problem.

Theorem 5.2.1. *Suppose that $u_E \in W^{1,\infty}(\Omega)^2$. Let $\{z_n\} \subset W^{1,\infty}(\Omega)^2$ be a sequence of observations of u_E , and let $\{\epsilon_n\}$ and $\{\gamma_n\}$ be two sequences of real numbers such that*

1. $\epsilon_n^2 \leq \gamma_n \leq \epsilon_n \ \forall n \in \mathbb{N}$, and $\epsilon_n^2/\gamma_n \rightarrow 0$ as $n \rightarrow \infty$,
2. $\|u_E - z_n\|_{W^{1,\infty}(\Omega)} \leq \epsilon_n \ \forall n \in \mathbb{N}$,
3. $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

For each $n \in \mathbb{N}$, let $m_n \in H^1(\Omega)^2$ be the unique solution of the optimization problem

$$\min_{m \in H^1(\Omega)^2} \frac{1}{2} \|\ell_m^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 + \frac{\gamma_n}{2} \|m\|_{H^1(\Omega)}^2.$$

Then there is an $\tilde{m} \in S_E$ such that $\|m_n - \tilde{m}\|_{H^1(\Omega)} \rightarrow 0$. Further, $\|\tilde{m}\|_{H^1(\Omega)} \leq \|m^*\|_{H^1(\Omega)}$ for all $m^* \in S_E$.

Proof. Using (3.21) and the fact that $\ell_{m^*}^E(u_E) = \ell^E$ in $(H^1(\Omega)^2)^*$ for all $m^* \in S_E$, we have

$$\begin{aligned}
\gamma_n \|m_n\|_{H^1(\Omega)}^2 &\leq \|\ell_{m^*}^E(z_n) - \ell^E\|_{H^1(\Omega)^*}^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2 \\
&= \|\ell_{m^*}^E(z_n - u_E)\|_{H^1(\Omega)^*}^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2 \\
&\leq \|m^*\|_{L^2(\Omega)}^2 \|\nabla(z_n - u_E)\|_{L^\infty(\Omega)}^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2 \\
&\leq \|m^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2 \\
&\leq \gamma_n \|m^*\|_{L^2(\Omega)}^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2,
\end{aligned} \tag{5.11}$$

and consequently

$$\|m_n\|_{H^1(\Omega)}^2 \leq \|m^*\|_{L^2(\Omega)}^2 + \|m^*\|_{H^1(\Omega)}^2 \quad \forall n \in \mathbb{N}, m^* \in S_E.$$

Therefore $\{m_n\}$ is bounded in $H^1(\Omega)$, and by Corollary 3.2.1, there exists a subsequence of $\{m_n\}$, which we still denote by $\{m_n\}$, and a vector $\tilde{m} \in H^1(\Omega)$ such that $m_n \rightharpoonup \tilde{m}$ in $H^1(\Omega)^2$ and $m_n \rightarrow \tilde{m}$ in $L^2(\Omega)^2$. Now from (3.21) we have

$$\begin{aligned}
\|\ell_{m_n}^E(u_E) - \ell_{m_n}^E(z_n)\|_{H^1(\Omega)^*}^2 &= \|\ell_{m_n}^E(u_E - z_n)\|_{H^1(\Omega)^*}^2 \\
&\leq \|m_n\|_{L^2(\Omega)}^2 \|\nabla(u_E - z_n)\|_{L^\infty(\Omega)}^2 \\
&\leq \|m_n\|_{L^2(\Omega)}^2 \epsilon_n^2,
\end{aligned} \tag{5.12}$$

and by the definition of m_n we also have

$$\begin{aligned}
\|\ell_{m_n}^E(z_n) - \ell^E\|_{H^1(\Omega)^*}^2 &\leq \|\ell_{m^*}^E(z_n) - \ell^E\|_{H^1(\Omega)^*}^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2 \\
&= \|\ell_{m^*}^E(z_n - u_E)\|_{H^1(\Omega)^*}^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2 \\
&\leq \|m^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \gamma_n \|m^*\|_{H^1(\Omega)}^2.
\end{aligned} \tag{5.13}$$

Combine (5.12), (5.13) and the fact that a convergent sequence is necessarily bounded

to see that

$$\begin{aligned}
\|\ell_{m_n}^E(u_E) - \ell^E\|_{H^1(\Omega)^*}^2 &\leq 2\|\ell_{a_n}^E(u_E) - \ell_{m_n}^E(z_n)\|_{H^1(\Omega)^*}^2 + 2\|\ell_{m_n}^E(z_n) - \ell^E\|_{H^1(\Omega)^*}^2 \\
&\leq 2\|m_n\|_{L^2(\Omega)}^2 \epsilon_n^2 + 2\left(\|m^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \epsilon_n \|m^*\|_{H^1(\Omega)}^2\right) \\
&\leq C(\epsilon_n^2 + \epsilon_n).
\end{aligned}$$

Consequently, in view of (4.12), we conclude that

$$\|\ell_{\tilde{m}}^E(u_E) - \ell^E\|_{H^1(\Omega)^*} = \lim_{n \rightarrow \infty} \|\ell_{m_n}^E(u_E) - \ell^E\|_{H^1(\Omega)^*} = 0,$$

showing $\tilde{m} \in S_E$. Since $m_n \rightharpoonup \tilde{m}$ in $H^1(\Omega)$, by Theorem 3.1.5 we have

$$\|\tilde{m}\|_{H^1(\Omega)} \leq \varliminf_{n \rightarrow \infty} \|m_n\|_{H^1(\Omega)},$$

and by the fourth inequality in (5.11) and the fact that $\tilde{m} \in S_E$ we have

$$\overline{\lim}_{n \rightarrow \infty} \|m_n\|_{H^1(\Omega)}^2 \leq \lim_{n \rightarrow \infty} \left(\frac{\epsilon_n^2}{\gamma_n} \|\tilde{m}\|_{L^2(\Omega)}^2 + \|\tilde{m}\|_{H^1(\Omega)}^2 \right) = \|\tilde{m}\|_{H^1(\Omega)}^2.$$

Hence $\|m_n\|_{H^1(\Omega)} \rightarrow \|\tilde{m}\|_{H^1(\Omega)}$, and by a standard functional analysis result, we conclude that $m_n \rightarrow \tilde{m}$ in $H^1(\Omega)$. Further, in view of (5.11), we conclude

$$\|\tilde{m}\|_{H^1(\Omega)}^2 = \lim_{n \rightarrow \infty} \|m_n\|_{H^1(\Omega)}^2 \leq \lim_{n \rightarrow \infty} \left(\frac{\epsilon_n^2}{\gamma_n} \|m^*\|_{L^2(\Omega)}^2 + \|m^*\|_{H^1(\Omega)}^2 \right) = \|m^*\|_{H^1(\Omega)}^2,$$

for all $m^* \in S_E$, which completes the proof. \square

Now we will derive an error bound on the residual $R(z_E) = \sigma(m_\gamma, z_E) - \sigma(m^*, z_E)$, where $m_\gamma = (\mu_\gamma, \rho_\gamma)$ is the solution of the optimization problem (4.3). With a little algebra, the 2-tensor $R(z_E)$ can be written more concisely as

$$R(z_E) = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix},$$

where

$$\begin{aligned} R_{11} &= (\rho_\gamma - \rho^*)\text{tr}(\epsilon(z_E)) + (\mu_\gamma - \mu^*)(\epsilon(z_E)_{11} - \epsilon(z_E)_{22}), \\ R_{12} &= R_{21} = 2(\mu_\gamma - \mu^*)\epsilon(z_E)_{12}, \\ R_{22} &= (\rho_\gamma - \rho^*)\text{tr}(\epsilon(z_E)) - (\mu_\gamma - \mu^*)(\epsilon(z_E)_{11} - \epsilon(z_E)_{22}). \end{aligned}$$

Let us assume $z_E \in W^{1,\infty}(\Omega)$ so that $R(z_E) \in L^2(\Omega)^{2 \times 2}$. Let R_1 and R_2 be the first and second rows of $R(z_E)$, respectively. Then, by Theorem 5.1.2 we have the decomposition

$$\begin{aligned} R_1 &= \nabla q_1 + \text{curl} \phi_1, \\ R_2 &= \nabla q_2 + \text{curl} \phi_2. \end{aligned}$$

Therefore, we see that $R(z_E)$ has the decomposition $R(z_E) = \nabla q + \text{curl} \phi$ where $q = (q_1, q_2)$ and $\phi = (\phi_1, \phi_2)$. Similarly, for the finite-dimensional case, we decompose the first and second rows of residual $R_h(z_E) = \sigma(m_{h,\gamma}, z_E) - \sigma(m^*, z_E)$ as

$$\begin{aligned} R_{h,1} &= \nabla q_{h,1} + \text{curl} \phi_{h,1}, \\ R_{h,2} &= \nabla q_{h,2} + \text{curl} \phi_{h,2}. \end{aligned}$$

Then we have the decomposition $R_h(z_E) = \nabla q_h + \text{curl} \phi_h$ where $q_h = (q_{h,1}, q_{h,2})$ and $\phi_h = (\phi_{h,1}, \phi_{h,2})$. In Theorem 5.2.4, we derive an upper bound on $\|\nabla q_h\|_{L^2(\Omega)}$, where $m_{h,\gamma}$ is the solution of the optimization problem (4.15).

The following theorem gives an error error bound on the rotation-free part of the residual $R(z_E)$, i.e. ∇q .

Theorem 5.2.2. *Assume that $m^* \in H^1(\Omega)^2$ and $u_E, z_E \in W^{1,\infty}(\Omega)^2$, and let m_γ be the solution of the optimization problem (4.3). If $\|u_E - z_E\|_{W^{1,\infty}(\Omega)} \leq \epsilon$, then we have the error estimate*

$$\|\nabla q\|_{L^2(\Omega)} \leq C(\epsilon + \sqrt{\gamma}),$$

where C depends on $\|m^*\|_{H^1(\Omega)}$ but is independent of γ and ϵ .

Proof. By Theorem 5.1.2, for all $v \in H^1(\Omega)^2$ we have

$$\begin{aligned}
(\nabla q, \nabla v)_{L^2(\Omega)} &= (R, \nabla v)_{L^2(\Omega)} \\
&= (\sigma(m_\gamma, z_E), \nabla v)_{L^2(\Omega)} - (\sigma(m^*, u_E), \nabla v)_{L^2(\Omega)} \\
&\quad + (\sigma(m^*, u_E - z_E), \nabla v)_{L^2(\Omega)} \\
&= \left\langle \ell_{m_\gamma}^E(z_E) - \ell^E, v \right\rangle_* + (\sigma(m^*, u_E - z_E), \nabla v)_{L^2(\Omega)},
\end{aligned}$$

In particular, taking $v = q$ then applying Corollary 3.3.1, yields

$$\begin{aligned}
\|\nabla q\|_{L^2(\Omega)}^2 &\leq \|\ell_{m_\gamma}^E(z_E) - \ell^E\|_{H^1(\Omega)^*} \|q\|_{H^1(\Omega)} + \|\sigma(m^*, u_E - z_E)\|_{L^2(\Omega)} \|\nabla q\|_{L^2(\Omega)} \\
&\leq \left(\|\ell_{m_\gamma}^E(z_E) - \ell^E\|_{H^1(\Omega)^*} + \|\sigma(m^*, u_E - z_E)\|_{L^2(\Omega)} \right) \|q\|_{H^1(\Omega)} \\
&\leq C \left(\|\ell_{m_\gamma}^E(z_E) - \ell^E\|_{H^1(\Omega)^*} + \|\sigma(m^*, u_E - z_E)\|_{L^2(\Omega)} \right) \|\nabla q\|_{L^2(\Omega)}.
\end{aligned}$$

Consequently, we have

$$\|\nabla q\|_{L^2(\Omega)}^2 \leq C \left(\|\ell_{m_\gamma}^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 + \|\sigma(m^*, u_E - z_E)\|_{L^2(\Omega)}^2 \right). \quad (5.14)$$

First, we have the bound

$$\|\sigma(m^*, u_E - z_E)\|_{L^2(\Omega)}^2 \leq C \|m^*\|_{L^2(\Omega)}^2 \|u_E - z_E\|_{W^{1,\infty}(\Omega)}^2 \leq C \epsilon^2. \quad (5.15)$$

Using (4.8) and noting the facts that $\ell_{m^*}^E(u_E) = \ell^E$ in $H^1(\Omega)^*$ and that m_γ solves (4.3), we obtain

$$\begin{aligned}
\|\ell_{m_\gamma}^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 &\leq \|\ell_{m^*}^E(z_E - u_E)\|_{H^1(\Omega)^*}^2 + \gamma \|m^*\|_{H^1(\Omega)}^2 \\
&\leq C \|m^*\|_{L^2(\Omega)}^2 \|z_E - u_E\|_{W^{1,\infty}(\Omega)}^2 + \gamma \|m^*\|_{H^1(\Omega)}^2 \\
&\leq C (\epsilon^2 + \gamma).
\end{aligned} \quad (5.16)$$

Combine (5.14), (5.15), and (5.16), and the result follows. \square

Using (4.9) instead of (4.8) in the proof above, it is easy to prove the following result, which allows the error in the data to be measured in the weaker H^1 norm instead of the $W^{1,\infty}$ norm, and eliminates the requirement that $u_E \in W^{1,\infty}(\Omega)^2$.

Theorem 5.2.3. *Assume that $m^* \in L^\infty(\Omega) \cap H^1(\Omega)$ and $z_E \in W^{1,\infty}(\Omega)$, and let*

m_γ be the solution of the optimization problem (4.3). If $\|u_E - z_E\|_{H^1(\Omega)} \leq \epsilon$, then we have the error estimate

$$\|\nabla q\|_{L^2(\Omega)} \leq C(\epsilon + \sqrt{\gamma}),$$

where C depends on $\|m^*\|_{L^\infty(\Omega)}$ and $\|m^*\|_{H^1(\Omega)}$ but is independent of γ and ϵ .

Remark 5.2.1. If desired, in Theorem 5.2.2 ∇q can be defined as the rotation-free part of the residual $R(u_E) = \sigma(m_\gamma, u_E) - \sigma(m^*, u_E)$; the same result can be concluded provided $\gamma \geq C\epsilon^2$.

