

Linear and Nonlinear Inverse Problems with Practical Applications

Jennifer L Mueller and Samuli Siltanen

August 29, 2013

Contents

0 Preface	11
I Linear Inverse Problems	15
1 Introduction	17
2 Naïve Reconstructions and Inverse Crimes	23
2.1 Convolution	23
2.1.1 Continuum model for one-dimensional convolution	24
2.1.2 Discrete convolution model	26
2.1.3 Naïve deconvolution and inverse crimes	28
2.1.4 Naïve reconstruction without inverse crime	30
2.2 Heat propagation	31
2.2.1 Diffusion processes	31
2.2.2 A finite difference discrete model	35
2.2.3 Naïve reconstruction of the initial temperature	37
2.3 Tomographic X-ray projection data	41
2.3.1 A simple example: probing two aluminum slabs	41
2.3.2 From photon count data to line integral data	42
2.3.3 Continuous tomographic data: the Radon transform .	44
2.3.4 Discrete tomographic data	48
2.3.5 Naïve reconstruction	52
2.3.6 Naïve reconstruction without inverse crime	54
3 Ill-Posedness in Inverse Problems	57
3.1 Forward map and Hadamard's conditions	57
3.2 Ill-posedness of the backward heat equation	58
3.3 Ill-posedness in the continuous case	63
3.3.1 The Singular Value Expansion (SVE)	67

3.3.2	Discretization	69
3.4	Regularized inversion	71
3.5	The singular value decomposition for matrices	73
3.6	SVD for the guiding examples	77
4	Truncated singular value decomposition	79
4.1	Minimum norm solution	79
4.2	Truncated SVD	81
4.3	Measuring the quality of reconstructions	83
4.4	TSVD for the guiding examples	83
4.4.1	TSVD for deconvolution	83
4.4.2	TSVD for backward heat propagation	85
4.4.3	TSVD for X-ray tomography	87
5	Tikhonov Regularization	89
5.1	Classical Tikhonov regularization	89
5.1.1	Tikhonov regularization for the deconvolution problem	91
5.1.2	Tikhonov regularization for backward heat propagation	91
5.2	Normal equations and stacked form	94
5.3	Generalized Tikhonov regularization	96
5.3.1	Generalized Tikhonov regularization for the deconvolution problem	98
5.4	Choosing the regularization parameter	101
5.4.1	Morozov's discrepancy principle	101
5.4.2	The L-curve method	102
5.5	Large-scale implementation	110
5.5.1	Conjugate direction methods	110
5.5.2	Conjugate gradient	112
5.5.3	Preconditioning	113
6	Total variation regularization	115
6.1	What is total variation?	116
6.2	Quadratic programming	119
6.3	Sparsity-based parameter choice	121
6.4	Large-scale implementation	123
6.4.1	TV regularization for tomography	126
7	Besov space regularization using wavelets	129
7.1	An introduction to wavelets	129
7.2	Besov spaces and wavelets	133

7.3	Using B_{11}^1 regularization to promote sparsity	134
8	Discretization-invariance	137
8.1	Tikhonov regularization and discretizations	139
8.2	Total variation regularization and discretizations	142
8.3	Besov norm regularization and discretizations	142
9	Practical X-ray tomography with limited data	147
9.1	Sparse full-angle tomography	151
9.1.1	Simulated data	151
9.1.2	Real data: total variation regularization	154
9.1.3	Real data: sparsity-promoting tomography	158
9.2	Limited-angle tomography	160
9.3	Low-dose 3D dental X-ray imaging	164
9.3.1	Tomosynthesis	164
9.3.2	Tuned-aperture computed tomography	167
9.3.3	Panoramic dental imaging	167
9.3.4	The VT device for dental implant planning	172
10	Projects	177
10.1	Image deblurring	178
10.2	Inversion of the Laplace transform	179
10.3	Backward parabolic problem	180
II	Nonlinear Inverse Problems	183
11	Nonlinear inversion	185
11.1	Analysis of nonlinear ill-posedness	187
11.1.1	Uniqueness	187
11.1.2	Continuous dependence on data	187
11.1.3	Existence	188
11.2	Nonlinear regularization	189
11.3	Computational inversion	191
11.3.1	Iterative nonlinear regularization	192
11.3.2	Direct nonlinear regularization	192
11.4	Examples of nonlinear inverse problems	193
11.4.1	Glottal inverse filtering	193
11.4.2	Inverse problems in hydrology	194
11.4.3	Inverse medium scattering in \mathbb{R}^3	196

11.4.4 Spectral inversion	200
11.4.5 Diffuse optical tomography	204
12 Electrical impedance tomography	209
12.1 Applications of EIT	211
12.2 Derivation from Maxwell's equations	213
12.3 Continuum-model boundary measurements	215
12.3.1 The Dirichlet-to-Neumann map	215
12.3.2 The Neumann-to-Dirichlet map	216
12.4 Nonlinearity of EIT	217
12.5 Ill-posedness of EIT	217
12.5.1 Alessandrini's example	218
12.5.2 Ill-posedness of EIT: a simulation study	221
12.6 Electrode models	225
12.6.1 The gap model	225
12.6.2 The shunt model	226
12.6.3 The complete electrode model	227
12.7 Current patterns and distinguishability	227
12.7.1 Popular choices of current patterns	227
12.7.2 Distinguishability and optimal current patterns	230
12.8 Further reading	236
12.8.1 Three-dimensional EIT	236
12.8.2 Reconstruction at the boundary	237
12.8.3 Non-circular boundary	237
12.8.4 Complex-valued conductivity	237
12.8.5 Lack of the dummy load measurement	238
12.8.6 Counterexamples and invisibility	240
13 Simulation of noisy EIT data	243
13.1 Eigenvalue data for symmetric σ	244
13.2 Continuum model data and FEM	246
13.2.1 Computing the DN matrix directly	246
13.2.2 Computing the ND matrix directly	248
13.2.3 Computing the DN matrix using the ND matrix	249
13.3 Complete electrode model and FEM	251
13.3.1 Variational formulation	251
13.3.2 Finite element approximation	253
13.3.3 Computing the DN matrix	257
13.4 Adding noise to EIT data matrices	258

14 Complex geometrical optics solutions	261
14.1 Calderón's pioneering work	264
14.2 The $\bar{\partial}$ operator and its kin	268
14.3 CGO solutions for the Schrödinger equation	269
14.3.1 The generalized Lippmann-Schwinger equation	270
14.3.2 Faddeev Green's function in dimension two	272
14.3.3 Reduction to a periodic integral equation	277
14.3.4 Fast solver for the periodic equation	278
14.3.5 Numerical examples	280
14.4 CGO solutions for the Beltrami equation	285
14.4.1 The real-linear Beltrami equation	285
14.4.2 Reformulation with a complex-linear equation	286
14.4.3 Reduction to a periodic integral equation	288
14.4.4 Fast solver for the periodic equation	289
14.4.5 Numerical examples	289
15 A regularized D-bar method for direct EIT	293
15.1 Reconstruction with infinite-precision data	294
15.1.1 The scattering transform	297
15.1.2 From Λ_σ to $\psi _{\partial\Omega}$ using a boundary integral equation .	298
15.1.3 From $\psi _{\partial\Omega}$ to \mathbf{t} via integration over the boundary .	301
15.1.4 From \mathbf{t} to m : the D-bar equation	301
15.1.5 From m to σ	303
15.2 Regularization via nonlinear low-pass filtering	303
15.3 Numerical solution of the boundary integral equation	307
15.4 Numerical solution of the D-bar equation	310
15.4.1 Reduction to a periodic equation	312
15.4.2 Numerical solution of the periodic equation	314
15.5 Regularized reconstructions	317
15.5.1 Ground truth: computing \mathbf{t} directly from σ	317
15.5.2 Simulation of continuum-model EIT data	319
15.5.3 Computation of the scattering transform from data .	319
15.5.4 Solving the D-bar equation	320
16 Other direct solution methods for EIT	327
16.1 D-bar methods with approximate scattering transforms . . .	327
16.1.1 The \mathbf{t}^{exp} approximation	328
16.1.2 The \mathbf{t}^0 approximation	337
16.2 Calderón's method	346

16.2.1 Calderón’s method as an approximation of the D-bar method	348
16.3 The Astala-Pääivärinta method	351
16.3.1 Complex geometric optics solutions	351
16.3.2 The transport equation	353
16.3.3 The boundary integral equation	354
16.3.4 The low-pass transport matrix	358
16.3.5 Final reconstruction steps	360
16.3.6 Numerical results	360
16.4 The enclosure method of Ikehata	366
17 Projects	371
17.1 Enclosure method for EIT	371
17.1.1 Simulation of EIT data	371
17.1.2 Calculation of the indicator function	372
17.1.3 Experiments with non-noisy data	372
17.1.4 Experiments with noisy data	373
17.1.5 Visualization of reconstructions	373
17.2 The D-bar method with Born approximation	373
17.2.1 Simulation of data	374
17.2.2 Computation of the scattering transform	375
17.2.3 Solution of the D-bar equation	376
17.2.4 Computation and visualization of reconstructions	376
17.3 Calderón’s method	377
17.3.1 Computational aspects	377
17.4 Inverse obstacle scattering	378
17.4.1 Numerical computation of far field patterns	379
A Banach Spaces and Hilbert Spaces	431
B Mappings and Compact Operators	433
C Fourier Transform and Sobolev Spaces	437
C.1 Sobolev spaces on domains $\Omega \subset \mathbb{R}^n$	437
C.2 Fourier series and spaces $H^s(\partial\Omega)$	442
C.2.1 Matrix approximation of operators	444
C.3 Traces of functions in $H^m(\Omega)$	447

<i>CONTENTS</i>	9
-----------------	---

D Iterative solution of linear equations	449
---	------------

D.0.1 The minimum norm residual approach	451
--	-----

D.0.2 Application to the solution of the D-bar equation	452
---	-----

D.0.3 A simple example	453
----------------------------------	-----

Chapter 0

Preface

Inverse problems arise from the need to interpret indirect and incomplete measurements. As an area of contemporary mathematics, the field of inverse problems is strongly driven by applications and has been growing steadily in the past 30 years. This growth has been fostered both by advances in computation and by theoretical breakthroughs. Modern digital sensors provide vast amounts of data related to diverse areas including engineering, geophysics, medicine, biology, physics, chemistry, and finance. As a result, the need for computational inversion can be expected to increase in the future.