Now we give an error estimate for the finite element solution of the equation error approach for estimating the Lamé moduli.

Theorem 5.2.4. Assume that $m^* \in H^r(\Omega)^2$ for some $1 \leq r \leq k+1$, and that $u_E, z_E \in W^{1,\infty}(\Omega)^2$. Let $m_{h,\gamma}$ be the solution of the optimization problem (4.15). If $\|u_E - z_E\|_{W^{1,\infty}(\Omega)} \leq \epsilon$, then we have the estimate

$$\|\nabla q_h\|_{L^2(\Omega)} \leq C(h^r + \epsilon + \sqrt{\gamma}),$$

where C depends on $\|m^*\|_{H^r(\Omega)}$ and $\|z_E\|_{W^{1,\infty}(\Omega)}$ but is independent of γ , ϵ , and h .

Proof. Let \tilde{m}_h be the H^1 -projection of m^* into $V_h^k \times V_h^k$. Using the same method as in the proof of Theorem 5.2.2, we obtain

$$\begin{aligned} \|\nabla q_h\|_{L^2(\Omega)}^2 &\leq C \left(\|\ell_{m_{h,\gamma}}^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{\tilde{m}_h}^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 + \gamma \|\tilde{m}_h\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{(\tilde{m}_h - m^*)}^E(z_E)\|_{H^1(\Omega)^*}^2 + \|\ell_{m^*}^E(z_E) - \ell^E\|_{H^1(\Omega)^*}^2 + \gamma \|m^*\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{(\tilde{m}_h - m^*)}^E(z_E)\|_{H^1(\Omega)^*}^2 + \epsilon^2 + \gamma \right). \end{aligned}$$

We need to bound the first term on right-hand side of the above inequality. Using the same reasoning that led to (5.9), we conclude that

$$\|\tilde{m}_h - m^*\|_{L^2(\Omega)} \leq Ch^r, \quad \text{for } 1 \leq r \leq k+1. \quad (5.17)$$

Finally, using (5.17) and (4.4) we have

$$\|\ell_{(\tilde{m}_h - m^*)}^E(z_E)\|_{H^1(\Omega)^*} \leq \|\nabla z_E\|_{L^\infty(\Omega)} \|\tilde{m}_h - m^*\|_{L^2(\Omega)} \leq Ch^r,$$

which completes the proof. \square

With a minor modification of the proof above, one can easily conclude the following result, which allows the error in the data to be measured in the H^1 norm instead of the $W^{1,\infty}$ norm, and eliminates the requirement that $u_E \in W^{1,\infty}(\Omega)$.

Theorem 5.2.5. *Assume that $m^* \in L^\infty(\Omega)^2 \cap H^r(\Omega)^2$ for some $1 \leq r \leq k+1$, and that $z_E \in W^{1,\infty}(\Omega)^2$. Let $m_{h,\gamma}$ be the solution of the optimization problem (4.15). If $\|u_E - z_E\|_{H^1(\Omega)} \leq \epsilon$, then*

$$\|\nabla q_h\|_{L^2(\Omega)} \leq C(h^r + \epsilon + \sqrt{\gamma}),$$

where C depends on $\|m^*\|_{L^\infty(\Omega)}$, $\|m^*\|_{H^r(\Omega)}$, and $\|z_E\|_{W^{1,\infty}(\Omega)}$ but is independent of γ , ϵ , and h .

5.2.1 Numerical examples

We consider the finite element solution using the space of continuous piecewise linear polynomials over the unit square $[0, 1]^2$ with a uniform triangulation similar to the one shown in Figure 5.1. The observation z_E of u_E is generated using a formula similar to (5.10). Further, let δ denote the relative error (measured in the H^1 norm) in the data z_E .

Example 5.2.1. *We consider the problem of recovering the Lamé moduli μ^* and λ^* from the displacement u_E given by*

$$u_E(x, y) = \begin{bmatrix} 5x + y \\ x^2 + 4y^3 \end{bmatrix}.$$

Here the body f_E and the traction g_E are chosen so that the exact parameters are

$$\mu^*(x, y) = \cos(5(x + \exp(y))) + 2,$$

$$\lambda^*(x, y) = x + y + 1.$$

Table 5.3

Error results in estimating μ^* in Example 5.2.1, the noise level $\delta = 0$ and regularization parameter $\gamma = h^2$.

h	$\ \mu_{h,\gamma} - \mu^*\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \mu_{h,\gamma} - \mu^*\ _{L^\infty(\Omega)}$
1/4	$3.857 \cdot 10^{-1}$	$1.855 \cdot 10^{-1}$	1.1167
1/8	$1.603 \cdot 10^{-1}$	$7.710 \cdot 10^{-2}$	$5.413 \cdot 10^{-1}$
1/16	$7.240 \cdot 10^{-2}$	$3.480 \cdot 10^{-2}$	$2.337 \cdot 10^{-1}$
1/32	$2.730 \cdot 10^{-2}$	$1.310 \cdot 10^{-2}$	$1.143 \cdot 10^{-1}$
1/55	$1.130 \cdot 10^{-2}$	$5.400 \cdot 10^{-3}$	$7.300 \cdot 10^{-2}$

Table 5.4

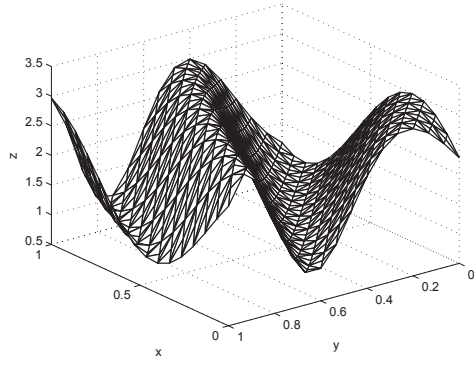
Error results in estimating λ^* in Example 5.2.1, the noise level $\delta = 0$ and regularization parameter $\gamma = h^2$.

h	$\ \lambda_{h,\gamma} - \lambda^*\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \lambda_{h,\gamma} - \lambda^*\ _{L^\infty(\Omega)}$
1/4	$3.047 \cdot 10^{-1}$	$1.493 \cdot 10^{-1}$	1.1452
1/8	$1.589 \cdot 10^{-1}$	$7.790 \cdot 10^{-2}$	$5.253 \cdot 10^{-1}$
1/16	$7.860 \cdot 10^{-2}$	$3.850 \cdot 10^{-2}$	$2.875 \cdot 10^{-1}$
1/32	$3.030 \cdot 10^{-2}$	$1.490 \cdot 10^{-2}$	$1.539 \cdot 10^{-1}$
1/55	$1.260 \cdot 10^{-2}$	$6.200 \cdot 10^{-3}$	$9.850 \cdot 10^{-2}$

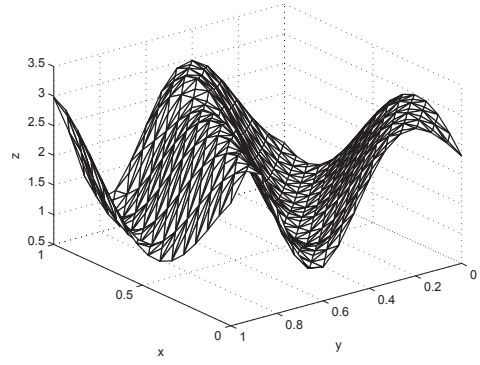
First we consider the case in which $\delta = 0$, and so z_E represents the interpolant of u_E . The computed errors are presented in Tables 5.3 and 5.4. Next, we plot the exact and the recovered parameters with the presence of noise, results for various noise levels are shown in Figures 5.4 and 5.5.

From the results in Table 5.3, we notice an $O(h^{1.3})$ convergence in the L^2 norm and of $O(h)$ in the L^∞ norm in estimating the μ^* . Table 5.4 shows that the convergence in estimating λ^* is of $O(h^{1.2})$ in the L^2 norm and of $O(h^{0.9})$ in the L^∞ norm.

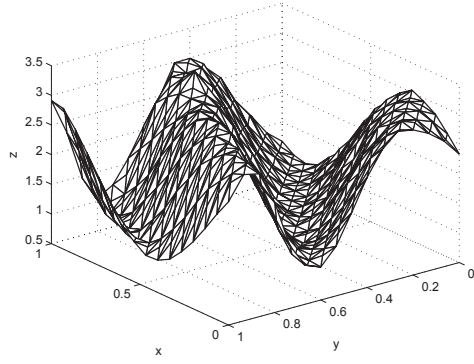
The effect of the noise is clear as the plots in Figures 5.4 and 5.5 show. However, when using exact data, the recovered parameters are in a good match with the exact



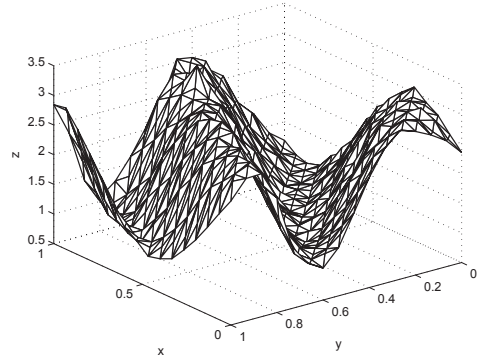
(a) Exact modulus μ^* .



(b) $\delta = 0, \alpha = 5.5 \cdot 10^{-4}$



(c) $\delta = 0.005, \alpha = 7.5 \cdot 10^{-4}$



(d) $\delta = 0.01, \alpha = 10^{-3}$

Figure 5.4: Recovered μ^* in Example 5.2.1 with different noise and regularization parameters α .

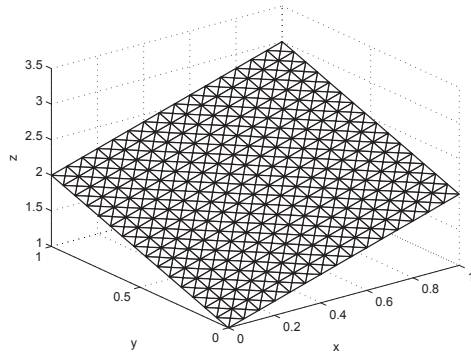
parameters. We point out that the regularization parameter γ was chosen by trial and error using the knowledge of the exact solutions, and it is nearly optimal.

5.3 The Dirichlet problem

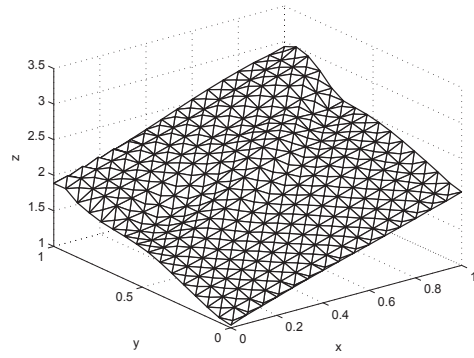
In this section, let S_D be the set of all $a \in H^1(\Omega)$ satisfying (3.17), that is

$$S_D = \{a \in H^1(\Omega) \mid \ell_a^D(u_D) = \ell^D \text{ in } H^{-1}(\Omega)\}.$$

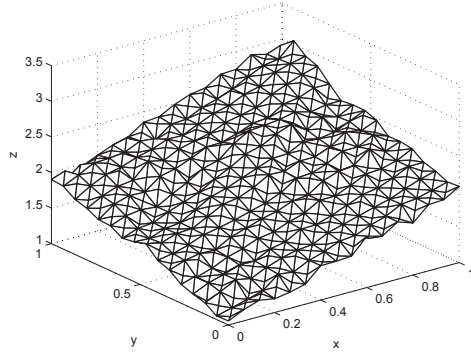
We shall make the assumption that S_D is nonempty. Throughout this section, let a_D be any element in S_D .



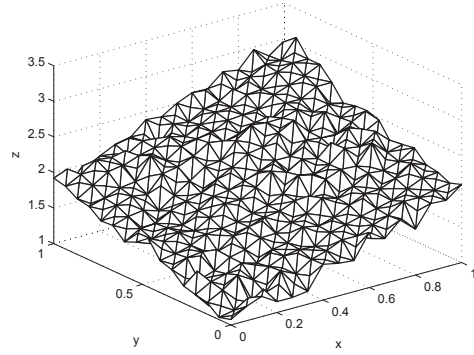
(a) Exact modulus λ^* .



(b) $\delta = 0, \alpha = 5.5 \cdot 10^{-4}$



(c) $\delta = 0.005, \alpha = 7.5 \cdot 10^{-4}$



(d) $\delta = 0.01, \alpha = 10^{-3}$

Figure 5.5: Recovered λ^* in Example 5.2.1 with different noise and regularization parameters α .

First we prove the convergence of equation error method applied to the inverse problem related to the BVP (1.2).

Theorem 5.3.1. *Suppose that $u_D \in W^{1,\infty}(\Omega)$. Let $\{z_n\} \subset W^{1,\infty}(\Omega)$ be a sequence of observations of u_D , and let $\{\epsilon_n\}$ and $\{\alpha_n\}$ be two sequences of real numbers such that*

1. $\epsilon_n^2 \leq \alpha_n \leq \epsilon_n \ \forall n \in \mathbb{N}$, and $\epsilon_n^2/\alpha_n \rightarrow 0$ as $n \rightarrow \infty$,
2. $\|u_D - z_n\|_{W^{1,\infty}(\Omega)} \leq \epsilon_n \ \forall n \in \mathbb{N}$,
3. $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

For each $n \in \mathbb{N}$, let $a_n \in H^1(\Omega)$ be the unique solution (guaranteed by Theorem

4.2.3) of the optimization problem

$$\min_{b \in H^1(\Omega)} \frac{1}{2} \|\ell_b^D(z_n) - \ell^D\|_{H^{-1}(\Omega)}^2 + \frac{\alpha_n}{2} \|b\|_{H^1(\Omega)}^2.$$

Then there is an $\tilde{a} \in S_D$ such that $\|a_n - \tilde{a}\|_{H^1(\Omega)} \rightarrow 0$. Further, $\|\tilde{a}\|_{H^1(\Omega)} \leq \|a^*\|_{H^1(\Omega)}$ for all $a^* \in S_D$.