We write this book with the goal of providing a practical introduction to inverse problems from both a computational and theoretical perspective. A solid theoretical framework is mandatory for understanding why *ill-posed inverse problems* require a different set of solution methods than well-posed problems. Ill-posedness is related to interpretation tasks that are extremely sensitive to measurement and modelling errors. On the other hand, solving an inverse problem involves the implementation of a computational algorithm that recovers useful information from measured data (with the word “useful” understood in the context of a particular application). A successful inversion algorithm is robust against measurement noise, computationally effective, and mathematically justified by appropriate analysis and theorems.

Much of the literature on computational inversion considers tailored methods for linear problems (such as filtered back-projection for X-ray tomography) and generic iterative methods for nonlinear problems (such as Tikhonov regularization with nonlinear objective function). However, in this book we do exactly the opposite: we discuss a unified solution framework for linear problems and tailored direct methods for nonlinear problems. Our rationale is the following:

- Linear inverse problems are all essentially alike since they are completely described by the singular value expansion of the forward map. Thus it makes sense to apply a general methodology designed for complementing measurement data with *a priori* information, for example by enforcing non-negativity or by promoting sparsity in a basis.
- Nonlinear inverse problems are all different and need dedicated solution methods. One way to proceed is to use the results of the analytic-geometric inverse problems research tradition to construct regularized algorithms.

The book is organized into two parts. The first part, *Linear Inverse Problems*, is suitable for a one-semester undergraduate course or for a part of a graduate course. We present both continuous and discrete inverse problems to instruct how the ill-posedness is inherent in the idealized inverse problem and how it shows up in the real-life problem and in its discretization. With this approach we hope the reader will develop a deeper understanding of the connection between the mathematical theory, the computational model, and the practical problem arising from the application.

The guiding examples in Part I are the problem of image deblurring, X-ray tomography, and backwards heat propagation. We discuss how to realistically simulate measurement data for all three. A dangerous pitfall in algorithm development and testing is the act of committing an *inverse crime*, that is, obtaining a great reconstruction due to the fact that the simulated data resonates in some helpful way with the reconstruction algorithm. We explain how this can occur and how to avoid it.

The use of Besov spaces and wavelets as a means of regularization is included in addition to the classical methods of truncated singular value decomposition, Tikhonov regularization, and total variation regularization. Also, this is the first book to discuss the recently introduced sparsity-based parameter choice rules. Many practical problems demand the use of very large data sets, and appropriate large-scale variants of the above reconstruction methods are addressed as well.

Part I requires knowledge of basic linear algebra and matrix computations, some knowledge of PDE's, basic analysis, and some programming skills. Some material for Part I is provided in the appendices.

Part II addresses nonlinear inversion, and it is a suitable text for a graduate course in applied mathematics. Also, we have received many requests for a text on the D-bar method, and Part II is designed to fulfill this need. We hope that researchers in electrical impedance tomography (EIT) will welcome this exposition.

The guiding example for Part II is EIT, although several other examples are also discussed briefly. Taking one guiding example is in accordance with the above rationale of treating nonlinear inverse problems as unique cases needing tailored solution strategies. We hope that the detailed discussion of the regularized D-bar method for EIT serves as a model for further research regarding other nonlinear inverse problems.

Actually, we see the regularized D-bar method as a topic that combines and unifies several schools of thought. Namely, there appear to be rather separated research traditions in the field of inverse problems, including the following three. The *analytic-geometric tradition* treats inverse problems as coefficient recovery tasks for partial differential equations. The main questions studied are uniqueness and stability proofs for recovering coefficient functions from limited but infinite-precision information. The *regularization tradition* studies the construction of continuous maps from the data space to the model space providing approximate reconstructions from indirect measurement data containing errors. The main questions studied are convergence rates of the reconstructions at the asymptotic limit of zero measurement noise. The *engineering tradition* is involved with writing robust computational algorithms that recover useful information from practical data.

In the regularized D-bar method, the reconstruction technique is defined using complex geometric optics solutions and a nonlinear Fourier transform introduced by the analytic-geometric tradition, the regularization strategy is provided by a nonlinear low-pass filtering step that can be analysed according to the standards of the regularization tradition, and the final result is a robust imaging algorithm applicable to *in vivo* medical EIT.

Part II requires some background in partial differential equations, complex analysis, functional analysis, the finite element method and numerical solution methods for linear systems. However, we do include short introductions to these topics in the text and in the appendices.

Throughout the book we provide exercises and project works involving MATLAB programming. Selected pieces of software can be downloaded from the website www.inverseproblemsbook.org.

We also wish to take this opportunity to thank the many people who have provided support, encouragement, and help with this text. We thank David Isaacson for his guidance and encouragement early in our careers. We are grateful to Kari Astala, Jutta Bikowski, Sarah Hamilton, Claudia Natalia Lara Herrera, Masaru Ikehata, Kim Knudsen, Ville Kolehmainen, Matti Lassas, Jon Newell, Allan Perämäki and Lassi Päävärinta for their help in reporting our joint works in this book. We thank Per Christian Hansen, An-

dreas Hauptmann, Olli Koskela and Jussi Määttä for their helpful comments regarding the manuscript, and Aki Kallonen, Keijo Hämäläinen, Martti Kalke, Jyrki Saarinen and PaloDEX Group for providing measurement data and image material concerning X-ray imaging. We thank Kati Niinimäki, Esa Niemi and Miguel Fernando Montoya Vallejo for creating some of the figures, and Ethan Murphy for providing some of the figures from his thesis. Furthermore, we thank Simon Arridge, Guillaume Bal, Martin Burger, Fioralba Cakoni, Martin Hanke, Tasio Helin, Slava Kurylev, Peter Maass, Adrian Nachman, Hanna Pikkarainen, Ronny Ramlau, Otmar Scherzer, Carola Schönlieb, John Schotland, Mikko Sillanpää and Erkki Somersalo for valuable discussions. We thank Alexandra Bellow for providing the two photos of Alberto Calderón included in Chapter 14. Thanks to Elizabeth Greenspan from SIAM for her help, patience and support during the preparation of the manuscript. Part of the time writing the book was spent at Mathematical Sciences Research Institute in Berkeley, California, and at Isaac Newton Institute in Cambridge, UK, whose hospitalities are greatly appreciated. Finally, we thank all our families for their love and support.

Jennifer Mueller
Samuli Siltanen

Fort Collins, Colorado
Helsinki, Finland
January, 2012

Part I

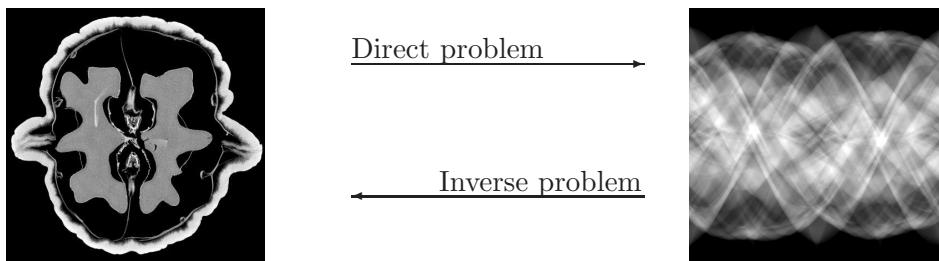
Linear Inverse Problems

Chapter 1

Introduction

Inverse problems are the opposites of direct problems. Informally, in a direct problem one finds an effect from a cause, and in an inverse problem one is given the effect and wants to recover the cause. The most usual situation giving rise to an inverse problem is the need to interpret indirect physical measurements of an unknown object of interest.

For example, in X-ray tomography the direct problem is to determine the X-ray projection images we would get from a physical body whose internal structure we know precisely. The corresponding inverse problem is to reconstruct the inner structure of an unknown physical body from the knowledge of X-ray images taken from different directions. Here is a two-dimensional example:

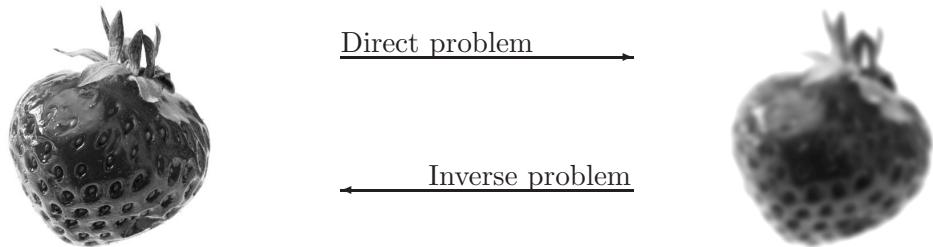


Here the slice through a walnut (left) is the cause and the collection of X-ray data (right) is the effect. The tomographic data is shown in the traditional *sinogram* form, which will be discussed in detail in Section 2.3.5. The slice image on the left is courtesy of Keijo Hämäläinen and Aki Kallonen from University of Helsinki, Finland.

Variants of the above tomographic problem appear also in monitoring ozone profiles in upper atmosphere using spaceborne star occultation mea-

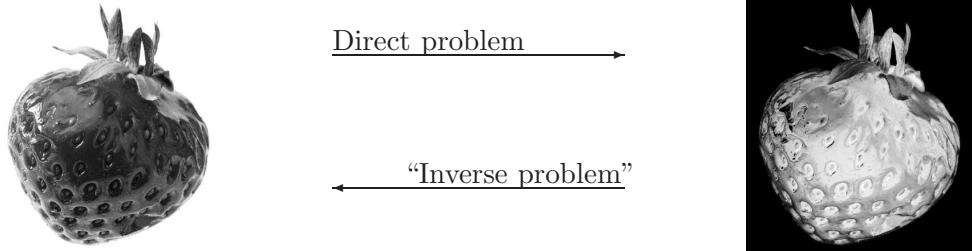
surements [295], identifying molecules based on electron microscope imaging [138], interpretation of Doppler weather radar measurements [298], and measuring the temperature distribution of hot gases flowing through the window of a burning house using metal wires [30]. This demonstrates the general nature of mathematics: the same underlying problem may be found in very different application areas.