Proof. Using (4.5) and the fact that $\ell_{a^*}^D(u_D) = \ell^D$ in $H^{-1}(\Omega)$ for all $a^* \in S_D$, we have

$$\begin{aligned} \alpha_n \|a_n\|_{H^1(\Omega)}^2 &\leq \|\ell_{a^*}^D(z_n) - \ell^D\|_{H^{-1}(\Omega)}^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2 \\ &= \|\ell_{a^*}^D(z_n - u_D)\|_{H^{-1}(\Omega)}^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2 \\ &\leq \|a^*\|_{L^2(\Omega)}^2 \|\nabla(z_n - u_D)\|_{L^\infty(\Omega)}^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2 \\ &\leq \|a^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2 \\ &\leq \alpha_n \|a^*\|_{L^2(\Omega)}^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2, \end{aligned} \quad (5.18)$$

and consequently

$$\|a_n\|_{H^1(\Omega)}^2 \leq \|a^*\|_{L^2(\Omega)}^2 + \|a^*\|_{H^1(\Omega)}^2 \quad \forall n \in \mathbb{N}, a^* \in S_D.$$

Therefore $\{a_n\}$ is bounded in $H^1(\Omega)$, and by Corollary 3.2.1, there exists a subsequence of $\{a_n\}$, which we still denote by $\{a_n\}$, and a vector $\tilde{a} \in H^1(\Omega)$ such that $a_n \rightharpoonup \tilde{a}$ in $H^1(\Omega)$ and $a_n \rightarrow \tilde{a}$ in $L^2(\Omega)$. Now from (4.5) we have

$$\begin{aligned} \|\ell_{a_n}^D(u_D) - \ell_{a_n}^D(z_n)\|_{H^{-1}(\Omega)}^2 &= \|\ell_{a_n}^D(u_D - z_n)\|_{H^{-1}(\Omega)}^2 \\ &\leq \|a_n\|_{L^2(\Omega)}^2 \|\nabla(u_D - z_n)\|_{L^\infty(\Omega)}^2 \\ &\leq \|a_n\|_{L^2(\Omega)}^2 \epsilon_n^2, \end{aligned} \quad (5.19)$$

and by the definition of a_n we also have

$$\begin{aligned} \|\ell_{a_n}^D(z_n) - \ell^D\|_{H^{-1}(\Omega)}^2 &\leq \|\ell_{a^*}^D(z_n) - \ell^D\|_{H^{-1}(\Omega)}^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2 \\ &= \|\ell_{a^*}^D(z_n - u_D)\|_{H^{-1}(\Omega)}^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2 \\ &\leq \|a^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \alpha_n \|a^*\|_{H^1(\Omega)}^2. \end{aligned} \quad (5.20)$$

Combine (5.19), (5.20) and the fact that a convergent sequence is necessarily bounded

to see that

$$\begin{aligned}
\|\ell_{a_n}^D(u_D) - \ell^D\|_{H^{-1}(\Omega)}^2 &\leq 2\|\ell_{a_n}^D(u_D) - \ell_{a_n}^D(z_n)\|_{H^{-1}(\Omega)}^2 + 2\|\ell_{a_n}^D(z_n) - \ell^D\|_{H^{-1}(\Omega)}^2 \\
&\leq 2\|a_n\|_{L^2(\Omega)}^2 \epsilon_n^2 + 2\left(\|a^*\|_{L^2(\Omega)}^2 \epsilon_n^2 + \epsilon_n \|a^*\|_{H^1(\Omega)}^2\right) \\
&\leq C(\epsilon_n^2 + \epsilon_n).
\end{aligned}$$

Consequently, in view of (4.11), we conclude that

$$\|\ell_{\tilde{a}}^D(u_D) - \ell^D\|_{H^{-1}(\Omega)} = \lim_{n \rightarrow \infty} \|\ell_{a_n}^D(u_D) - \ell^D\|_{H^{-1}(\Omega)} = 0,$$

showing $\tilde{a} \in S_D$. Since $a_n \rightharpoonup \tilde{a}$ in $H^1(\Omega)$, by Theorem 3.1.5 we have

$$\|\tilde{a}\|_{H^1(\Omega)} \leq \varliminf_{n \rightarrow \infty} \|a_n\|_{H^1(\Omega)},$$

and by the fourth inequality in (5.18) and the fact that $\tilde{a} \in S_D$ we have

$$\overline{\lim}_{n \rightarrow \infty} \|a_n\|_{H^1(\Omega)}^2 \leq \lim_{n \rightarrow \infty} \left(\frac{\epsilon_n^2}{\alpha_n} \|\tilde{a}\|_{L^2(\Omega)}^2 + \|\tilde{a}\|_{H^1(\Omega)}^2 \right) = \|\tilde{a}\|_{H^1(\Omega)}^2.$$

Hence $\|a_n\|_{H^1(\Omega)} \rightarrow \|\tilde{a}\|_{H^1(\Omega)}$, and by a standard functional analysis result, we conclude that $a_n \rightarrow \tilde{a}$ in $H^1(\Omega)$. Further, in view of (5.18), we conclude

$$\|\tilde{a}\|_{H^1(\Omega)}^2 = \lim_{n \rightarrow \infty} \|a_n\|_{H^1(\Omega)}^2 \leq \lim_{n \rightarrow \infty} \left(\frac{\epsilon_n^2}{\alpha_n} \|a^*\|_{L^2(\Omega)}^2 + \|a^*\|_{H^1(\Omega)}^2 \right) = \|a^*\|_{H^1(\Omega)}^2,$$

for all $a^* \in S_D$, which completes the proof. \square

To derive an upper bound on the error, we measure the error in the quotient norm $\|\cdot\|_Q$ introduced in Section 4.1. To this end, we need the following result (see [34, Proposition 1, p. 215] and [35, Theorem 3.2]).

Theorem 5.3.2. *Let Ω be a connected open set in \mathbb{R}^2 , with Lipschitz-continuous boundary $\partial\Omega$. Then every function \vec{u} of $L^2(\Omega)^2$ admits the following orthogonal decomposition:*

$$\vec{u} = \nabla q + \vec{v},$$

where $q \in H_0^1(\Omega)$ is the only solution of

$$(\nabla q, \nabla \mu)_{L^2(\Omega)} = (\vec{u}, \nabla \mu)_{L^2(\Omega)} \quad \forall \mu \in H_0^1(\Omega), \quad (5.21)$$

and $\vec{v} \in L^2(\Omega)^2$ with $\operatorname{div} \vec{v} = 0$.

Now we are ready to state our first error estimate for the equation error method applied to BVP (1.2).

Theorem 5.3.3. *Assume that $a_D \in H^1(\Omega)$ and $u_D, z_D \in W^{1,\infty}(\Omega)$, and let a_α be the solution of the optimization problem (4.2). If $\|u_D - z_D\|_{W^{1,\infty}(\Omega)} \leq \epsilon$, then for the error vector $\vec{e} = (a_\alpha - a_D)\nabla z_D$ we have*

$$\|\dot{e}\|_Q \leq C(\epsilon + \sqrt{\alpha}),$$

where C depends on $\|a_D\|_{H^1(\Omega)}$ but is independent of α and ϵ .

Proof. Obviously $\vec{e} \in L^2(\Omega)^2$ and so by Theorem 5.3.2 we have the decomposition

$$\vec{e} = \nabla q + \vec{v}.$$

Since by Theorem 5.1.4 $\nabla q \in R$, we have $\operatorname{proj}_{R^\perp} \vec{e} = \vec{v}$, and consequently

$$\|\dot{e}\|_Q = \|\vec{e} - \vec{v}\|_{L^2(\Omega)} = \|\nabla q\|_{L^2(\Omega)}.$$

Now by (5.21) and the fact that $\ell_{a_D}^D(u_D) = \ell^D$ in $H^{-1}(\Omega)$, we have

$$\begin{aligned} (\nabla q, \nabla \mu)_{L^2(\Omega)} &= (\vec{e}, \nabla \mu)_{L^2(\Omega)} \\ &= (a_\alpha \nabla z_D, \nabla \mu)_{L^2(\Omega)} - (a_D \nabla z_D, \nabla \mu)_{L^2(\Omega)} \\ &= (a_\alpha \nabla z_D, \nabla \mu)_{L^2(\Omega)} - (a_D \nabla u_D, \nabla \mu)_{L^2(\Omega)} \\ &\quad + (a_D \nabla u_D, \nabla \mu)_{L^2(\Omega)} - (a_D \nabla z_D, \nabla \mu)_{L^2(\Omega)} \\ &= \langle \ell_{a_\alpha}^D(z_D) - \ell^D, \mu \rangle_{-1} + (a_D \nabla(u_D - z_D), \nabla \mu)_{L^2(\Omega)} \quad \forall \mu \in H_0^1(\Omega). \end{aligned}$$

If we take $\mu = q$ and apply Poincaré inequality (Theorem 3.3.1), we get

$$\begin{aligned}\|\nabla q\|_{L^2(\Omega)}^2 &\leq \|\ell_{a_\alpha}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}\|q\|_{H^1(\Omega)} \\ &\quad + \|a_D\|_{L^2(\Omega)}\|\nabla(u_D - z_D)\|_{L^\infty(\Omega)}\|\nabla q\|_{L^2(\Omega)} \\ &\leq C\left(\|\ell_{a_\alpha}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}\|\nabla q\|_{L^2(\Omega)} + \epsilon\|\nabla q\|_{L^2(\Omega)}\right).\end{aligned}$$

Consequently, we have

$$\|\nabla q\|_{L^2(\Omega)}^2 \leq C\left(\|\ell_{a_\alpha}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \epsilon^2\right). \quad (5.22)$$

Using (4.5) and noting that $\ell_{a_D}^D(u_D) = \ell^D$ in $H^{-1}(\Omega)$, we obtain

$$\|\ell_{a_D}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)} = \|\ell_{a_D}^D(z_D - u_D)\|_{H^{-1}(\Omega)} \leq \epsilon\|a_D\|_{L^2(\Omega)}. \quad (5.23)$$

Combine (5.22), (5.23) and the fact that a_α is the minimizer of (4.2) to see that

$$\|\nabla q\|_{L^2(\Omega)}^2 \leq C\left(\|\ell_{a_D}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \alpha\|a_D\|_{H^1(\Omega)}^2 + \epsilon^2\right) \leq C(\epsilon^2 + \alpha),$$

and the result follows. \square

Using (4.7) instead of (4.5) in the proof above, it is easy to prove the following result, which allows the error in the data to be measured in the weaker H^1 norm instead of the $W^{1,\infty}$ norm, and eliminates the requirement that $u_D \in W^{1,\infty}(\Omega)$.

Theorem 5.3.4. *Assume that $a_D \in L^\infty(\Omega) \cap H^1(\Omega)$ and $z_D \in W^{1,\infty}(\Omega)$, and let a_α be the solution of the optimization problem (4.2). If $\|u_D - z_D\|_{H^1(\Omega)} \leq \epsilon$, then for the error vector $\vec{e} = (a_\alpha - a_D)\nabla z_D$ we have*

$$\|\vec{e}\|_Q \leq C(\epsilon + \sqrt{\alpha}),$$

where C depends on $\|a_D\|_{L^\infty(\Omega)}$ and $\|a_D\|_{H^1(\Omega)}$ but is independent of α and ϵ .

Remark 5.3.1. *If desired, the error vector in Theorem 5.3.3 can be defined as $\vec{e} = (a_\alpha - a_D)\nabla u_D$ provided $\alpha \geq C\epsilon^2$; the analysis changes little.*

Now we give an error estimate for the finite element solution of the equation error method applied to the inverse problem associated with the BVP (1.2).

Theorem 5.3.5. Assume that $a_D \in H^r(\Omega)$ for some $1 \leq r \leq k+1$, and that $u_D, z_D \in W^{1,\infty}(\Omega)$. Let $a_{h,\alpha}$ be the solution of the optimization problem (4.14). Suppose that $\|u_D - z_D\|_{W^{1,\infty}(\Omega)} \leq \epsilon$, then for the error vector $\vec{e}_h = (a_{h,\alpha} - a_D)\nabla z_D$ we have

$$\|\dot{e}_h\|_Q \leq C (h^r + \epsilon + \sqrt{\alpha}),$$

where C depends on $\|a_D\|_{H^r(\Omega)}$ and $\|z_D\|_{W^{1,\infty}(\Omega)}$ but is independent of α , ϵ , and h .

Proof. Let \tilde{a}_h be the H^1 -projection of a_D into V_h^k . Following the same outlines in the proof of Theorem 5.3.3, we see that

$$\begin{aligned} \|\dot{e}_h\|_Q^2 &\leq C \left(\|\ell_{a_{h,\alpha}}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{\tilde{a}_h}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \alpha \|\tilde{a}_h\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{(\tilde{a}_h - a_D)}^D(z_D)\|_{H^{-1}(\Omega)}^2 + \|\ell_{a_D}^D(z_D) - \ell^D\|_{H^{-1}(\Omega)}^2 + \alpha \|a_D\|_{H^1(\Omega)}^2 + \epsilon^2 \right) \\ &\leq C \left(\|\ell_{(\tilde{a}_h - a_D)}^D(z_D)\|_{H^{-1}(\Omega)}^2 + \epsilon^2 + \alpha \right). \end{aligned}$$

The first term on the right-hand side of the above inequality can be bounded using (4.5) and (5.9) as follows

$$\|\ell_{(\tilde{a}_h - a_D)}^D(z_D)\|_{H^{-1}(\Omega)} \leq \|\nabla z_D\|_{L^\infty(\Omega)} \|\tilde{a}_h - a_D\|_{L^2(\Omega)} \leq Ch^r,$$

which completes the proof. \square

With minor adjustments to the proof above, one can easily conclude the following result.

Theorem 5.3.6. Assume that $a_D \in L^\infty(\Omega) \cap H^r(\Omega)$ for some $1 \leq r \leq k+1$, and that $z_D \in W^{1,\infty}(\Omega)$. Let $a_{h,\alpha}$ be the solution of the optimization problem (4.14). If $\|u_D - z_D\|_{H^1(\Omega)} \leq \epsilon$, then for the error vector $\vec{e}_h = (a_{h,\alpha} - a_D)\nabla z_D$ we have

$$\|\dot{e}_h\|_Q \leq C (h^r + \epsilon + \sqrt{\alpha}),$$

where C depends on $\|a_D\|_{L^\infty(\Omega)}$, $\|a_D\|_{H^r(\Omega)}$, and $\|z_D\|_{W^{1,\infty}(\Omega)}$ but is independent of α , ϵ , and h .