Another example comes from image processing. Define the direct problem as finding out how a given sharp photograph would look like if the camera was incorrectly focused. The inverse problem known as *deblurring* is finding the sharp photograph from a given blurry image.



Here the cause is the sharp image and the effect is the blurred image. A famous example of deblurring is the correction algorithm used for Hubble space telescope images after finding out a flaw in the construction of its main lens.

There is an apparent symmetry in the above explanation: without further restriction of the definitions, direct problem and inverse problem would be in identical relation with each other. For example, we might take as the direct problem the determination of a positive photograph from the knowledge of the negative photograph.



In this case the corresponding “inverse problem” would be inverting a given photograph to arrive at the negative. Here both problems are easy and stable, and one can move between them repeatedly.

However, we concentrate on *ill-posed inverse problems*, where the inverse problem is more difficult to solve than the direct problem. To explain this

we need the notion of a *well-posed problem* introduced by Jacques Hadamard (1865-1963):

H_1 : **Existence.** There should be at least one solution.

H_2 : **Uniqueness.** There should be at most one solution.

H_3 : **Stability.** The solution must depend continuously on data.

Now denote by \mathcal{A} the *forward map*, defined conceptually by $\mathcal{A}(\text{cause}) = \text{effect}$. The direct problem must be well-posed, in other words \mathcal{A} should be a well-defined, single-valued and continuous function. The inverse problem is ill-posed if \mathcal{A}^{-1} does not exist or is not continuous; then at least one of the conditions H_1-H_3 fails for \mathcal{A}^{-1} . In the positive-negative photograph example above both \mathcal{A} and \mathcal{A}^{-1} are well-defined and continuous, so it is not an ill-posed inverse problem.

Our general mathematical model is constructed as follows. We consider indirect linear measurements of the form

$$\mathbf{m} = \mathcal{A}f + \varepsilon, \quad (1.1)$$

where f is a piecewise continuous function defined on a subset of \mathbb{R}^d and $\mathbf{m} \in \mathbb{R}^k$ is a vector of numbers given by a measurement device. Models of the form (1.1) arise from various situations in technology or physics; the linear operator \mathcal{A} may be related, for example, to a partial differential equation or to an integral equation. We will discuss examples of practically relevant operators \mathcal{A} in Chapter 2.

The vector $\varepsilon \in \mathbb{R}^k$ in (1.1) models errors coming from measurement noise, which is inevitable in practical situations. Sometimes ε is modelled as a random variable with certain statistics. However, in this book we think of ε as a deterministic but unknown error; the information we have on ε is an inequality $\|\varepsilon\| \leq \delta$ with a known constant $\delta > 0$. Such a number δ can often be found by calibration of the measurement device; higher-quality device typically gives smaller δ . Our deterministic approach does include the possibility that ε is a fixed realization of a random process; this is actually a quite accurate model of many realistic measurements.

The reason for the term “indirect measurement” is the following. We are interested in the function f but cannot measure its values directly. However, f is connected to another physical quantity \mathbf{m} , which is available for measurement. The connection is modelled by the linear operator \mathcal{A} . Now the direct problem is “Given f , determine $\mathbf{m} = \mathcal{A}f$ ”. The corresponding

inverse problem is

$$\begin{aligned} \text{Given noisy measurement } \mathbf{m} = \mathcal{A}\mathbf{f} + \varepsilon \text{ and } \delta > 0 \\ \text{with } \|\varepsilon\| \leq \delta, \text{ extract information about } \mathbf{f}. \end{aligned} \quad (1.2)$$

Part I of this book is about practical extraction of information from indirect linear measurements using computational methods. Consequently, we need to introduce a finite-dimensional approximation $\mathbf{f} \in \mathbb{R}^n$ to the function f and to build a matrix model for the linear operator \mathcal{A} .

In case of discrete linear inverse problems we consider measurements of the form

$$\mathbf{m} = \mathcal{A}\mathbf{f} + \varepsilon, \quad (1.3)$$

where $\mathbf{m} \in \mathbb{R}^k$ and $\mathbf{f} \in \mathbb{R}^n$. Moreover, A is a matrix of size $k \times n$ (k rows and n columns). Strictly speaking, we abuse notation by using \mathbf{m} in both (1.1) and (1.3) although they are different models. Whenever there is a danger of confusion, we will denote the measurement from the finite model (1.3) by $\mathbf{m}^{(n)}$.

Once the computational model (1.3) has been constructed, it is tempting to try to solve the inverse problem (1.2) by the naïve reconstruction

$$\mathbf{f} \approx A^{-1}\mathbf{m}. \quad (1.4)$$

However, in the case of ill-posed inverse problems the approach (1.4) will fail. In Sections 2.1–2.3 we describe some important indirect measurements and demonstrate the failure of the naïve reconstruction (1.4) numerically.

Regularization is what really needs to be done for successful and noise-robust solution of linear inverse problems. We discuss the theory and implementation of various regularization methods in Chapters 3–7. We demonstrate the properties of the various methods using the example problems developed in Chapter 2. Large-scale computational methods are emphasized throughout the text because practical applications often lead to very high-dimensional problems.

Discretization of the continuum model (1.1) using discrete models of the form (1.3) involves choosing the dimension n of the discrete vector \mathbf{f} . It is desirable to design computational inversion methods that give consistent results at different resolutions n . This so-called *discretization-invariance* is discussed in Chapter 8.

Practical examples of linear inversion are described in Chapter 9 in the case of X-ray tomography.

It is impossible to cover all useful and important material related to computational inversion in this book. We list here some further reading that complements our approach.

Regularization theory is discussed in the classical texts [438, 338] and in the more recent books [131, 253, 468]. The mathematical foundations of inverse problems are explained more generally in the books [235, 268].

Some of the computational inversion methods presented in this book (such as truncated singular value decomposition, Tikhonov regularization and total variation regularization) are discussed also in [461, 193, 195, 194, 352]. Useful methodologies that are not covered here due to restrictions of space include truncated iterative solvers [192, 190, 64], approximate inverses [406], statistical inversion [246, 433, 66], and variational methods [368, 404].

Application-oriented texts are available for inverse problems in medical imaging [248, 351, 132], geophysical inversion [329, 433], and signal processing [402].

Chapter 2

Naïve Reconstructions and Inverse Crimes

This book is about developing computational solution methods for real-life inverse problems. The design of reconstruction algorithms is best done by first testing the code extensively with simulated data because every new aspect of the code can be systematically tested. Working directly with measured data may lead to very hard debugging problems as the source of difficulties can be hard to track.

What happens if proper simulation of errors is neglected? For example, using the same computational grid for the data simulation and reconstruction sometimes results in perfect reconstructions from noise-free data. Such a situation is not realistic and is referred to as an *inverse crime*. Excellent inversion results may be obtained, but these are not representative of any realistic inverse problem, since noise is present in any experimental setting. Such studies are inconclusive at best since robustness against modeling and measurement errors is not tested.

In this chapter, we will introduce these concepts in the context of the three guiding examples in Part I: deconvolution, the backward heat equation, and x-ray tomography.

2.1 Convolution

Linear convolution is a useful process for modeling a variety of practical measurements. *Deconvolution*, the corresponding inverse problem, is related to many engineering problems such as removing unwanted echoes from sound recordings or sharpening a misfocused photograph.

One-dimensional deconvolution will serve as a basic example throughout Part I of the book. Two-dimensional deconvolution is a project topic in Section 10.

2.1.1 Continuum model for one-dimensional convolution

We build a computational model for one-dimensional convolution with periodic boundary conditions. We consider 1-periodic functions $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $f(x) = f(x + n)$ with any integer $n \in \mathbb{Z}$. Essentially the function f is defined on an interval of length 1 such as $[0, 1]$ or $[-\frac{1}{2}, \frac{1}{2}]$ with the endpoints identified; another way of thinking about this is to consider $f(x)$ defined on a circle with radius $(2\pi)^{-1}$ and x being the arc length variable.

The reason for considering periodic functions is that we can avoid some technicalities related to boundary conditions that would obscure the main message about ill-posedness. Also, the Fourier transform and the wavelet transform are easily defined and implemented in the periodic setting.

The continuum measurement model concerns a 1-periodic signal $f : \mathbb{R} \rightarrow \mathbb{R}$ blurred by a 1-periodic *point spread function* (PSF) ψ . Other common names for the point spread function include *device function*, *impulse response*, *blurring kernel*, *convolution kernel* and *transfer function*.

Let us first construct the PSF using a building block ψ_0 defined in the interval $[-a, a] \subset \mathbb{R}$ with some constant $0 < a < 1/2$:

$$\psi_0(x) = C_a(x + a)^2(x - a)^2, \text{ for } -a \leq x \leq a, \quad (2.1)$$

where the constant $C_a := (\int_{-a}^a (x + a)^2(x - a)^2 dx)^{-1}$ is chosen to enforce the following normalization:

$$\int_{-a}^a \psi_0(x) dx = 1. \quad (2.2)$$

The periodic point spread function is defined by copying $\psi_0(x)$ to every interval $[n - a, n + a]$ with $n \in \mathbb{Z}$ and setting $\psi(x)$ to zero outside those intervals. The resulting ψ is a non-negative and even function:

$$\psi(x) \geq 0 \quad \text{and} \quad \psi(x) = \psi(-x) \quad \text{for all } x \in \mathbb{R}. \quad (2.3)$$

See Figure 2.1 for a plot of the point spread function with $a = 0.04$.

We remark that instead of (2.2) one often requires $\int_{-a}^a \psi_0(x)^2 dx = 1$. However, we prefer (2.2) since then constant functions remain unchanged in convolution with ψ ; this will be convenient below when we compare plots of reconstructions to the plot of the true signal by showing them in the same figure.