Table 5.5

Results of Example 5.3.1 with $\delta = 0$ and regularization parameter $\alpha = h^2$.

h	$\ a_{h,\alpha} - a_D\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
1/8	$2.557 \cdot 10^{-1}$	$1.202 \cdot 10^{-1}$	$2.223 \cdot 10^{-1}$
1/16	$1.412 \cdot 10^{-1}$	$6.640 \cdot 10^{-2}$	$1.094 \cdot 10^{-1}$
1/32	$7.080 \cdot 10^{-2}$	$3.330 \cdot 10^{-2}$	$4.250 \cdot 10^{-2}$
1/64	$3.500 \cdot 10^{-2}$	$1.640 \cdot 10^{-2}$	$1.510 \cdot 10^{-2}$

5.3.1 Numerical examples

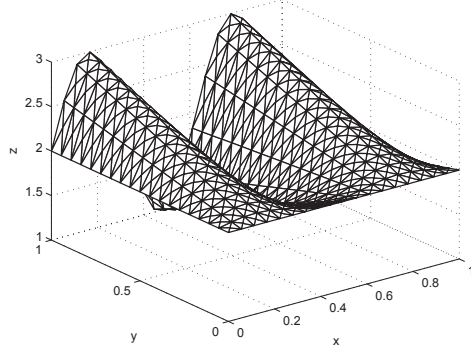
We consider the finite element solution using the space of continuous piecewise linear polynomials over the unit square $[0, 1]^2$ with a uniform triangulation similar to the one shown in Figure 5.1. The observation z_D of u_D is generated using a formula similar to (5.10). Let δ denotes the relative error (measured in the H^1 norm) in the measured data z_D .

Example 5.3.1. *We consider the problem with $u_D(x, y) = \sin(\pi x) \sin(\pi y)$, where f_D is chosen so that the exact parameter is*

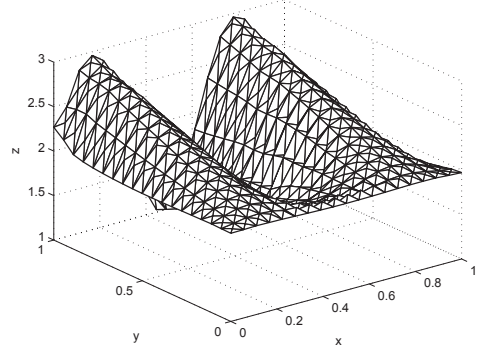
$$a_D(x, y) = 2 + y \sin(10x).$$

First we consider the case in which $\delta = 0$, so that z_D represents the interpolant of u_D . The computed errors are presented in Table 5.5. Next, we plot the exact and the recovered parameter with the presence of noise, results for various noise levels δ are shown in Figure 5.6.

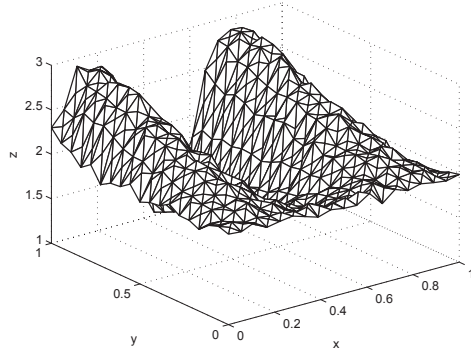
Table 5.5 shows a convergence rate of $O(h^{0.96})$ in the L^2 norm and of $O(h^{1.3})$ in the quotient norm. The plots in Figure 5.6 indicate how critical the noise on the quality of the recovered solution. However, even with noisy data, the recovered solution still resembles the exact parameter as it can be seen from the plots. In practice, the data should be smoothed or ‘denoised’ by an appropriate method then the equation error method might be used after. We shall consider some heuristic methods in Chapter 6.



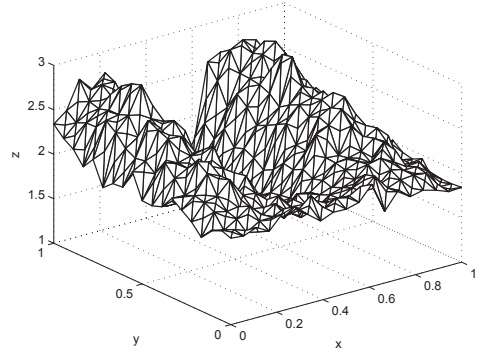
(a) Exact coefficient



(b) $\delta = 0, \alpha = 5 \cdot 10^{-5}$



(c) $\delta = 0.01, \alpha = 10^{-4}$



(d) $\delta = 0.05, \alpha = 5 \cdot 10^{-4}$

Figure 5.6: Recovered parameter in Example 5.3.1 with different noise and regularization parameters α .

Chapter 6

Heuristic results

The examples presented in Chapter 5 indicate a strong relationship between the noise level in the data and the quality of the computed solution by the equation error method. Because the equation error method is sensitive to noisy data (due to the term ∇z), one should consider smoothing before the equation error is applied. Moreover, in all the examples of Chapter 5, the values of regularization parameters were chosen to be nearly optimal (by trial and error), using the knowledge of the exact solution. But obviously without any a priori information about the exact coefficient, a good choice of the regularization parameter is vital in this type of analysis.

The purpose of this chapter is to give some heuristic results concerning data smoothing and parameter choice strategies. We present some numerical experiments to assess the proposed methods.

6.1 Parameter choice methods

In this section we present a popular heuristic method for choosing the regularization parameter in the equation error method, namely the **L-curve method**. This method is *purely a posteriori* or *error-free* parameter choice rule in the sense that it does not require a knowledge of the level of the noise in the data.

For simplicity, we will present this method in the context of Tikhonov regularization. We will adopt the notation of Section 2.1. Let x_{α,y^δ} denote the solution obtained by Tikhonov regularization; that is, x_{α,y^δ} is the solution of the minimization

problem

$$\min_{x \in X} \|Tx - y^\delta\|_Y^2 + \alpha \|x\|_X^2.$$

Intuitively, a desired value of α is the one that make a good compromise between the minimization of $\|Tx_{\alpha, y^\delta} - y^\delta\|_Y$ (data fitting) and keeping $\|x_{\alpha, y^\delta}\|_X$ small at the same time (enforcing stability).

The L-curve method is motivated by the following observation [18]. For small enough values of α , the residual $\|Tx_{\alpha, y^\delta} - y^\delta\|_Y$ varies a little while $\|x_{\alpha, y^\delta}\|_X$ blows up as α decreases. On the other hand, for moderate to large values of α , $\|x_{\alpha, y^\delta}\|_X$ varies by not too much as α decreases, but $\|Tx_{\alpha, y^\delta} - y^\delta\|_Y$ changes at a relatively faster rate. Thus, for a large range of values of α , the graph of $\|x_{\alpha, y^\delta}\|_X^2$ versus $\|Tx_{\alpha, y^\delta} - y^\delta\|_Y^2$ tends to exhibit an ‘L’ shaped curve, especially when plotted in a log-log scale. The L-curve method is to choose the value of α that corresponds to the corner of the L-curve. Here the corner is defined as the point of maximum curvature. If we set

$$f(\alpha) = \log \|Tx_{\alpha, y^\delta} - y^\delta\|_Y^2, \quad g(\alpha) = \log \|x_{\alpha, y^\delta}\|_X^2,$$

then the curvature is

$$\kappa(\alpha) = \frac{f''(\alpha)g'(\alpha) - f'(\alpha)g''(\alpha)}{(f'(\alpha)^2 + g'(\alpha)^2)^{3/2}}. \quad (6.1)$$

We refer to [18] and references therein for more justifications and discussions.

6.1.1 Numerical experiments

In this section we consider two experiments to assess L-curve for the equation error approach. The corner of the L-curve was determined visually (by trial and error) and using the curvature formula (6.1). We use $e_h = (a_{\beta, h} - a_N)\nabla u$ instead of $e_h = (a_{\beta, h} - a_N)\nabla z_N$ in the quotient norm defined in Chapter 5, similarly for the Dirichlet case.

Example 6.1.1. *In this experiment, we are trying to recover the coefficient a_N from noisy data z_N . The exact data is given by*

$$u_N(x, y) = x(1 - y), \quad \Omega = [0, 1] \times [0, 1].$$

The functions g_N and f_N are chosen so that the exact coefficient is

$$a_N(x, y) = \frac{9 \exp [-(6x - 3)^2 - (6y - 2)^2] (4 - 6x)^2 - \exp [-(6x - 2)^2 - (6y - 3)^2]}{3} \\ - \exp [-(6x - 3)^2 - (6y - 3)^2] (2(6x - 3) - 10(6x - 3)^3 - 10(6y - 3)^5) .$$

Table 6.1 shows the error results using optimal values of the regularization parameter. Here and throughout, optimal means the value of the parameter that gives least L^2 error. We repeat the experiment but the regularization parameter is chosen by the L -curve; error results are presented in Table 6.2. Figure 6.1 shows plots for the exact and recovered parameters for several noise levels. Plots for the L -curves are given in Figure 6.2.

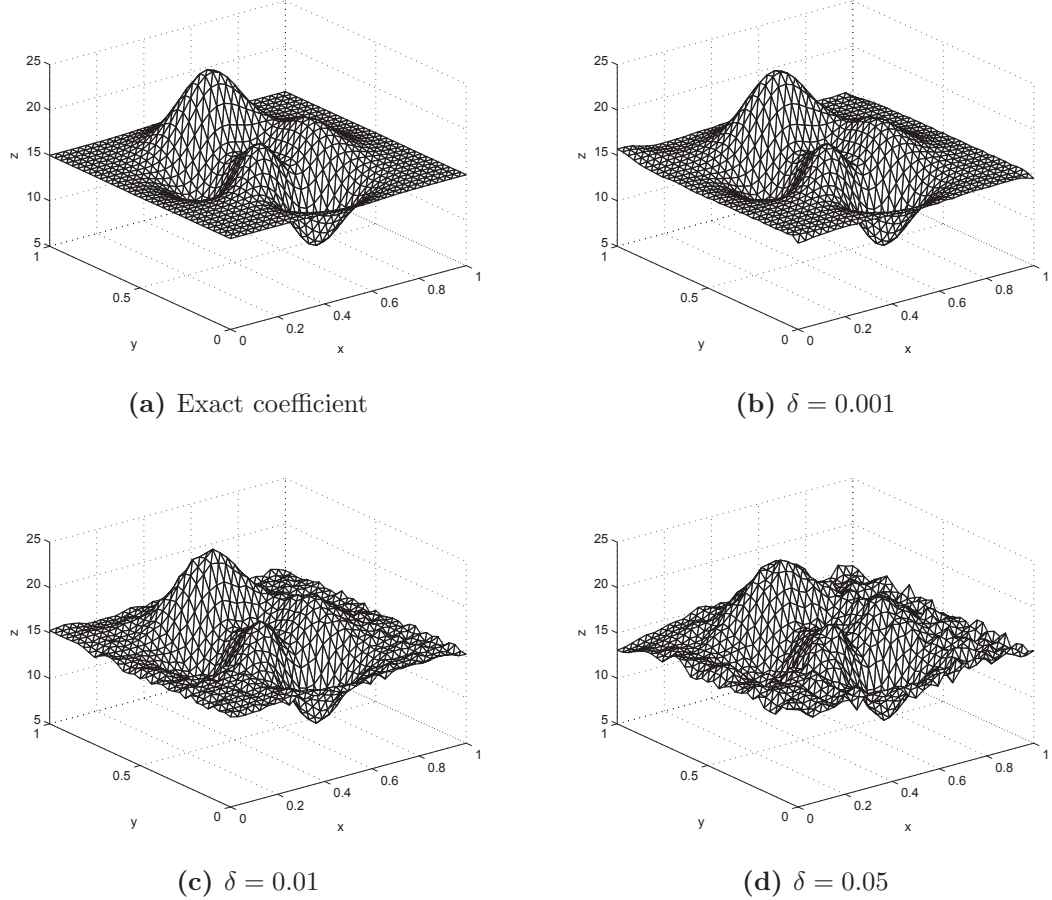


Figure 6.1: Recovered parameter in Example 6.1.1 using optimal β .

Table 6.1

Results of Example 6.1.1 for several noise levels using optimal choice of regularization parameter β .

δ	β	$\ a_{h,\beta} - a_N\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
0.001	$7 \cdot 10^{-6}$	$9.429 \cdot 10^{-2}$	$6.091 \cdot 10^{-3}$	$2.236 \cdot 10^{-2}$
0.005	$1 \cdot 10^{-5}$	$1.266 \cdot 10^{-1}$	$8.179 \cdot 10^{-3}$	$3.954 \cdot 10^{-2}$
0.010	$2 \cdot 10^{-5}$	$1.658 \cdot 10^{-1}$	$1.017 \cdot 10^{-2}$	$6.103 \cdot 10^{-2}$
0.050	$2 \cdot 10^{-4}$	$4.685 \cdot 10^{-1}$	$3.026 \cdot 10^{-2}$	$1.905 \cdot 10^{-1}$
0.100	$3 \cdot 10^{-4}$	1.042	$6.733 \cdot 10^{-2}$	$3.531 \cdot 10^{-1}$

Table 6.2

Results of Example 6.1.1 for several noise levels, regularization parameter β is computed using the L-curve method.

δ	β	$\ a_{h,\beta} - a_N\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
0.001	$3 \cdot 10^{-7}$	$1.531 \cdot 10^{-1}$	$9.891 \cdot 10^{-3}$	$2.595 \cdot 10^{-2}$
0.005	$6 \cdot 10^{-6}$	$1.379 \cdot 10^{-1}$	$8.911 \cdot 10^{-3}$	$4.312 \cdot 10^{-2}$
0.010	$1 \cdot 10^{-5}$	$1.886 \cdot 10^{-1}$	$1.218 \cdot 10^{-2}$	$7.058 \cdot 10^{-2}$
0.050	$2 \cdot 10^{-3}$	1.138	$7.357 \cdot 10^{-2}$	$4.723 \cdot 10^{-1}$
0.100	$5 \cdot 10^{-3}$	1.549	$1.000 \cdot 10^{-1}$	$6.949 \cdot 10^{-1}$

Example 6.1.2. In this experiment, we are trying to recover the coefficient a_D from noisy data z_D . The exact data is given by

$$u_D(x, y) = x(x - 1) \sin(\pi y), \quad \Omega = [0, 1] \times [0, 1].$$

The function f_D are chosen so that the exact coefficient is

$$a_D(x, y) = \begin{cases} 1, & 0 \leq x < 0.4, \\ 10x - 3, & 0.4 \leq x < 0.6, \\ 3, & 0.6 \leq x \leq 1. \end{cases}$$

Table 6.3 shows the error results using optimal values of the regularization parameter.

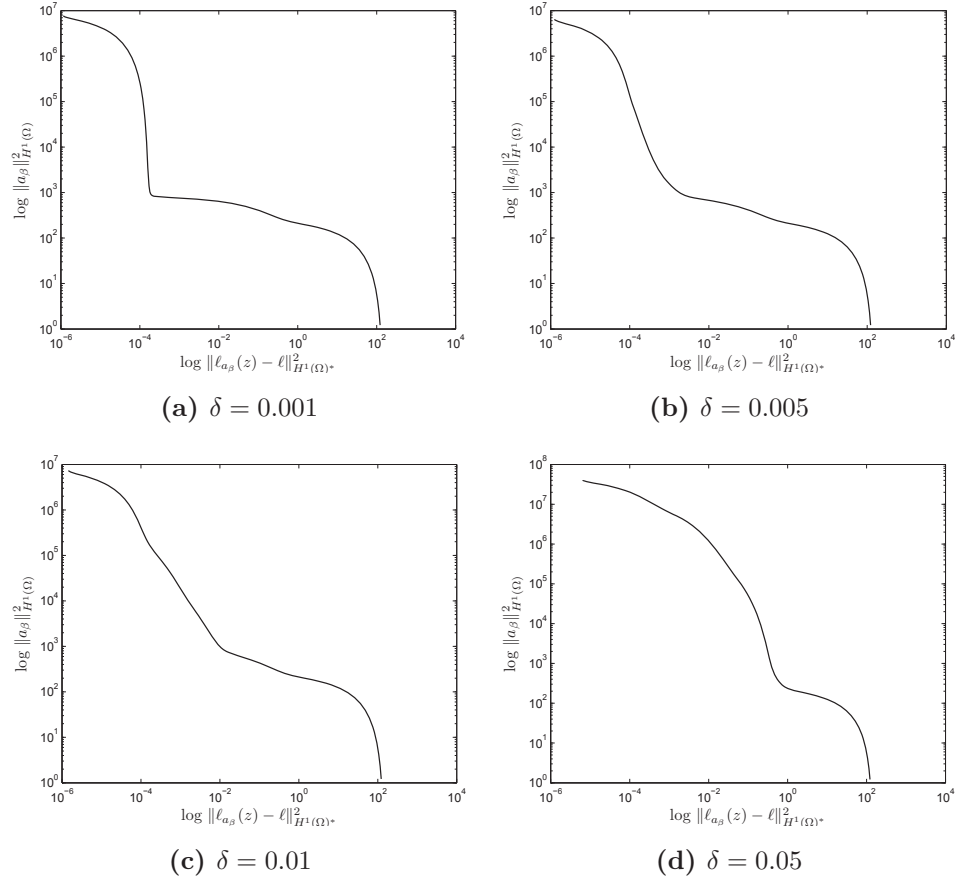


Figure 6.2: L-curves for Example 6.1.1 with different noise levels.

We repeat the experiment but the regularization parameter is chosen by the L-curve; error results are presented in Table 6.4. Figure 6.4 shows plots for the exact and recovered parameters for several noise levels. Plots for the L-curves are given in Figure 6.3.

Remark 6.1.1. Although Figure 6.3 (a) does not exhibit the characteristic L shape, there is a local maximum of the curvature in the region indicated by the circle. We choose the corresponding value of α as the regularization parameter.

The above experiments indicates that L-curve method gives reasonable results as a parameter choice strategy. However, there are still unanswered questions about the L-curve; for example, does always the L-curve have a ‘corner’? We quote “In spite of its use in several applications, there still lacks a sound mathematical foundation of the L-curve method.” [18, page 111].

Table 6.3

Results of Example 6.1.2 for several noise levels using optimal choice of regularization parameter α .

δ	α	$\ a_{h,\alpha} - a_D\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
0.001	$4 \cdot 10^{-7}$	$1.415 \cdot 10^{-2}$	$6.417 \cdot 10^{-3}$	$4.241 \cdot 10^{-3}$
0.005	$3 \cdot 10^{-6}$	$2.567 \cdot 10^{-2}$	$1.163 \cdot 10^{-2}$	$7.011 \cdot 10^{-3}$
0.010	$6 \cdot 10^{-6}$	$3.536 \cdot 10^{-2}$	$1.603 \cdot 10^{-2}$	$1.006 \cdot 10^{-2}$
0.050	$3 \cdot 10^{-5}$	$8.979 \cdot 10^{-2}$	$4.070 \cdot 10^{-2}$	$2.890 \cdot 10^{-2}$
0.100	$7 \cdot 10^{-5}$	$1.585 \cdot 10^{-1}$	$7.184 \cdot 10^{-2}$	$5.211 \cdot 10^{-2}$

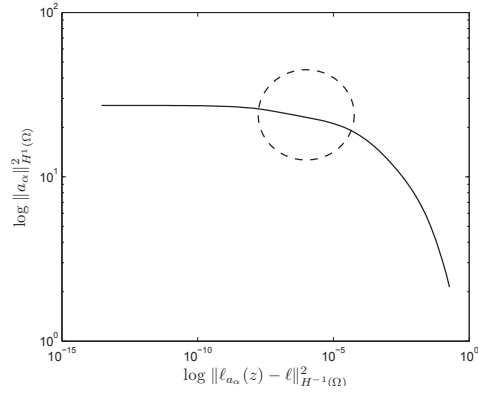
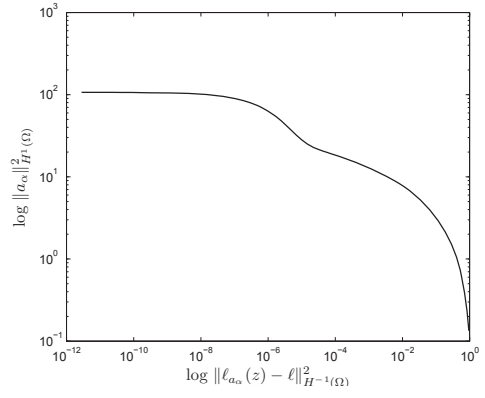
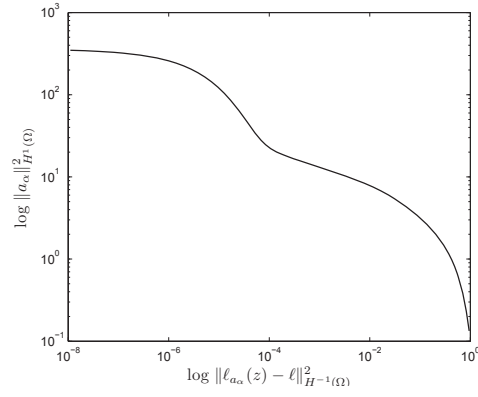
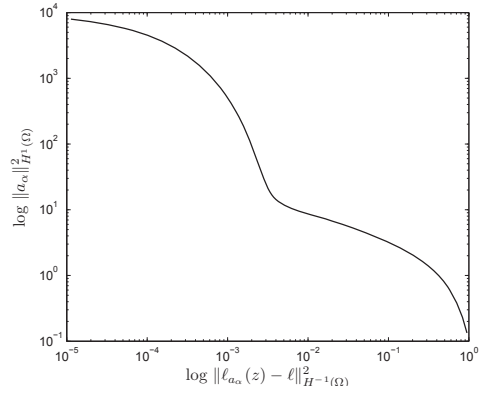
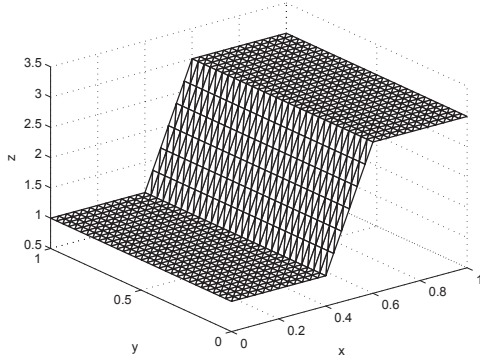
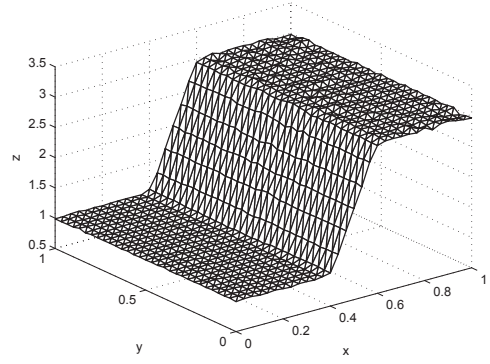
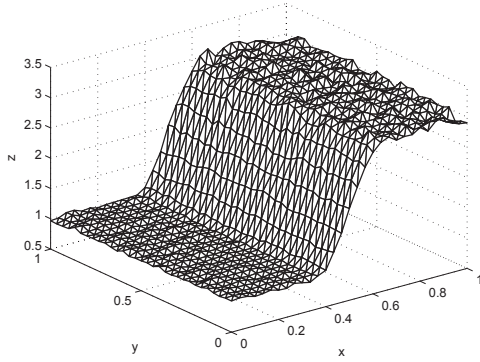
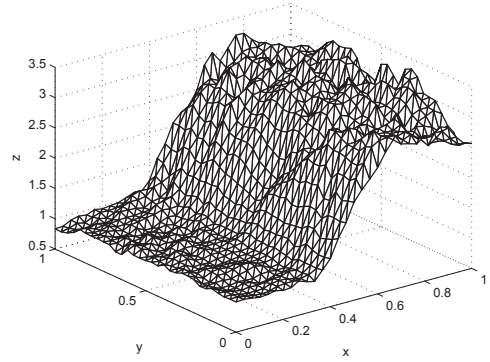
**(a)** $\delta = 0.001$ **(b)** $\delta = 0.005$ **(c)** $\delta = 0.01$ **(d)** $\delta = 0.05$ **Figure 6.3:** L-curves for Example 6.1.2 with different noise levels.

Table 6.4

Results of Example 6.1.2 for several noise levels. The regularization parameter α is chosen using the L-curve method.

δ	α	$\ a_{h,\alpha} - a_D\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
0.001	$2 \cdot 10^{-7}$	$1.451 \cdot 10^{-2}$	$6.577 \cdot 10^{-3}$	$4.130 \cdot 10^{-3}$
0.005	$2 \cdot 10^{-6}$	$2.608 \cdot 10^{-2}$	$1.182 \cdot 10^{-2}$	$6.715 \cdot 10^{-3}$
0.010	$8 \cdot 10^{-6}$	$3.648 \cdot 10^{-2}$	$1.653 \cdot 10^{-2}$	$1.065 \cdot 10^{-2}$
0.050	$2 \cdot 10^{-4}$	$1.491 \cdot 10^{-1}$	$6.761 \cdot 10^{-2}$	$5.041 \cdot 10^{-2}$
0.100	$7 \cdot 10^{-4}$	$1.782 \cdot 10^{-1}$	$8.079 \cdot 10^{-2}$	$6.150 \cdot 10^{-2}$

**(a)** Exact coefficient**(b)** $\delta = 0.001$ **(c)** $\delta = 0.01$ **(d)** $\delta = 0.1$ **Figure 6.4:** Recovered parameter in Example 6.1.2 using optimal α .

6.2 Data smoothing

The aim of data smoothing is to remove or reduce the noise from a data set to get a better estimate to the true (exact) data. One hopes that the smoothed data still captures important patterns and features in the true data, so the data is not over-smoothed, but unwanted rapidly changing points and fine scale structures are ignored, so that the data is not under-smoothed. Consequently, by appropriately smoothing the data, we stabilize algorithms that are sensitive to noise, such as the equation error method.

The sensitivity of the equation error method comes from the need to differentiate the noisy data. Differentiation is actually an ill-posed process as one can demonstrate easily. Consider a function $y \in C[0, 1]$, and let

$$y_n(t) = y(t) + 1/\sqrt{n} \sin(n\pi t).$$

Then

$$\|y_n - y\|_{C[0,1]} = 1/\sqrt{n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus y_n converges to y uniformly on $[0, 1]$. However, this is not the case for the derivatives:

$$\|y'_n - y'\|_{C[0,1]} = \pi \sqrt{n} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

The above results still hold even if we use the L^2 norm instead:

$$\|y_n - y\|_{L^2(0,1)} = 1/\sqrt{2n} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

$$\|y'_n - y'\|_{L^2(0,1)} = \pi \sqrt{n/2} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Since in practice only noisy data is available, we expect numerical instability in the equation error method due to the need to differentiate the data in its computations. Thus, the aim of data smoothing here is to reduce harmful variations in the given noisy data but at the same time keeping most important information relevant to the true data.

In this section we present two data smoothing techniques which can be used in conjunction with the equation error method. The first of which is the *cubic smoothing spline*, and the second is smoothing by the *Laplacian operator*.