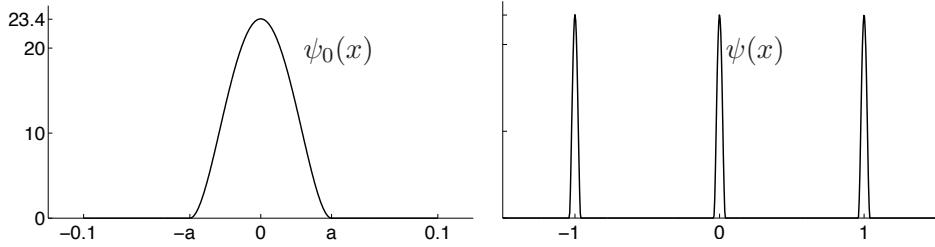


Figure 2.1: Point spread function according to (2.4) with $a = 0.04$ for one-dimensional convolution. Left: the continuously differentiable building block $\psi_0(x)$ used for constructing the periodic PSF. Right: the periodic PSF $\psi(x)$.

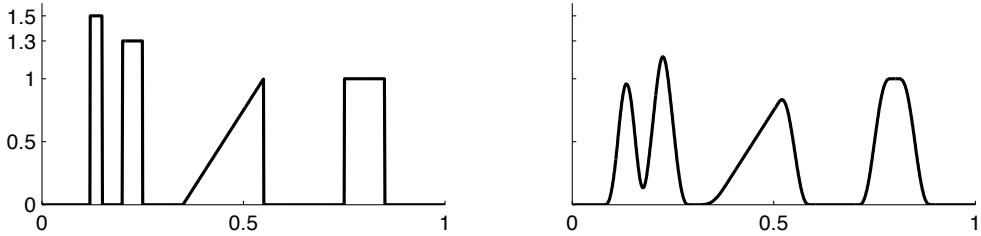


Figure 2.2: Effect of convolution on a piecewise continuous function. Left: target function $f(x)$. Right: the function $(\psi * f)(x)$.

Definition 2.1.1 *The continuum model of convolution, or blurring, is given by the following integral:*

$$(\psi * f)(x) = \int_{-a}^a \psi(x') f(x - x') dx'. \quad (2.4)$$

Note that formula (2.4) is not of the form (1.1) since the left hand side is not a k -dimensional vector. However, suppose the function f is defined on an interval $[b, b + 1]$, and assume that we have a device that measures the values of the convolution function $(\psi * f)(x)$ at a collection of k equally spaced points $\tilde{x}_1 = b, \tilde{x}_2 = b + \frac{1}{k}, \tilde{x}_3 = b + \frac{2}{k}, \dots, \tilde{x}_k = b + \frac{k-1}{k}$ and define

$$\mathbf{m} := [(\psi * f)(\tilde{x}_1), (\psi * f)(\tilde{x}_2), \dots, (\psi * f)(\tilde{x}_k)]^T \in \mathbb{R}^k. \quad (2.5)$$

Then $\mathcal{A}f = \mathbf{m}$ is of the form (1.1).

2.1.2 Discrete convolution model

Next we need to discretize the continuum model to arrive at a finite-dimensional measurement model of the form (1.3). Define

$$x_j = b + \frac{j-1}{n} \quad \text{for } j = 1, 2, \dots, n; \quad (2.6)$$

then the 1-periodic real-valued function $f(x)$ is represented by a vector \mathbf{f} containing values at the grid points:

$$\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n]^T = [f(x_1), f(x_2), \dots, f(x_n)]^T \in \mathbb{R}^n. \quad (2.7)$$

Furthermore, denote $\Delta x := x_2 - x_1 = 1/n$.

We can approximate the integral appearing in (2.4) by numerical quadrature. For any reasonably well-behaved function $g : [b, b+1] \rightarrow \mathbb{R}$ we have

$$\int_b^{b+1} g(x) dx \approx \Delta x \sum_{j=1}^n g(x_j), \quad (2.8)$$

the approximation becoming better as n increases.

For convenience, let us take $k = n$ and measure the convolution at the same points (2.6) as where the unknown function f is sampled. This is not necessary in general, but it will lead to a square-shaped matrix A , making it easy to illustrate naïve reconstructions and inverse crimes.

Let us construct an $n \times n$ matrix A so that $A\mathbf{f} \in \mathbb{R}^k$ approximates $\mathcal{A}f$ defined by (2.4). We define a discrete PSF denoted by

$$\mathbf{p} = [\mathbf{p}_{-\nu}, \mathbf{p}_{-\nu+1}, \dots, \mathbf{p}_{-1}, \mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{\nu-1}, \mathbf{p}_\nu]^T$$

as follows. Recall that $\psi_0(x) \equiv 0$ for $|x| > a > 0$. Take $\nu > 0$ to be the smallest integer satisfying the inequality $(\nu + 1)\Delta x > a$ and set

$$\tilde{\mathbf{p}}_j = \psi_0(j\Delta x) \quad \text{for } j = -\nu, \dots, \nu.$$

For example, with $a = 0.04$ as in Figure 2.1 and $n = 64$, we get $\nu = 2$. By (2.8) the normalization condition (2.2) almost holds: $\Delta x \sum_{j=-\nu}^{\nu} \tilde{p}_j \approx 1$. However, in practice it is a good idea to normalize the discrete PSF explicitly by the formula

$$\mathbf{p} = \left(\Delta x \sum_{j=-\nu}^{\nu} \tilde{\mathbf{p}}_j \right)^{-1} \tilde{\mathbf{p}}; \quad (2.9)$$

then it follows that

$$\Delta x \sum_{j=-\nu}^{\nu} \mathbf{p}_j = 1. \quad (2.10)$$

Now

$$\begin{aligned} \int_{-a}^a \psi(x') f(x_j - x') dx' &\approx \Delta x \sum_{\ell=-\nu}^{\nu} \psi(x_\ell) f(x_j - x_\ell) \\ &\approx \Delta x \sum_{\ell=-\nu}^{\nu} \mathbf{p}_\ell \mathbf{f}_{j-\ell}. \end{aligned}$$

Hence discrete convolution is defined by the formula

$$(\mathbf{p} * \mathbf{f})_j = \sum_{\ell=-\nu}^{\nu} \mathbf{p}_\ell \mathbf{f}_{j-\ell}, \quad (2.11)$$

where $\mathbf{f}_{j-\ell}$ is defined using periodic boundary conditions for the cases $j-\ell < 1$ and $j-\ell > n$. Then

$$\Delta x (\mathbf{p} * \mathbf{f}) \approx \mathcal{A}f, \quad (2.12)$$

and we define the measurement vector $\mathbf{m} = [\mathbf{m}_1, \dots, \mathbf{m}_k]^T$ by

$$\mathbf{m}_j = \Delta x (\mathbf{p} * \mathbf{f})_j + \varepsilon_j. \quad (2.13)$$

We would like to write formula (2.13) using a matrix A so that we would arrive at the desired model (1.3). To this end, set

$$\begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_k \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix}.$$

The answer is to build a circulant matrix having the elements of \mathbf{p} appearing systematically on every row of A .

Let us illustrate the structure of the convolution matrix A by an example in the case $n = 64$. As observed above, if $a = 0.04$ then $\nu = 2$, and the PSF takes the form $p = [p_{-2} \ p_{-1} \ p_0 \ p_1 \ p_2]^T$. According to (2.11) we have

$$\begin{aligned} (\mathbf{p} * \mathbf{f})_1 &= \mathbf{p}_0 \mathbf{f}_1 + \mathbf{p}_{-1} \mathbf{f}_2 + \mathbf{p}_{-2} \mathbf{f}_3 + \mathbf{p}_2 \mathbf{f}_{n-1} + \mathbf{p}_1 \mathbf{f}_n, \\ (\mathbf{p} * \mathbf{f})_2 &= \mathbf{p}_1 \mathbf{f}_1 + \mathbf{p}_0 \mathbf{f}_2 + \mathbf{p}_{-1} \mathbf{f}_3 + \mathbf{p}_{-2} \mathbf{f}_4 + \mathbf{p}_2 \mathbf{f}_n, \\ (\mathbf{p} * \mathbf{f})_3 &= \mathbf{p}_2 \mathbf{f}_1 + \mathbf{p}_1 \mathbf{f}_2 + \mathbf{p}_0 \mathbf{f}_3 + \mathbf{p}_{-1} \mathbf{f}_4 + \mathbf{p}_{-2} \mathbf{f}_5, \\ &\vdots \\ (\mathbf{p} * \mathbf{f})_n &= \mathbf{p}_{-1} \mathbf{f}_1 + \mathbf{p}_{-2} \mathbf{f}_2 + \mathbf{p}_2 \mathbf{f}_{n-2} + \mathbf{p}_1 \mathbf{f}_{n-1} + \mathbf{p}_0 \mathbf{f}_n. \end{aligned}$$

Consequently the matrix A looks like this:

$$A = \Delta x \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_{-1} & \mathbf{p}_{-2} & 0 & 0 & 0 & \cdots & \mathbf{p}_2 & \mathbf{p}_1 \\ \mathbf{p}_1 & \mathbf{p}_0 & \mathbf{p}_{-1} & \mathbf{p}_{-2} & 0 & 0 & \cdots & 0 & \mathbf{p}_2 \\ \mathbf{p}_2 & \mathbf{p}_1 & \mathbf{p}_0 & \mathbf{p}_{-1} & \mathbf{p}_{-2} & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{p}_2 & \mathbf{p}_1 & \mathbf{p}_0 & \mathbf{p}_{-1} & \mathbf{p}_{-2} & \cdots & 0 & 0 \\ \vdots & & & & & & \ddots & & \\ \vdots & & & & & & \ddots & & \\ 0 & 0 & \cdots & & \mathbf{p}_2 & \mathbf{p}_1 & \mathbf{p}_0 & \mathbf{p}_{-1} & \mathbf{p}_{-2} \\ \mathbf{p}_{-2} & 0 & \cdots & & 0 & \mathbf{p}_2 & \mathbf{p}_1 & \mathbf{p}_0 & \mathbf{p}_{-1} \\ \mathbf{p}_{-1} & \mathbf{p}_{-2} & \cdots & & 0 & 0 & \mathbf{p}_2 & \mathbf{p}_1 & \mathbf{p}_0 \end{bmatrix}; \quad (2.14)$$

note the systematic band-diagonal structure, which characterizes A as a circulant matrix. Linear systems involving circulant matrices can be quickly solved using Fast Fourier Transforms, a topic we will return to later.