We start with smoothing by the Laplace operator (**SLO**). The motivation of this technique comes from the fact that any solution w of the Laplace equation

$$-\Delta w = f$$

is usually much smoother than the function f . Suppose u is a given function that vanishes on the boundary of Ω and z is a given (possibly noisy) observation of u . Then, the SLO method is to solve the BVP

$$\begin{aligned} -\nu \Delta u_s + u_s &= z \text{ in } \Omega, \\ u_s &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{6.2}$$

for u_s , where $\nu \geq 0$. Now, if $\nu = 0$, then $u_s = z$ and we have done nothing regarding smoothing. However, if we choose $\nu > 0$, then we hope that u_s is smoother than z . Now the question is how to choose a good value of ν ? On the one hand, if ν is too large, then $\|u_s\|_{H^1(\Omega)}$ is too small relative to $\|z\|_{H^1(\Omega)}$ (this can be shown using Fourier analysis), and so u_s is very smooth compared to z . Thus, in this case, u_s may fail to capture important features such as the curvature and pointwise values of u . On the other hand, if ν is too small, then $\|u_s\|_{H^1(\Omega)} \sim \|z\|_{H^1(\Omega)}$, and so u_s is probably still as rough as z . From this discussion, we are motivated to choose largest $\nu > 0$ such that

$$\|u_s - z\|_{H^1(\Omega)} \sim \|u - z\|_{H^1(\Omega)}.$$

In practice, we usually do not know $\|u - z\|_{H^1(\Omega)}$ but an upper bound, say ϵ . Thus, the SLO method is to choose u_s that corresponds to the largest value of ν such that

$$\|u_s - z\|_{H^1(\Omega)} \sim \epsilon.$$

It is not clear how to generalize the SLO for data that satisfies Neumann boundary conditions. However, we expect to have good smoothing properties at least inside Ω but poor estimates near the boundary.

We consider the following numerical experiment.

Example 6.2.1. Suppose the exact data is given by the function

$$u(t) = t(1 - t), \quad 0 \leq t \leq 1,$$

and the noisy data is given by

$$z(t) = u(t) + \frac{1}{70} \sin(30\pi t) u(t).$$

Figure 6.5 shows a plot for u and z on the same axes. We smooth the noisy data z using the SLO; plot for smoothed data u_s versus the exact data u is shown in Figure 6.6. Next we plot derivatives u' , z' , and u'_s ; see Figure 6.7. Finally, quantitatively we have the following errors in both the L^2 and H^1 norms:

$$\|u - z\|_{L^2(0,1)} = 0.001844, \quad \|u - u_s\|_{L^2(0,1)} = 0.005485,$$

$$\|u - z\|_{H^1(0,1)} = 0.173927, \quad \|u - u_s\|_{H^1(0,1)} = 0.0269627.$$

From these statistics, we see that slight increase in the relative L^2 error is compensated with a noticeable reduction in the H^1 relative error. This is important especially when a good approximation of the derivative is desired as in the case of the equation error method.

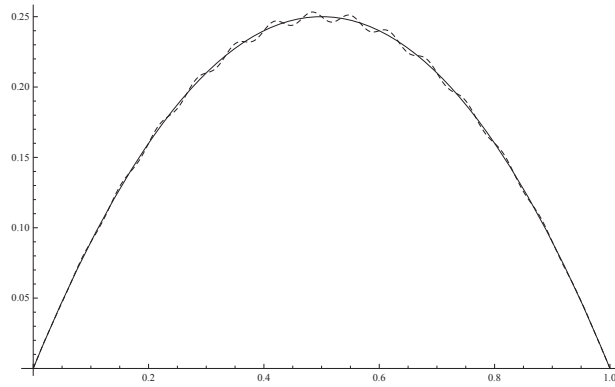


Figure 6.5: Exact data u (solid), noisy data z (dashed).

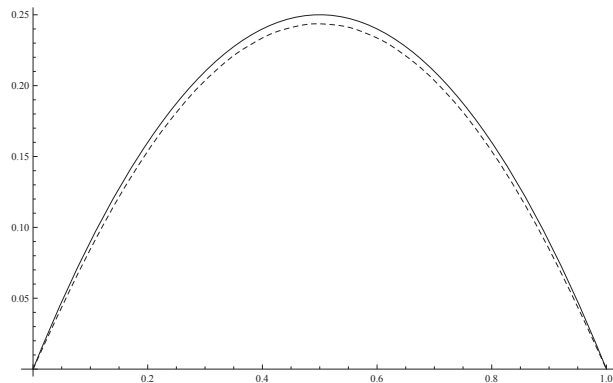


Figure 6.6: Exact data u (solid), smoothed data u_s by the SLO (dashed).

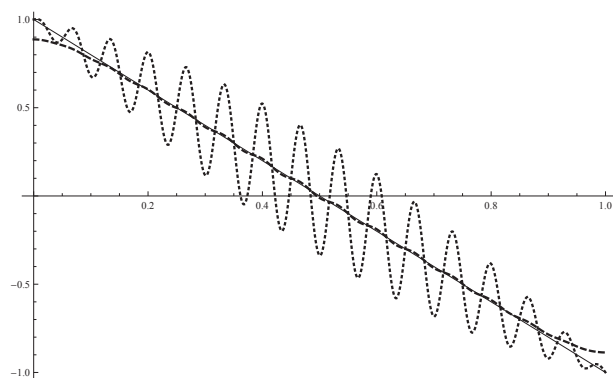


Figure 6.7: u' (solid), u'_s (dashed), and z' (dotted).

Another strategy to smooth noisy data is using smoothing splines. In particular, we will consider the *cubic smoothing spline*. For a given data set

$$\{(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)\}$$

with $x_0 < x_1 < \dots < x_n$, and a specified tolerance **Tol**, the method of smoothing spline can be formulated as: find a cubic spline function¹ f that minimizes

$$\int_{x_0}^{x_n} |f''(x)|^2 dx \quad \text{subject to} \quad \sum_{i=0}^n |f(x_i) - y_i|^2 \leq \text{Tol}.$$

Thus for small **Tol**, f resembles the spline interpolant of the data set. However, for large **Tol**, then f is a very smooth spline function and so is almost flat. Therefore,

¹That is, a twice continuously differentiable piecewise polynomial of degree three.

choosing the appropriate value of `Tol` for a given problem is critical, as we discussed for the case of choosing ν in the SLO method. We point out that the above formulation of the smoothing spline can be written in several alternative variations, but we shall not discuss them here. Furthermore, cubic smoothing spline can be generalized for multivariable functions. There are some routines included in **Matlab** Curve Fitting Toolbox which we will use in the numerical examples in the next subsection.

6.2.1 Numerical experiments

In this section, we apply the smoothing techniques presented in the last section to assess their performance when they are used in conjunction with the equation error method.

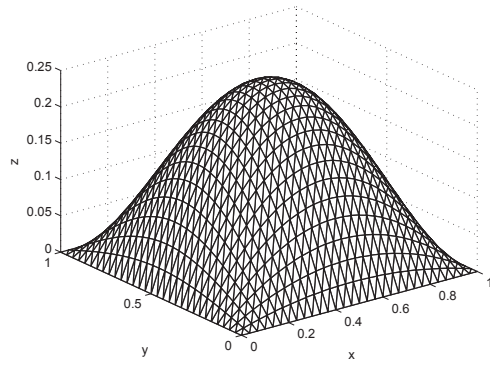
Example 6.2.2. *In the first test, we revisit Example 6.1.2. We use the SLO method to smooth out the data before applying the equation error method. The smoothing parameter ν is chosen such that*

$$\|u_s - z_D\|_{H^1(\Omega)} \sim \|u_D - z_D\|_{H^1(\Omega)},$$

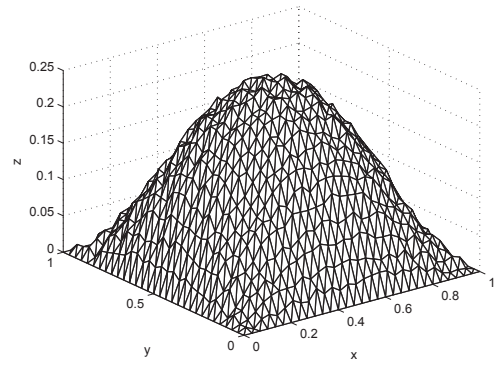
where here u_s is the smoothed data using the SLO method. Figure 6.8 shows the plots for the exact data, noisy data, and the denoised data when the noise level in the data is $\delta = 0.5$. Next we use the denoised data to recover the parameter a_D using the equation error method; results are presented in Table 6.5.

Table 6.5
Results of Example 6.2.2 using denoised by the SLO method.

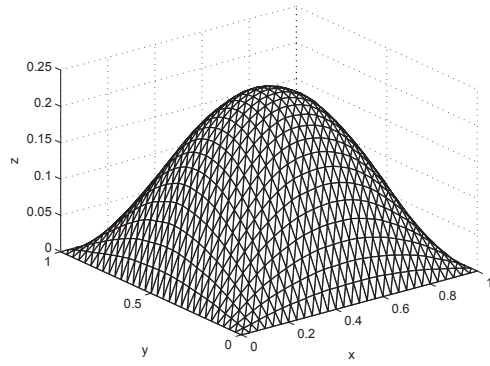
δ	α	$\ a_{h,\alpha} - a_D\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
0.001	$2 \cdot 10^{-7}$	$1.284 \cdot 10^{-2}$	$5.824 \cdot 10^{-3}$	$3.740 \cdot 10^{-3}$
0.005	$1 \cdot 10^{-6}$	$1.903 \cdot 10^{-2}$	$8.629 \cdot 10^{-3}$	$6.139 \cdot 10^{-3}$
0.010	$2 \cdot 10^{-6}$	$2.469 \cdot 10^{-2}$	$1.119 \cdot 10^{-2}$	$9.292 \cdot 10^{-3}$
0.050	$6 \cdot 10^{-6}$	$5.591 \cdot 10^{-2}$	$2.534 \cdot 10^{-2}$	$2.808 \cdot 10^{-2}$
0.100	$1 \cdot 10^{-4}$	$7.482 \cdot 10^{-2}$	$3.391 \cdot 10^{-2}$	$3.592 \cdot 10^{-2}$



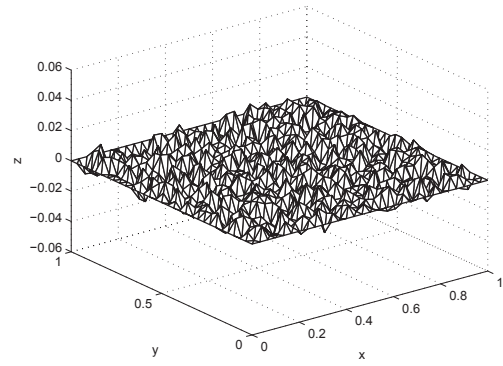
(a) Exact data u_D



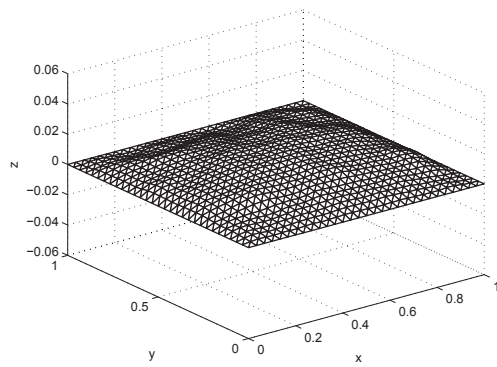
(b) Noisy data z_D



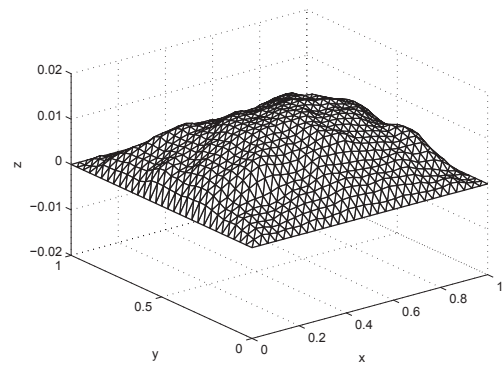
(c) Smoothed data u_s



(d) Error before smoothing



(e) Error after smoothing



(f) Error after smoothing (zoomed)

Figure 6.8: Smoothing the data of Example 6.2.2 with noise level $\delta = 0.5$.

The results in Table 6.3 and Table 6.5 reveal a noticeable improvement in recovering the parameter a_D before and after smoothing the data by the SLO method. We point out that the regularization parameter is chosen to be optimal, as in the case of nonsmoothed data.

In the next example we revisit Example 6.1.1; here we smooth the data using the cubic smoothing spline. We should point out that the noisy data is first generated by adding uniformly distributed noise, and then the experiment is repeated by adding Gaussian noise of mean zero and standard deviation one; the results are almost identical in both tests (under the same noise level).

Example 6.2.3. *In this test, we revisit Example 6.1.1. We use the cubic smoothing spline to smooth out the data before applying the equation error method. The parameter Tol is chosen so that*

$$\sum_{i=1}^n |u(x_i) - z_N(x_i)|^2 \sim \text{Tol}$$

where $\{x_i\}$ are the nodes in the finite element triangulation of the domain $[0, 1]^2$. Figure 6.10 shows the plots for the exact data, noisy data, and the denoised data when the noise level in the data is $\delta = 0.5$. Next, we use the denoised data to recover the parameter a_N using the equation error method; results are presented in Table 6.6. In Figure 6.9 we plot the recovered parameter before and after smoothing the data for relatively large noise levels.

Table 6.6

Results of Example 6.2.3 using denoised data by cubic smoothing spline.

δ	β	$\ a_{h,\beta} - a_N\ _{L^2(\Omega)}$	L^2 Relative Error	$\ \dot{e}_h\ _Q$
0.001	$3 \cdot 10^{-6}$	$4.561 \cdot 10^{-2}$	$2.946 \cdot 10^{-3}$	$1.058 \cdot 10^{-2}$
0.005	$3 \cdot 10^{-6}$	$4.564 \cdot 10^{-2}$	$2.948 \cdot 10^{-3}$	$1.056 \cdot 10^{-2}$
0.010	$3 \cdot 10^{-6}$	$4.568 \cdot 10^{-2}$	$2.951 \cdot 10^{-3}$	$1.054 \cdot 10^{-2}$
0.050	$3 \cdot 10^{-6}$	$4.603 \cdot 10^{-2}$	$2.973 \cdot 10^{-3}$	$1.041 \cdot 10^{-2}$
0.100	$3 \cdot 10^{-6}$	$4.651 \cdot 10^{-2}$	$3.004 \cdot 10^{-3}$	$1.032 \cdot 10^{-2}$

The error results in Table 6.6 shows a remarkable improvement in the quality of

the recovered parameter after denoising the data (Table 6.6 should be compared with Table 6.1). Similarly, the plots in Figure 6.9 indicates a better stability behavior for the equation error method even with relatively large noise levels in the original data.