Returning to the general case of \mathbf{p} defined by (2.9), the approximation formula (2.12) can be written in the form

$$A\mathbf{f} \approx \mathcal{A}\mathbf{f}. \quad (2.15)$$

Figure 2.3 shows data computed by the discrete model $A\mathbf{f}$ and compares the result to the continuous data $(\psi * f)(x)$ defined by (2.4).

Now let's add a little noise to the data. For example, we might take $k = 64 = n$ and construct the measurement noise in a probabilistic manner by taking a realization of a random vector with 64 independently distributed Gaussian elements having standard deviation $\sigma = 0.01 \cdot \max |f(x)|$. This corresponds to a relative noise level of 1%.

2.1.3 Naïve deconvolution and inverse crimes

We illustrate numerically the failure of the following naïve reconstruction attempt:

$$\mathbf{f} \approx A^{-1}\mathbf{m} \approx A^{-1}(A\mathbf{f} + \varepsilon) = \mathbf{f} + A^{-1}(\varepsilon). \quad (2.16)$$

In the case of no added noise ($\varepsilon = 0$) we use the data shown in the left plot of Figure 2.4 and get the left plot in Figure 2.5. The naïve reconstruction seems perfect! However, there is a catch. This apparently accurate reconstruction is not to be trusted; it is an example of an *inverse crime*. We will show how to avoid inverse crime in Section 2.1.4.

If we apply naïve reconstruction (2.16) to the slightly noisy data shown in the right plot of Figure 2.4, we get the result shown in the right plot

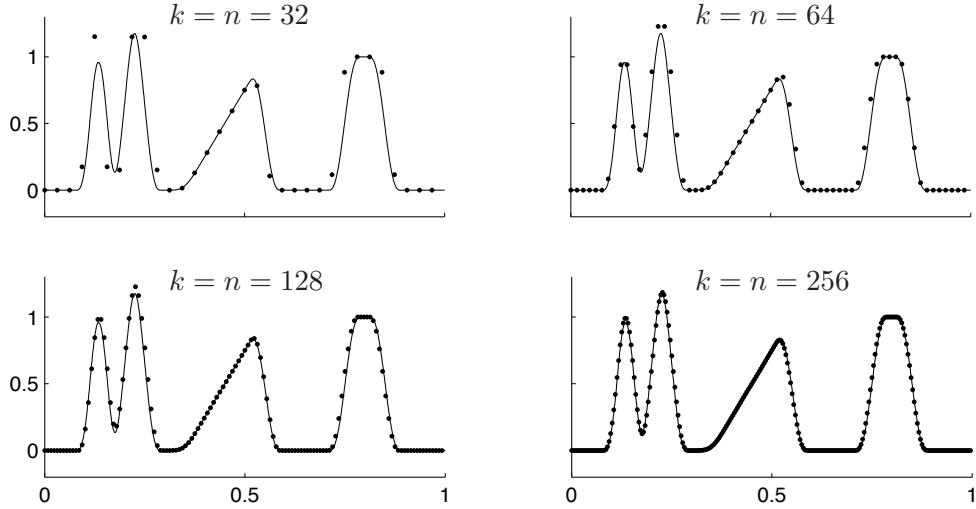


Figure 2.3: Illustration of the approximation $Af \approx A f$ of formula (2.15) for different choices of $k = n$. The actual function $(\psi * f)(x)$ defined by (2.4) is shown with a thin solid line, and the data points are indicated as dots. Note how the discrete approximation becomes better as the discretization is refined.

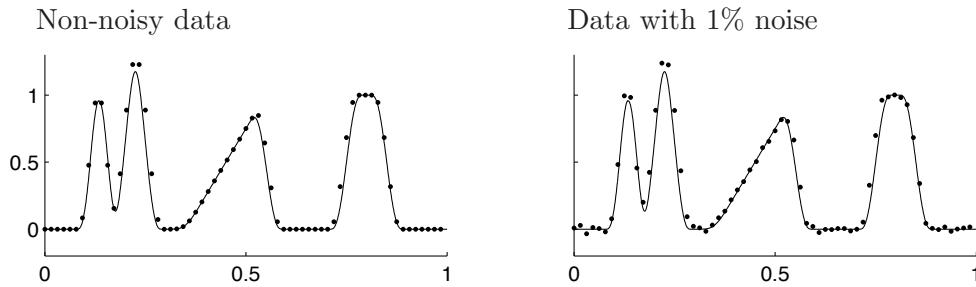


Figure 2.4: Illustration of simulated measurement noise. The actual function $(\psi * f)(x)$ defined by (2.4) is shown with a thin solid line, and the data points are indicated as dots. Left: non-noisy discrete data Af with $n = 64 = k$. Right: the same data corrupted with 1% white noise.

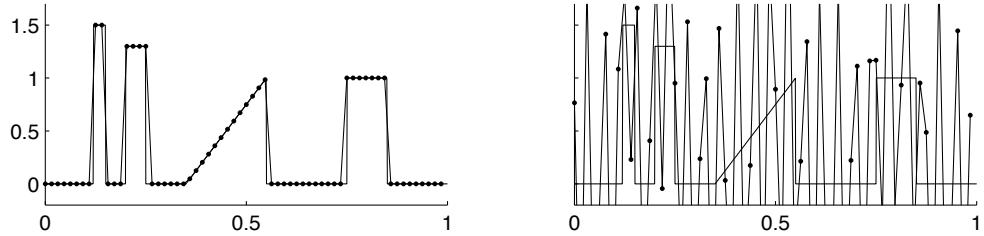


Figure 2.5: Two naïve deconvolutions by applying the inverse matrix A^{-1} to data. The original target function $f(x)$ is shown with a thin solid line, and the reconstruction is shown as dots. Left: naïve reconstruction (involving inverse crime) from the non-noisy discrete data $A\mathbf{f}$ with $n = k = 64$ shown in the left plot in Figure 2.4. Right: naïve reconstruction from the noisy data shown in the right plot of Figure 2.4.

in Figure 2.5. It is completely useless. This example shows how sensitive inverse problems are to the smallest errors in the measurement. We need to introduce *regularization* to overcome extreme sensitivity to measurement errors.

2.1.4 Naïve reconstruction without inverse crime

In the case of the deconvolution problem, we first simulate the measurements by convolving our known function f with a known discretized point spread function. In reality, when a blurred signal or image is encountered, the point spread function that “caused” the blurring is both unknown and can unlikely be expressed in simple terms. Thus, using the same point spread function for simulating a blurred signal and deconvolving the signal constitutes a serious inverse crime. Using the same PSF *and* the same discretization mesh is an inverse felony!

We show one simple way to avoid inverse crime. We use a modified point spread function by taking $a = 0.041$ in (2.1) when simulating data. We compute the function $(\psi * f)(x)$ defined in (2.4) approximately at 1000 uniformly spaced points in the interval $[0, 1]$ using trapezoidal rule with 400 quadrature points for the evaluation of the integral. Finally, we interpolate the values of $\psi * f$ at the 64 grid points using splines.

Now the data has been simulated completely differently than using the 64×64 model matrix A as was (criminally) done in Section 2.1.3.

We apply naïve inversion (2.16) to the crime-free data and show the results in Figure 2.6. Compare the left plots in Figures 2.5 and 2.6. Proper simulation of crime-free data reveals the ill-posedness of the deconvolution

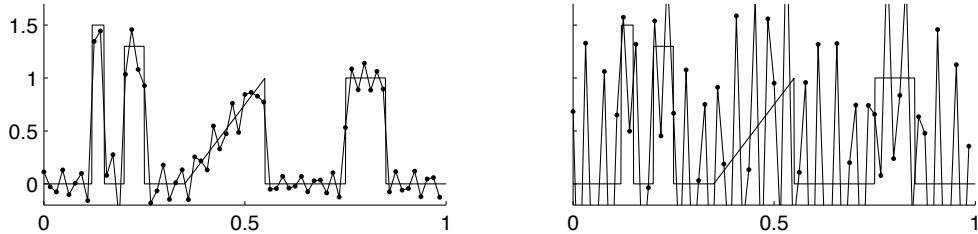


Figure 2.6: Two naïve deconvolutions by applying the inverse matrix A^{-1} to data generated avoiding inverse crime. The original target function $f(x)$ is shown with a thin solid line, and the reconstruction is shown as dots. Left: naïve reconstruction from non-noisy discrete data with $n = k = 64$. Right: naïve reconstruction from noisy data. Compare to Figure 2.5.

problem: the slightest perturbations in the data are amplified in naïve reconstruction using (2.16).

Exercise 2.1.1 Determine whether the PSF ψ is a $C^\infty(\mathbb{R})$ function.

Exercise 2.1.2 What is the effect of increasing the support of ψ_0 on ν ? Use the MATLAB programs `DCcontdatacomp.m` and `DCcontdataplot.m` to study the effect of increasing a on the convolved function. What do you observe?

Exercise 2.1.3 Plot a constant function of height 2 on $[0, 1]$ before and after convolution with ψ . Use the MATLAB program `DC2discretedatacomp.m` to add noise to the convolved function and `DC2naiveplot.m` to compute a naïve reconstruction. Plot your results.

2.2 Heat propagation

A classic ill-posed problem is that of determining the temperature distribution in a region from knowledge of the temperature distribution at the present time. This problem is known as the backward heat equation. We will begin with a discussion of the governing PDE's and their origins and then move to a simple discrete model.

2.2.1 Diffusion processes

The heat equation is the prototypical equation for modeling processes governed by pure diffusion. Following a probabilistic description as in, for

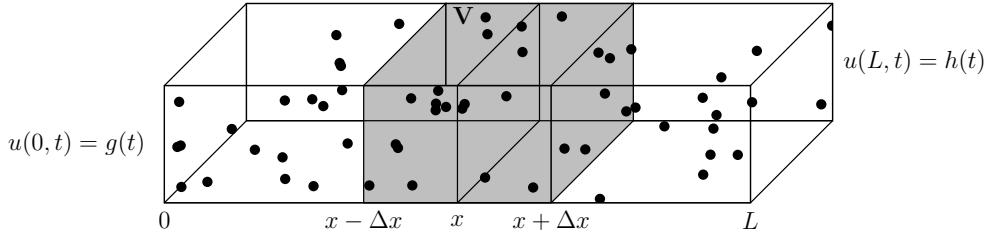


Figure 2.7: An illustration of the molecules in Brownian motion.

example [180], it can be derived by modeling the Brownian motion of the individual molecules in what we will assume to be a homogeneous material.

Suppose we have a material, such as depicted in Figure 2.7, containing n molecules, each of mass m , and suppose each molecule in this small volume is continually in motion. We will derive a model for one-dimensional spatial motion for simplicity, and so assume each molecule can only move to the left or to the right a distance Δx , representing an average displacement in time period Δt . To extend to higher dimensions, discrete motion in each of the three Cartesian coordinates would be permissible. Let p be the probability that the molecule moves to the right, and q the probability that the molecule moves to the left. Note that $p + q = 1$. Let $u(x, t)$ be the probability per unit length that a molecule is located in the interval $[x - \frac{1}{2}\Delta x, x + \frac{1}{2}\Delta x]$ at time t . The integral $\int_a^b u(x, t)dx$ is the probability that a molecule is located in $[a, b]$ at time t and $u(x, t)$ is a *probability density*.

The mass of molecules between $x - \Delta x$ and x is approximately

$$mn u(x - \frac{1}{2}\Delta x, t)\Delta x,$$

so the mass of the molecules crossing the plane at x from left to right at time t is approximately $pmnu(x - \frac{1}{2}\Delta x, t)\Delta x$. Similarly, the mass of the molecules crossing the plane at x from right to left at time t is approximately $qmnu(x + \frac{1}{2}\Delta x, t)\Delta x$. Thus, the net mass flux ϕ across the plane at x in the positive x -direction over a time interval Δt is approximately

$$\phi(x, t) \approx mn \frac{\Delta x}{\Delta t} \left(pu(x - \frac{1}{2}\Delta x, t) - qu(x + \frac{1}{2}\Delta x, t) \right).$$

Taking a Taylor series expansion for $u(x \pm \frac{1}{2}\Delta x, t)$ about (x, t) :

$$u(x \pm \frac{1}{2}\Delta x, t) = u(x, t) + u_x(x, t)(\pm \frac{1}{2}\Delta x) + \frac{1}{2!}u_{xx}(x, t)(\pm \frac{1}{2}\Delta x)^2 + \dots$$

gives the linear approximations

$$\begin{aligned} u(x + \frac{1}{2}\Delta x, t) &\approx u(x, t) + \frac{\Delta x}{2}u_x(x, t) \\ u(x - \frac{1}{2}\Delta x, t) &\approx u(x, t) - \frac{\Delta x}{2}u_x(x, t). \end{aligned}$$

Now the net mass flux across the plane at x is approximately

$$\begin{aligned} \phi(x, t) &\approx pmn\frac{\Delta x}{\Delta t} \left(u(x, t) + \frac{\Delta x}{2}u_x(x, t) \right) - qmn\frac{\Delta x}{\Delta t} \left(u(x, t) - \frac{\Delta x}{2}u_x(x, t) \right) \\ &= (p - q)mn\frac{\Delta x}{\Delta t}u(x, t) - \frac{mn(\Delta x)^2}{2\Delta t}u_x(x, t)(p + q). \end{aligned}$$

The concentration $c(x, t)$ of molecules between $x - \frac{1}{2}\Delta x$ and $x + \frac{1}{2}\Delta x$ is approximately

$$c(x, t) \approx \frac{1}{\Delta x}(mnu(x, t)\Delta x).$$

So in terms of concentration, our expression for net flux becomes:

$$\phi(x, t) \approx (p - q)\frac{\Delta x}{\Delta t}c(x, t) - \frac{1}{2}\frac{(\Delta x)^2}{\Delta t}c_x(x, t).$$

As $\Delta t \rightarrow 0$ assume that

$$\lim_{\Delta t \rightarrow 0} (p - q)\frac{\Delta x}{\Delta t} = \lambda,$$

where λ is a constant known as the drift constant, and

$$\lim_{\Delta t \rightarrow 0} \frac{1}{2}\frac{(\Delta x)^2}{\Delta t} = \frac{1}{2}D > 0,$$

where D is a constant known as the diffusion coefficient. Then in the limit as $\Delta t \rightarrow 0$ the net flux becomes

$$\phi(x, t) = \lambda c(x, t) - \frac{1}{2}Dc_x(x, t). \quad (2.17)$$

For a volume unit V , the quantity $\int_V c(x, t)dx$ represents the total mass in V , and conservation of mass implies that the time rate of change of the total mass in the volume V equals the flux across the boundary plus any mass created by sources f inside the volume V :

$$\frac{d}{dt} \int_V c(x, t)dx = - \int_{\partial V} \phi \cdot n ds + \int_V f dx. \quad (2.18)$$

By the divergence theorem

$$\int_{\partial V} \phi \cdot n ds = \int_V \nabla \cdot \phi dx,$$

and here in our 1-D model equation (2.18) takes the form

$$\frac{d}{dt} \int_{x-\Delta x}^{x+\Delta x} c(x, t) dx = - \int_{x-\Delta x}^{x+\Delta x} \frac{\partial}{\partial x} \phi(x, t) dx + \int_{x-\Delta x}^{x+\Delta x} f dx. \quad (2.19)$$

Since the interval $[x - \Delta x, x + \Delta x]$ is arbitrary, we have

$$c_t(x, t) + \phi_x(x, t) = f(x, t)$$

or, from (2.17),

$$c_t(x, t) + \lambda c_x(x, t) - \frac{1}{2} D c_{xx}(x, t) = f(x, t).$$

Since $c(x, t) = mnu(x, t)$, under the assumption of no sources or sinks ($f = 0$) and the assumption that $\lambda = 0$ (which can also be thought of as $p = q$), we have the familiar heat equation with initial and boundary conditions:

$$\begin{aligned} u_t - Du_{xx} &= 0, \quad 0 < x < L, \quad t > 0 \\ u(0, t) &= g(t), \quad t > 0 \\ u(L, t) &= h(t), \quad t > 0 \\ u(x, 0) &= f(x), \quad 0 < x < L \end{aligned} \quad (2.20)$$

The forward problem is to determine the temperature distribution $u(x, t)$ throughout the domain at time t from knowledge of $u(x, 0)$. If the endpoints of the bar are kept at zero temperature, we have $g(t) = h(t) = 0$, which we will henceforth take for simplicity. Problems with nonzero boundary conditions can be transformed to zero boundary conditions through a change of variables.

The backward problem is to determine the temperature distribution $u(x, t)$ at some prior time $t < T$ from knowledge of $u(x, T)$

$$u_t - Du_{xx} = 0, \quad 0 < x < L, \quad t > 0 \quad (2.21)$$

$$u(0, t) = 0, \quad t > 0$$

$$u(L, t) = 0, \quad t > 0$$

$$u(x, T) = m(x), \quad 0 < x < L \quad (2.22)$$

2.2.2 A finite difference discrete model

In the following section, we will look at the fundamental solution for the heat equation and discretize the integral equation for the solution of the backward problem. Here, we consider another elementary approach: a finite difference discretization with explicit time stepping.

Define a mesh on the spatial domain by

$$x_0 = 0, x_1 = \frac{L}{M+1}, x_2 = \frac{2L}{M+1}, \dots, x_{M+1} = L.$$

Then $\Delta x = 1/(M+1)$. Define a sequence of uniform time steps up to time T by

$$t_0 = 0, t_1 = \frac{T}{N+1}, t_2 = \frac{2T}{N+1}, \dots, t_{N+1} = T.$$

Then $\Delta t = 1/(N+1)$. Denote $u(x_i, t_j)$ by u_{ij} . The Taylor series expansion for $u(x + \Delta x, t)$ about (x, t) is

$$u(x + \Delta x, t) = u(x, t) + (\Delta x)u_x(x, t) + \frac{(\Delta x)^2}{2!}u_{xx}(x, t) + O((\Delta x)^3). \quad (2.23)$$

For a linear approximation to $u(x + \Delta x, t)$, the terms of order $(\Delta x)^2$ and higher are dropped. Solving for $u_x(x, t)$ in the linear approximation results in the *forward difference formula*

$$u_x(x, t) = \frac{1}{\Delta x}(u(x + \Delta x, t) - u(x, t)) + O(\Delta x)$$

or in subscript notation

$$(u_x)_{ij} \approx \frac{1}{\Delta x}(u_{i+1,j} - u_{ij}).$$

Similarly, in the time variable

$$(u_t)_{ij} \approx \frac{1}{\Delta t}(u_{i,j+1} - u_{ij}). \quad (2.24)$$

Alternatively, by evaluating the Taylor expansion with respect to t about (x, t) at $(x, t - \Delta t)$ one obtains the *backward difference formula*

$$(u_t)_{ij} \approx \frac{1}{\Delta t}(u_{ij} - u_{i,j-1}). \quad (2.25)$$

An approximation to the second spatial derivative of u can be obtained by adding (2.23) and the Taylor expansion in (2.23) evaluated at $(x - \Delta x, t)$ to obtain

$$(u_{xx})_{ij} \approx \frac{1}{(\Delta x)^2}(u_{i+1,j} - 2u_{ij} + u_{i-1,j}). \quad (2.26)$$

Combining the approximations (2.24) and (2.26), the finite difference discretization of the PDE (2.20) is

$$\frac{u_{i,j+1} - u_{i,j}}{\Delta t} = \frac{D}{\Delta x^2}(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}). \quad (2.27)$$