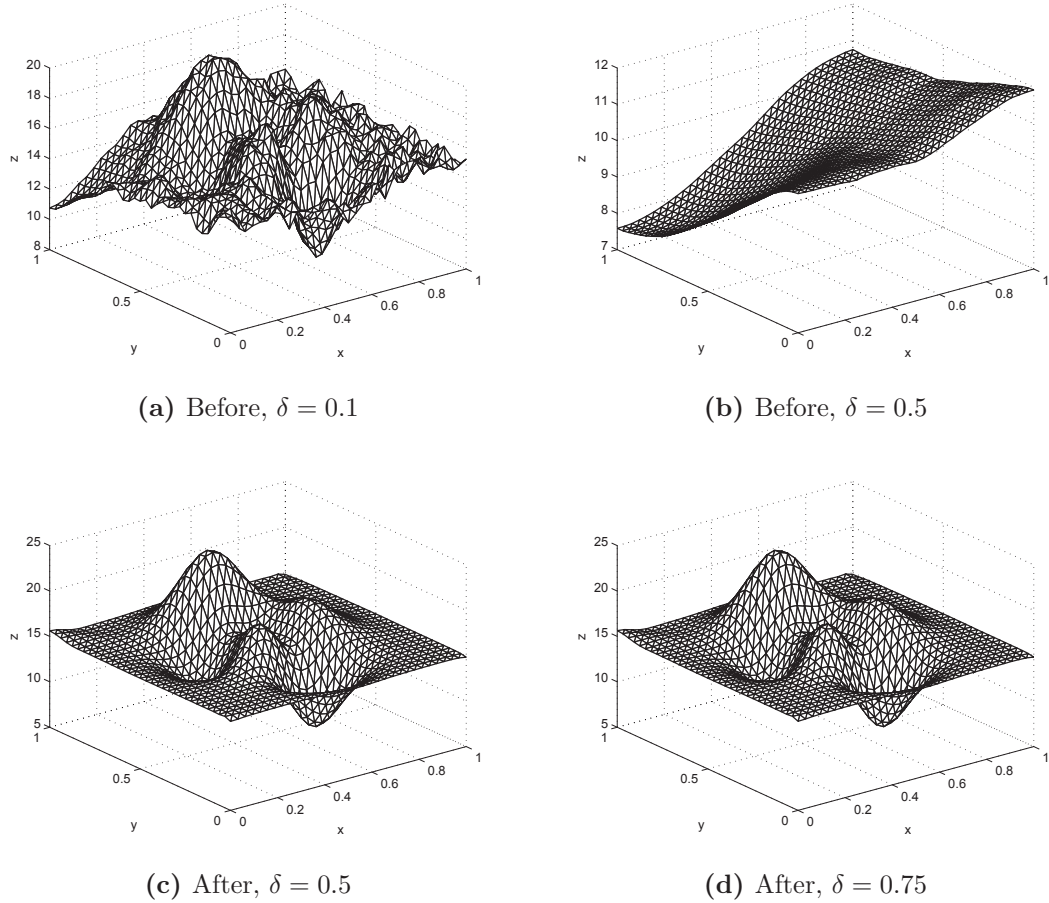
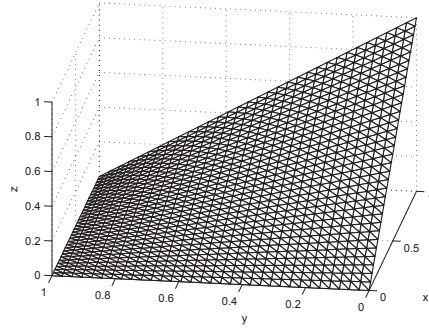


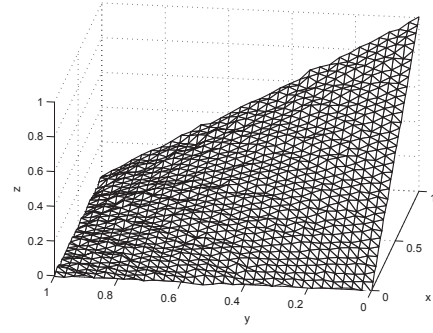
Figure 6.9: Recovered parameter in Example 6.2.3 before smoothing the data (a) and (b), and after smoothing the data using cubic smoothing spline (c) and (d).

From the last examples, we see that smoothing can significantly improve the quality of the recovered parameter by the equation error method. These observations are heuristic but worth pursuing in a more rigorous mathematical analysis.

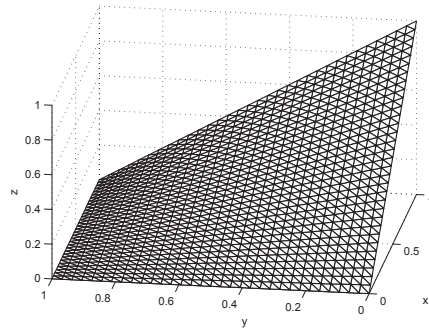
Some numerical and theoretical results, especially in the case of the recovery of constant coefficients, indicates that the OLS (output least squares) method is more stable and robust with respect to noisy data than the equation error method; see [43].



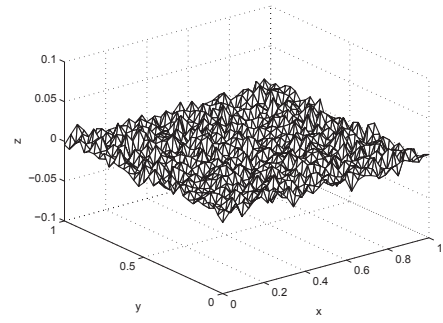
(a) Exact data u_N



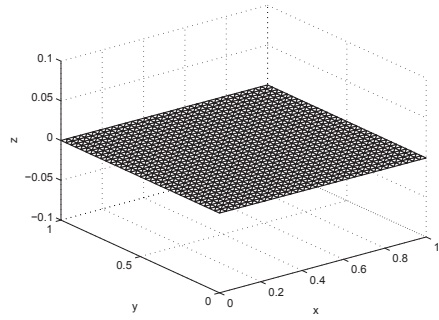
(b) Noisy data z_N



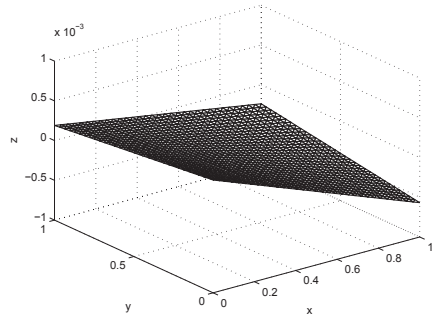
(c) Smoothed data u_s



(d) Error before smoothing



(e) Error after smoothing



(f) Error after smoothing (zoomed)

Figure 6.10: Smoothing the data of Example 6.2.3 with noise level $\delta = 0.5$ using cubic smoothing spline.

However, the OLS is computationally expensive since it requires numerically solving nonlinear and nonconvex optimization problem. As in most nonlinear optimization problems, a good initial estimate of the minimizer is vital. It can improve the speed and might prevent the algorithm from getting stuck at a local minimum or at a

‘flat valley’ (due to small change in the gradient that may cause the optimization algorithm to terminate at such point). Since the OLS is expected to perform better, it may be a good idea to use the equation error method to provide an initial guess for OLS iterations. We investigate this in the following experiment.

Example 6.2.4. *In this numerical experiment, we consider the recovery of the coefficient a_D where*

$$u_D(x, y) = (1 - x)x \sin(2\pi y), \quad \Omega = [0, 1]^2.$$

The load function f_D is chosen so that the true coefficient is

$$a_D(x, y) = 2 + x(1 + y^2).$$

The purpose in this experiment is to compare the speed of the OLS with and without initialization by the equation error method. We use a line search with descent directions obtained by a quasi-Newton method. The results in Table 6.7 compare the time and number of iterations until the OLS converges to a possible minimum. For the equation error method, the value of the regularization parameter is fixed to $\alpha = 0.001$. When the equation error is not used, the initial guess for the line search is taken to be the vector with all entries one. We use uniform triangulation as in Figure 6.11 consisting of $(n + 1)^2$ nodes. The following notations will be used:

- T_E : *time (in seconds) until the convergence of the OLS when initialized using the equation error method. This time also includes the time needed by the equation error method to produce the initial guess;*
- T_O : *time (in seconds) until the convergence of the OLS without using the equation error method;*
- NI_E : *Number of quasi-Newton iterations until convergence using the equation error method;*
- NI_O : *Number of quasi-Newton iterations until convergence without using the equation error method.*

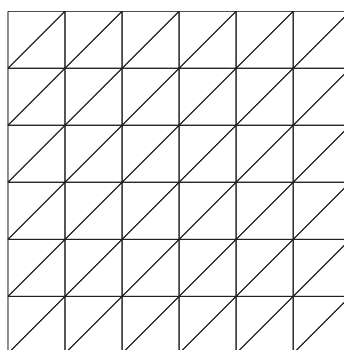
It is evident from the results in the Table 6.7 that the equation error method can dramatically enhance the robustness and speed of the OLS method. Furthermore,

Table 6.7

Comparing the times and number of iterations for the OLS with and without initialization from the equation error method. Asterisked numbers indicate failure to converge to the true parameter.

n	δ	T_E	T_O	NI_E	NI_O
10	0.001	3.308	6.24	4	10
	0.010	3.354	6.24	4	10
20	0.001	22.074	48.656	4	11
	0.010	22.339	49.203	4	11
30	0.001	95.238	412.870*	3	16
	0.010	118.310	440.568*	4	17

initializing the OLS using the equation error may prevent the OLS from the convergence to a false (nonstationary) minimizer as the one in Figure 6.12. For large-scale problems, it becomes critical to initialize the OLS by a good starting point; the equation error method is an excellent option since it is fast and produces reasonable estimates.

**Figure 6.11:** Triangulation of Ω .

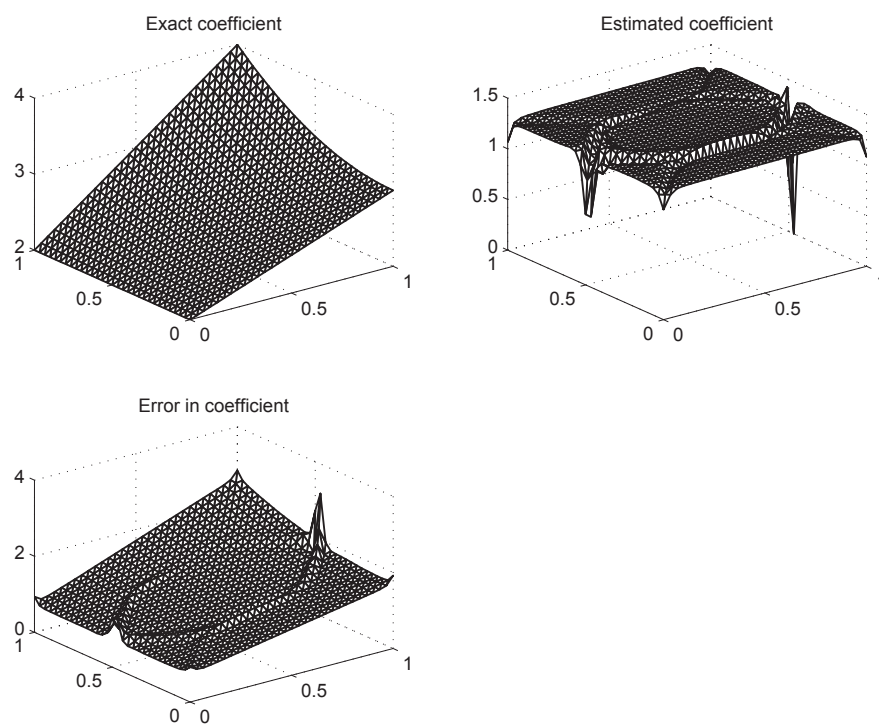


Figure 6.12: Recovered coefficient by the OLS, noise level $\delta = 0.001$.

Chapter 7

Conclusion

The last few decades have witnessed a remarkable growth in the number of publications on the theory of inverse and ill-posed problems. This is mainly due to the increasing number of applications in science and engineering, including physical, biomedical, and geophysical applications. In particular, parameter identification problems have gained extra attention, especially inverse elasticity problems, due to a promising biomedical applications.

The methods of output least-squares and equation error (and their variants) are among the most prominent methods in solving elliptic inverse problems. The output least-squares is very natural and its theoretical development is noticeably more advanced than that of the equation error method. However, thus far, there is a lack of theoretical results regarding convergence and error estimates for the equation error method.

In this work, we proved the convergence of the equation error method and derived error bounds on the computed solutions. We presented the existing theory in a way that is both more understandable and more amenable to generalizations. Though the equation error method looks less natural than the other methods, it is simple to implement in a practical algorithm, which is also less expensive than most of the existing algorithms. However, this simplicity comes with some flaws as our numerical experiments reveal. The main drawback in the equation error method is its sensitivity to noisy data. We proposed a partial remedy to these pitfalls by data smoothing using the Laplace operator or cubic smoothing splines, and the L-curve method as a parameter choice strategy. Our numerical experiments indicate noteworthy results.

However these methods are still heuristic and lack a sound mathematical foundation. For large-scale problems and when the results of the equation error method become in doubt, the output least-squares may serve as an alternative method due its better robustness properties (though, for nonconstant coefficients, this might be just heuristic). In this case, the equation error method can be used to initialize the output least-squares; our numerical experiments show a significant increase in the speed.

We hope to generalize our results for higher-dimensional problems (3D, in particular), and to other elliptic inverse problems such as the general linear elasticity equations (not necessarily isotropic). Furthermore, we hope to come up with *a posteriori* parameter choice rule that is appropriate for the equation error method. The proposed data smoothing techniques are worth pursuing and they will be the subject of the sequel to this work.