In the solution of the forward model, we step forward in time solving for $u_{i,j}$ for $i = 1, \dots, M$ and $j = 1, 2, \dots$ until we reach the desired time at which we want to compute the solution. An analysis of the error in making these approximations to the derivatives shows that the finite difference method will converge to the solution of the forward problem provided

$$0 < \frac{\Delta t}{(\Delta x)^2} < \frac{1}{2}. \quad (2.28)$$

Most texts on the numerical solution of PDE's contain a proof of this result. See, for example, [9]. The boundary conditions are assumed to be known, and they are included in the solution as follows:

$$u_{0,j} = g(j\Delta t) \quad (2.29)$$

$$u_{M+1,j} = h(j\Delta t) \quad (2.30)$$

The initial condition is included by setting

$$u_{i,0} = f(i\Delta x).$$

Now we can write the solution at time step t_{i+1} as a linear system as follows. For simplicity, let $d = \frac{D\Delta t}{(\Delta x)^2}$ and denote by A the banded matrix

$$A = \begin{bmatrix} 1-2d & d & 0 & \cdots & \cdots & 0 \\ d & 1-2d & d & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & d & 1-2d & d \\ 0 & \cdots & \cdots & 0 & d & 1-2d \end{bmatrix}.$$

Let \mathbf{u}^j denote the vector at the j th time step $\mathbf{u}^j = [u_{1,j}, \dots, u_{M,j}]^T$ and \mathbf{v}^j the vector at the j th time step $\mathbf{v}^j = [du_{0,j}, 0, \dots, 0, du_{M+1,j}]^T$. Provided the stability criterion (2.28) holds, the solution at the $(j+1)$ st time step is approximated by computing

$$\mathbf{u}^{j+1} = A\mathbf{u}^j + \mathbf{v}^j. \quad (2.31)$$

Example: Consider the forward problem

$$u_t - u_{xx} = 0, \quad 0 < x < \pi, \quad t > 0 \quad (2.32)$$

$$u(0, t) = 0, \quad t > 0 \quad (2.33)$$

$$u(L, t) = 0, \quad t > 0 \quad (2.34)$$

$$u(x, 0) = 10 \sin 2x, \quad 0 < x < \pi \quad (2.35)$$

One can show that the actual solution to the forward problem is $u(x, t) = 10e^{-4t} \sin 2x$. The relative errors for the solution computed out to times $T = 0.1, 0.2, 0.3, 0.4$ are given in Table 2.1. A plot of the evolution of the solution in time is found in Figure 2.8. Notice that the solution has decayed to nearly zero by time $T = 1$.

Time T	Δx	Δt	$\ u(x, T)\ _\infty$	Relative sup-norm error
0.1	0.0668	4.65e-4	6.67	2.24e-4
0.2	0.0806	0.002	4.49	0.0015
0.3	0.0668	0.0014	3.01	0.0016
0.4	0.1366	0.0037	2.01	0.0020

Table 2.1: Accuracy of the finite difference forward solver on the problem (2.32) – (2.35).

Data at time $T = 0.4$ was simulated using (2.31) with $\Delta x = 0.1366$ and $\Delta t = 0.0037$, which corresponds to $M + 1 = 24$ in our spatial discretization and $K = 108$ time steps. The non-noisy data and data with 2% noise is found in Figure 2.8.

2.2.3 Naïve reconstruction of the initial temperature

A naïve approach to the inverse problem of determining u at the previous time step would be to solve

$$\mathbf{u}^j = A^{-1}(\mathbf{u}^{j+1} - \mathbf{v}^j). \quad (2.36)$$

This casts the problem in a discrete form. Since the stepsize in time is limited by the stability criterion, it will take numerous time steps to reach the initial condition, but this can be achieved by iterating the method (2.36). The results of applying (2.36) and iterating backward in time with and without noisy data at $T = 0.4$ are shown in Figure 2.9. The results are displayed at the iterate at which the solution to the backward problem

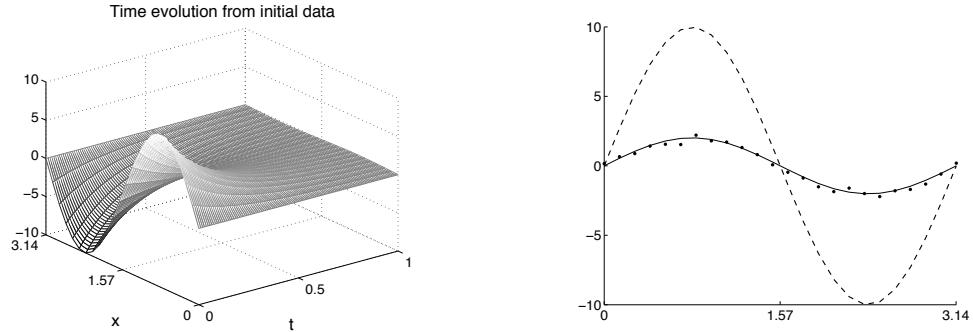


Figure 2.8: Left: illustration of the time evolution of the forward solution of the heat equation with initial condition $u_0(x) = 10 \sin 2x$ computed to time $T = 1$. Right: illustration of simulated measurement noise for the heat equation. The dashed line is the initial profile. The noise-free data at time $T = 0.4$ is the solid line. The computed solution corrupted with 2% white noise is plotted with large dots.

begins to become unstable. In the case of noise-free data, this occurs at approximately 22 backward steps, or at time $t = 0.3215$. However, the method is very sensitive to noise in the data, and with just 0.01% noise, it is only stable for approximately four backward steps, or $t = 0.3888$. In either case, a serious inverse crime is being committed here. The same method on the same mesh is being used to both generate and reconstruct the data, and the results are therefore better than they should be!

The approach (2.36) is not equivalent to using backward differences and time-stepping backward in that manner since A^{-1} is not equal to the matrix that arises from that approach. Let us next investigate that approach. Stepping backward in time from $j = K - 1$, we solve for each $u_{i,j}$ from

$$u_{i,j} = u_{i,j+1} - \frac{D\Delta t}{\Delta x^2}(u_{i+1,j+1} - 2u_{i,j+1} + u_{i-1,j+1}) \quad (2.37)$$

Plots of the results from four time steps backward with and without noisy data are found in Figures 2.10 and 2.11. We see that this method is somewhat more stable than (2.36), but it is still not useful for long times. In the computations resulting in Figure 2.10 an inverse crime is still being committed since the same mesh is used for the solution of the inverse problem as was used for the construction of the data by (2.31). By using a different mesh and time steps for the solution of (2.31) and (2.37) and interpolating the data to the mesh of (2.37), an inverse crime is avoided. This was the approach used in computing the results in Figure 2.11.

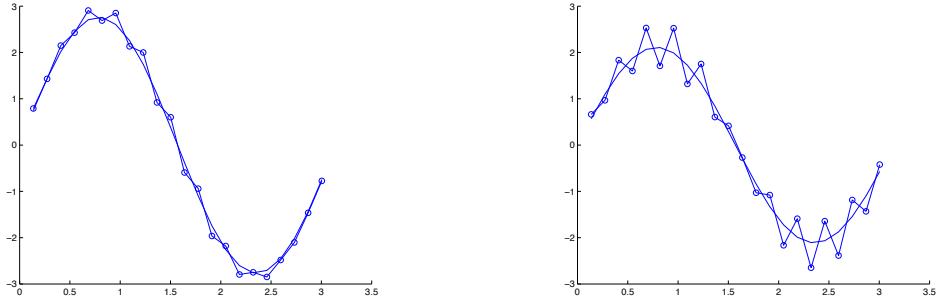


Figure 2.9: Left: reconstruction of the heat profile at time $t = 0.3215$ from noise-free data measured at $T = 0.4$. The solid line is the actual solution and the line with dots is the reconstruction computed by iterating (2.36) 22 steps. Right: reconstruction of the heat profile at time $t = 0.3888$ from data with 0.01% random noise measured at $T = 0.4$. The solid line is the actual solution and the line with dots is the reconstruction computed by iterating (2.36) 4 steps.

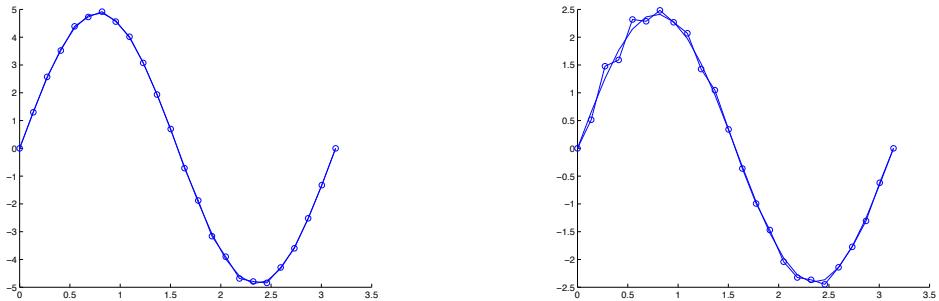


Figure 2.10: Left: reconstruction of the heat profile at time $t = 0.1794$ from noise-free data measured at $T = 0.4$. The solid line is the actual solution and the line with dots is the reconstruction computed by iterating (2.37) 60 steps. Here, an inverse crime was still committed since the same mesh and time steps were used to construct the data and compute the solution. The results are undeservingly good. Right: reconstruction of the heat profile at time $t = 0.3551$ from data with 0.01% random noise measured at $T = 0.4$. The solid line is the actual solution and the line with dots is the reconstruction computed by iterating (2.37) 13 steps. The same inverse crime is committed here.

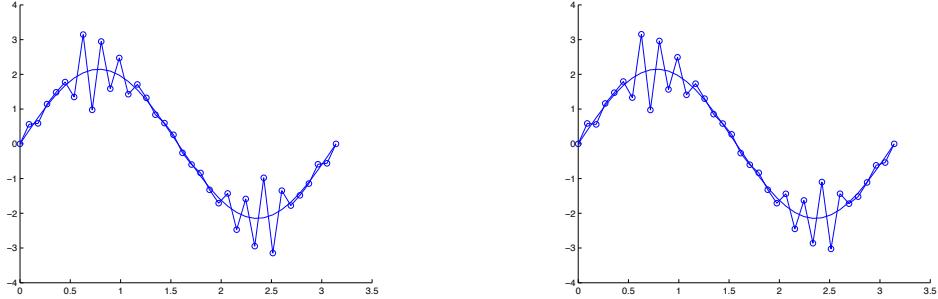


Figure 2.11: Left: reconstruction of the heat profile at time $t = 0.3844$ from noise-free data measured at $T = 0.4$. The solid line is the actual solution and the line with dots is the reconstruction computed by iterating (2.37) 8 steps. Equation (2.31) was used to compute the data on a different mesh from (2.37). Right: reconstruction of the heat profile at time $t = 0.3844$ from data with 0.01% random noise measured at $T = 0.4$. The solid line is the actual solution and the line with dots is the reconstruction computed by iterating (2.37) 8 steps.