Bibliography

- [1] Bear J. Dynamics of Fluids in Porous Media. 1st ed. New York(NY): American Elsevier; 1972.
- [2] Bruckner G, Handrock-Meyer S, Langmach H. An inverse problem from 2D ground-water modelling. *Inverse Problems*. 1998;14(4):835–851.
- [3] Knowles I. Parameter identification for elliptic problems. *Journal of Computational and Applied Mathematics*. 2001;131(1-2):175–194.
- [4] Sun N. Inverse Problems in Groundwater Modeling. 1st ed. Dordrecht: Kluwer Academic; 1999.
- [5] Evans LC. Partial Differential Equations. Providence(RI): American Mathematical Society; 2010.
- [6] Mcowen R. Partial Differential Equations: Methods and Applications. 2nd ed. Upper Saddle River(NJ): Prentice Hall; 2003.
- [7] Duvaut G, Lions J-L. Inequalities in Mechanics and Physics. 1st ed. Berlin: Springer; 1976.
- [8] Kärkkäinen T. An equation error method to recover diffusion from distributed observation. *Inverse Problems*. 1997;13(4):1033–1051.
- [9] Chen J, Gockenbach MS. A variational method for recovering planar Lamé moduli. *Mathematics and Mechanics of Solids*. 2002;7(2):191–202.
- [10] Cox SJ, Gockenbach MS. Recovering planar Lamé moduli from a single traction experiment. *Mathematics and Mechanics of Solids*. 1997;2(3):297–306.

- [11] Gockenbach MS. The output least-squares approach to estimating Lamé moduli. *Inverse Problems*. 2007;23(6):2437–2455.
- [12] Gockenbach MS, Jadamba B, Khan AA. Equation error approach for elliptic inverse problems with an application to the identification of Lamé parameters. *Inverse Problems in Science and Engineering*. 2008;16(3):349–367.
- [13] Ji L, McLaughlin J. Recovery of the Lamé parameter μ in biological tissues. *Inverse Problems*. 2004;20(1):1–24.
- [14] Oberai AA, Gokhale NH, Feijóo GR. Solution of inverse problems in elasticity imaging using the adjoint method. *Inverse Problems*. 2003;19(2):297–313.
- [15] Kohn RV, Lowe B. A variational method for parameter identification. *Mathematical Modelling and Numerical Analysis*. 1988;22(1):119–158.
- [16] Alessandrini G. On the identification of the leading coefficient of an elliptic equation. *Analisi Funzionale e Applicazioni*. 1985;4(1):87–111.
- [17] Acar R. Identification of the coefficient in elliptic equations. *SIAM Journal on Control and Optimization*. 1993;31(5):1221–1244.
- [18] Engl HW, Hanke M, Neubauer A. *Regularization of Inverse Problems*. 1st ed. Dordrecht: Kluwer Academic Publishers; 1996.
- [19] Groetsch C. *Inverse Problems in the Mathematical Sciences*. 1st ed. Braunschweig: Friedr. Vieweg & Sohn; 1993.
- [20] Kirsch A. *An Introduction to the Mathematical Theory of Inverse Problems*. 1st ed. New York(NY): Springer-Verlag; 2011.
- [21] Nair MT. *Linear Operator Equations: Approximation and Regularization*. 1st ed. Singapore: World Scientific; 2009.
- [22] Hadamard J. *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. 1st ed. New Haven(CT): Yale University Press; 1923.
- [23] Bartle RG. *A Modern Theory of Integration*. 1st ed. Providence(RI): American Mathematical Society; 2001.

- [24] Falk R, Error estimates for the numerical identification of a variable coefficient. *Mathematics of Computation*. 1983;40(162):537–546.
- [25] Richter G. An inverse problem for the steady state diffusion equation. *SIAM Journal on Applied Mathematics*. 1981;41(2):210–221.
- [26] Zou S. Numerical methods for elliptic inverse problems. *International Journal of Computer Mathematics*. 1998;70(2):211–232.
- [27] Hào DN, Quyen NT. Convergence rates for Tikhonov regularization of coefficient identification problems in Laplace-type equations. *Inverse Problems*. 2010;26(12):125014–23.
- [28] Atkinson K, Weimin H. *Theoretical Numerical Analysis: A Functional Analysis Framework*, 3rd ed. New York(NY): Springer; 2009.
- [29] Kreyszig E. *Introductory Functional Analysis with Applications*. 1st ed. New York(NY): Wiley; 1978.
- [30] Lebedev L, Vorovich I, Gladwell G. *Functional Analysis: Applications in Mechanics and Inverse Problems*. 1st ed. Dordrecht: Kluwer Academic; 2002.
- [31] Megginson RE. *An Introduction to Banach Space Theory*. 1st ed. New York(NY): Springer-Verlag; 1998.
- [32] Adams R. *Sobolev Spaces*. 1st ed. New York(NY): Academic Press; 1975.
- [33] Brenner S, Scott L. *The Mathematical Theory of Finite Element Methods*. 3rd ed. New York(NY): Springer; 2003.
- [34] Dautray R, Lions J-L. *Mathematical Analysis and Numerical Methods for Science and Technology*. vol 3. Berlin: Springer; 1990.
- [35] Girault V, Raviart P. *Finite Element Methods for Navier-Stokes Equations*. 1st ed. Berlin: Springer; 1986.
- [36] Grisvard P. *Elliptic Problems in Nonsmooth Domains*. 1st ed. Boston(MA): Pitman; 1985.

- [37] Nečas J. Direct Methods in the Theory of Elliptic Equations. 1st ed. Berlin: Springer; 2012.
- [38] Ciarlet PG. The Finite Element Method for Elliptic Problems. Philadelphia(PA): SIAM; 2002.
- [39] Gockenbach MS. Understanding and Implementing the Finite Element Method. 1st ed. Philadelphia(PA): SIAM; 2006.
- [40] Steinbach O. Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements. 1st ed. New York(NY): Springer; 2008.
- [41] Braess D. Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics. 3rd ed. Cambridge: Cambridge University Press; 2007.
- [42] Al-Jamal MF, Gockenbach MS. Stability and error estimates for an equation error method for elliptic equations. Inverse problems. 2012;28(9):095006–24.
- [43] Andreev A. Comparison of Numerical Methods for Elliptic Inverse Problems. Ann Arbor(MI): ProQuest LLC; 2006. Thesis (Ph.D.)–Michigan Technological University.
- [44] Gohberg I, Goldberg S. Basic Operator Theory. 1st ed. Cambridge(MA): Birkhäuser Boston; 1980.

Appendix A

A.1 Spectral theory of compact operators

In this appendix, we give a brief introduction to the spectral theory of compact operators. Compact operators enjoy interesting features and studying their spectral properties is fairly easy. The emphasis of this discussion will be on topics relevant to the theory of ill-posed and inverse problems. Most of the results in this presentation can be found in [18, 19, 21, 29, 44].

We start with following result and definition from the functional analysis.

Theorem A.1.1. *Let $T : X \rightarrow Y$ be a bounded linear operator, where X and Y are Hilbert spaces. Then there exists a unique operator $T^* : Y \rightarrow X$ such that*

$$(Tx, y)_Y = (x, T^*y)_X \quad \forall x \in X, y \in Y.$$

The operator T^ is called the (Hilbert) **adjoint** of T . Further, T^* is linear and bounded with $\|T^*\| = \|T\|$. Moreover, $T^{**} = (T^*)^* = T$.*

Definition A.1.1. *Let X be a Hilbert space. A bounded linear operator $T : X \rightarrow X$ is called **self-adjoint** if $T^* = T$.*

One can define eigenvalues and eigenvectors for operators exactly in the same way they are defined for matrices.

Definition A.1.2. *Let X be a complex Hilbert space, and let $L : X \rightarrow X$ be a bounded linear operator. A complex number λ is called an **eigenvalue** of L if there is a nonzero vector $x \in X$ such that*

$$Lx = \lambda x.$$

Any such vector x is called an **eigenvector** of L associated with the eigenvalue λ . The set of all eigenvalues of L is called the **point spectrum** of L .

From now on we assume X and Y are Hilbert spaces.

Theorem A.1.2. *Let $L : X \rightarrow X$ be a bounded linear operator.*

- (a) *If λ is an eigenvalue of L , then $|\lambda| \leq \|L\|$.*
- (b) *If L is compact, then*
 - (i) *The point spectrum of L is countable (perhaps finite or even empty).*
 - (ii) *The eigenspace $E_\lambda(L)$ of a nonzero eigenvalue λ is finite-dimensional, that is, the set $E_\lambda(L) = \{x \in X \mid Lx = \lambda x\}$ is finite-dimensional. The dimension of $E_\lambda(L)$ will be called the **multiplicity** of the eigenvalue λ .*
- (c) *If L is self-adjoint, then all eigenvalues of L are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal.*
- (d) *If L is self-adjoint and compact, then $-\|L\|$ or $\|L\|$ is an eigenvalue of L .*

Suppose that $T : X \rightarrow X$ is a self-adjoint compact operator. Then in view of Theorem A.1.2, we can arrange the nonzero eigenvalues of T in a sequence $\lambda_1, \lambda_2, \lambda_3, \dots$ such that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \dots$ and each eigenvalue is repeated according to its multiplicity. Let $\phi_1, \phi_2, \phi_3, \dots$ be a corresponding sequence of orthonormal eigenvectors. The set $\{(\lambda_n, \phi_n)\}$ is called an *eigensystem* of K .

Now we state the **spectral theorem** for self-adjoint compact operators.

Theorem A.1.3. *Let $\{(\lambda_n, \phi_n)\}$ be an eigensystem of a compact self-adjoint operator $T : X \rightarrow X$. Then*

- (a) *T has the spectral expansion*

$$Tx = \sum_n \lambda_n (\phi_n, x)_X \phi_n \quad \forall x \in X.$$

- (b) *The set $\{\phi_1, \phi_2, \dots\}$ form a basis for $\mathcal{N}(T)^\perp = \overline{\mathcal{R}(T)}$.*
- (c) *If $\{\lambda_n\}$ is infinite, then $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$.*

Let $K : X \rightarrow Y$ be a compact operator (but not necessarily self-adjoint). It can be shown that the operators $K^*K : X \rightarrow X$ and $KK^* : Y \rightarrow Y$ are self-adjoint, compact, and they have the same eigenvalues and these eigenvalues are nonnegative. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$ be an enumeration of the nonzero eigenvalues of K^*K , and let $\{\phi_n\}$ be a sequence of associated orthonormal eigenvectors of K^*K . Define $\sigma_n = \sqrt{\lambda_n}$, and set

$$\psi_n = \sigma_n^{-1} K \phi_n.$$

The set $\{(\sigma_n, \phi_n, \psi_n)\}$ is called a *singular system* for the compact operator K , and the numbers $\{\sigma_n\}$ are called the *singular values* of K . We state the following fundamental theorem.

Theorem A.1.4. *Let $K : X \rightarrow Y$ be a compact operator, and let $\{(\sigma_n, \phi_n, \psi_n)\}$ be a singular system for K . Then*

$$Kx = \sum_n \sigma_n (\phi_n, x)_X \psi_n \quad \forall x \in X.$$

*This representation is called the **singular value expansion** of K .*

We also mention the following important characterization.

Theorem A.1.5. *K has only finitely many singular values if and only if K is of finite-rank, i.e. $\mathcal{R}(K)$ is finite-dimensional.*

Theorem A.1.6. *Let $\{(\sigma_n, \phi_n, \psi_n)\}$ be a singular system for the compact operator K , $y \in Y$. Then*

(a) *$y \in D(K^\dagger)$ if and only if*

$$\sum_n \frac{|(\psi_n, y)_Y|^2}{\sigma_n^2} < \infty.$$

*This condition is called **Picard criterion**.*

(b) *For $y \in D(K^\dagger)$,*

$$K^\dagger y = \sum_n \frac{(\psi_n, y)_Y}{\sigma_n} \phi_n.$$

The following theorem introduces the notion of functions of compact self-adjoint operators. The general framework is known as *functional calculus*.

Theorem A.1.7. *Let $T : X \rightarrow X$ be a compact self-adjoint operator with spectral expansion*

$$Tx = \sum_n \lambda_n (\phi_n, x)_X \phi_n \quad \forall x \in X.$$

If $f : [-\|T\|^2, \|T\|^2] \rightarrow \mathbb{R}$ is piecewise continuous, then the operator $T_f : X \rightarrow X$ given by

$$T_f x = \sum_n f(\lambda_n) (\phi_n, x)_X \phi_n \quad \forall x \in X$$

is a self-adjoint bounded linear operator.

Let X and Y be Hilbert spaces, and let $K : X \rightarrow Y$ be compact with singular value expansion

$$Kx = \sum_n \sigma_n (\phi_n, x)_X \psi_n \quad \forall x \in X.$$

For $\mu > 0$, the operator $(K^*K)^\mu : X \rightarrow X$ is defined by

$$(K^*K)^\mu x = \sum_n (\sigma_n^2)^\mu (\phi_n, x)_X \phi_n \quad \forall x \in X.$$

In particular, it can be shown that $\mathcal{R}((K^*K)^{1/2}) = \mathcal{R}(K^*)$.

Most of the convergence results presented in Section 2.1, require a priori assumptions on the exact solution of $Tx = y$. Precisely, $T^\dagger y$ should satisfy the abstract smoothness assumption $T^\dagger y \in \mathcal{R}((T^*T)^\mu)$ for some $\mu > 0$. In the case T is compact, this condition has the following realization:

Theorem A.1.8. *Let K be compact with singular system $\{(\sigma_n, \phi_n, \psi_n)\}$. Then, for $\mu > 0$,*

$$K^\dagger y \in \mathcal{R}((K^*K)^\mu) \iff \sum_n \frac{|(\psi_n, y)_Y|^2}{\sigma_n^{2+4\mu}} < \infty.$$

Notice that when K is compact but not of finite-rank, then K has infinitely many singular values $\{\sigma_n\}$ and further $\sigma_n \rightarrow 0$ as $n \rightarrow \infty$. Now, Picard criterion states $y \in D(K^\dagger)$ if only if the ‘Fourier coefficients’ $(\psi_n, y)_Y$ decay faster than the singular values of σ_n . In Fourier series analysis, the rate at which the Fourier coefficients decay indicates how smooth the function is; the smoother the function the faster the Fourier coefficients converge to 0. Thus, in comparison, the requirement $K^\dagger y \in \mathcal{R}((K^*K)^\mu)$ expresses a smoothness assumption on the exact data y (at least in the abstract

sense), and hence the name ‘abstract smoothness assumption’. Moreover, in view of Theorem A.1.8, we see that the larger the μ , the ‘smoother’ the ‘function’ y is. For concrete examples and more explanations we recommend the book [18].