Exercise 2.2.1 Complete Table 3.2 for later times $T = 0.5, 0.6, \dots, 1.0$. Observe what happens to the solution when the stability criterion is violated.

Exercise 2.2.2 Compute reconstructions by the methods in this section for the same example but use data at time $T = 0.1$. Do not commit any inverse crimes. Is the method more stable than from the final time data at $T = 0.4$?

Exercise 2.2.3 Study the numerical forward solution of the heat equation by adding noise to the initial data $u_0(x) = 10 \sin 2x$ in the MATLAB program `InverseHeatCondDataSimulator.m` and computing the solution at time $T = 0.4$. Plot the difference between the computed $u(x, T)$ from a noisy and non-noisy initial condition. How does this differ from what we see in Figures 2.9 and 2.11?

Exercise 2.2.4 Use the MATLAB program `InvHeatCondNaiveSolver.m` and modify it to use the initial temperature distribution $f(x) = 10\chi_{[\pi/4, 3\pi/4]}(x)$, where χ is the characteristic function. Compute non-noisy and noisy simulated data, and compute naive reconstructions at four prior time steps using methods (2.36) and (2.37). Plot your results. Give the reason that the discontinuity in the initial condition can never be reconstructed.

2.3 Tomographic X-ray projection data

In tomographic X-ray imaging one takes X-ray projection images of an object from several different directions and attempts to recover the inner structure of the object from the data. We show how such a measurement can be written in the form $\mathbf{m} = \mathbf{Af} + \varepsilon$, and illustrate numerically how the naïve reconstruction approach (1.4) fails.

2.3.1 A simple example: probing two aluminum slabs

Let us first demonstrate the exponential attenuation law of X-rays using a very simple example, where two aluminum slabs are probed as shown in Figure 2.12. Typically, X-rays emanate from a roughly point-like location inside an X-ray tube. That point is called the *X-ray source* and shown as a black dot in the figures below. Three X-rays are sent traveling towards a detector, each consisting initially of 1000 photons. The detector is capable of counting how many photons arrive at each point. One of the rays arrives at the detector through empty space, delivering all 1000 photons. Another ray travels through an aluminum slab whose width is chosen to be the *half-thickness* of the X-radiation used here. This means that half of the photons entering the slab will be absorbed inside the slab. The third ray encounters two such aluminum slabs. We call these three X-rays the empty-space ray, the one-slab ray, and the two-slab ray, respectively.

The photon count data can now be transformed into line integral data via two simple steps. First, take the logarithm of each photon count. Then, realizing that the integral of the empty-space ray must be zero, subtract each logarithm from the logarithm corresponding to the empty-space ray. As seen from the actual numbers shown in Figure 2.12, the resulting attenuation data is zero for the empty-space ray, a positive number (0.693) for the one-slab ray, and twice that number (1.386) for the two-slab ray.

We have described the basic calibration process for ideal photon count data based on the exponential attenuation law. However, we ignored at least a couple of properties of real-world measurements. First, practical detectors (for instance charge-coupled devices, or CCD's) do not provide the actual photon count but rather an integer that is proportional to the photon count. However, this is not a serious problem as you can find out in Exercise 2.3.2 below. Second, the photon count is not a deterministic number; it is better modeled as a random variable with Poisson distribution. This results in random measurement noise in the data; we will discuss this below.

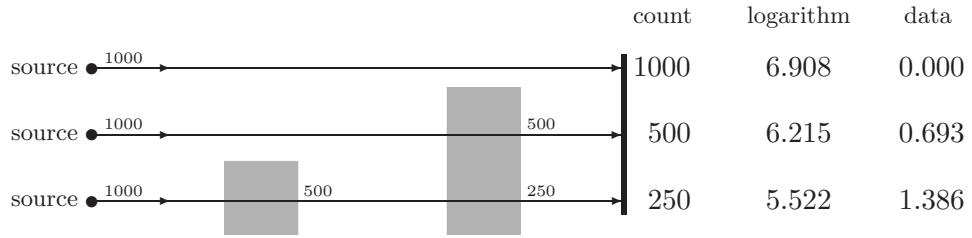


Figure 2.12: Simple experiment illustrating the attenuation of X-rays and interpretation of measurements. The three black dots show the positions of the X-ray source, and the horizontal lines depict X-rays. The gray boxes are slabs of attenuating material, and their width has been chosen to be the half-thickness of the X-radiation. The vertical thick line is the detector counting how many photons arrive at each point.

2.3.2 From photon count data to line integral data

The two-slab example in Section 2.3.1 is quite simple as it concerns only homogeneous material. Consider now an X-ray traveling through a phantom¹ representing a two-dimensional cross-section of a patient’s head along a straight line as shown in the left panel of Figure 2.13. We place the target slice inside the unit square defined by $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 1$. For the sake of the argument, assume that the X-ray travels along the horizontal path defined by $0 \leq x_1 \leq 1$ and $x_2 = \frac{1}{2}$.

Interaction between radiation and matter lowers the intensity of the ray. We think of the X-ray having initial intensity $I_0 := I(0)$ when entering the patient’s head and smaller intensity $I_1 := I(1)$ when exiting. Also, we denote by $I(x_1)$ the intensity of the X-ray at the point $(x_1, \frac{1}{2})$ while traveling from the source to the detector.

In contrast to the simple homogeneous slab example above, the cross-section of a head contains various tissues with different X-ray attenuation properties. We model this situation using a non-negative attenuation coefficient function $f(x_1, x_2)$, whose value gives the relative intensity loss of the X-ray within a small distance dx :

$$\frac{dI(x_1)}{I(x_1)} = -f(x_1, \frac{1}{2})dx_1.$$

¹A *phantom* can be either a physical calibration device or a mathematical model. Here it is a mathematical model to simulate an idealized cross-section of a human head. The use of “phantom” will be clear from the context.

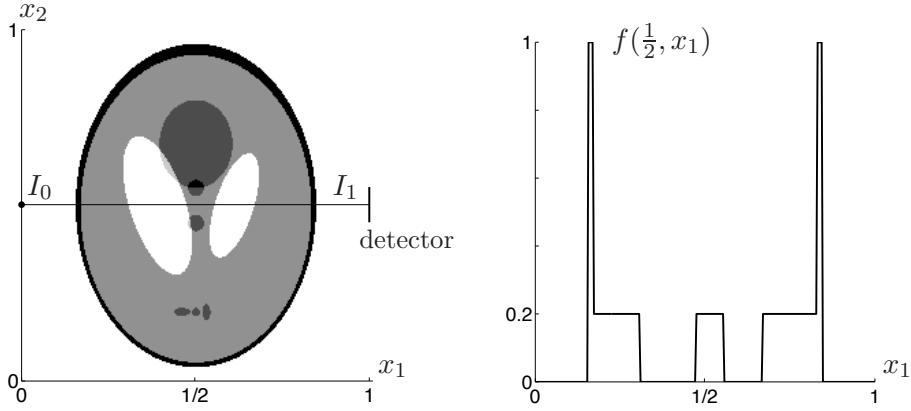


Figure 2.13: X-ray measurement. Left: an X-ray traveling through a simulated cross-section of a human head (a low-contrast version of the infamous Shepp-Logan phantom). Note that high attenuation is shown here as darker shade of gray and low attenuation as lighter shade. Right: plot of the attenuation coefficient along the path of the X-ray.

For example, bone has higher attenuation coefficient than brain tissue, and cerebrospinal fluid (white ovals in the left panel of Figure 2.13) provides practically zero attenuation. See the right panel in Figure 2.13 for a plot of the profile $f(x_1, \frac{1}{2})$.

Integration along the X-ray from source to detector gives

$$\int_0^1 f(x_1, \frac{1}{2}) dx_1 = - \int_0^1 \frac{I'(x_1)}{I(x_1)} dx_1 = \log I_0 - \log I_1. \quad (2.38)$$

Now the right hand side of (2.38) is known: I_0 by calibration and I_1 from the measurement. The left hand side of (2.38) consists of an integral of the unknown function f over a straight line, as wished.

Regarding noise, the quantity I_1 is a constant multiple of a Poisson-distributed random variable. It is typically sampled in practice using an analog-to-digital converter that produces integer output containing truncation errors and additional electronic noise. Taking logarithm of I_1 leads to a random variable with remarkably complicated statistics. However, it is usually quite plausible to model the measurement as

$$\log I_0 - \log I_1 = \int_0^1 f(x_1, \frac{1}{2}) dx_1 + \varepsilon, \quad (2.39)$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed random variable. The standard deviation σ of the noise can be estimated for example by measuring the same target repeatedly and calculating the standard deviation of the samples. This procedure is a reasonably accurate model when the photon count is large enough, see [413, Appendix].

We remark that in the above model we neglect the energy dependence of the attenuation function. Namely, most X-ray sources produce a multi-spectral beam, and an energy-dependent f may result in different measured line integrals depending on the propagation direction of the X-ray along the line. This is called *beam hardening*.

2.3.3 Continuous tomographic data: the Radon transform

In the previous section we described how to turn attenuation data from one single X-ray into line integral data concerning a nonnegative, compactly supported attenuation coefficient $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. The aim of tomographic imaging is to collect information about f using different angles of view.

Let us define the *Radon transform*, denoted by \mathfrak{R} , as follows. We interpret $\theta \in \mathbb{R}$ as an angle measured in radians, and denote by

$$\vec{\theta} := \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \in \mathbb{R}^2$$

the unit vector with angle θ with respect to the x_1 -axis. The Radon transform of the function f depends on the angular parameter θ and on a linear parameter $s \in \mathbb{R}$ in the following way:

$$\mathfrak{R}f(s, \theta) = \int_{x \cdot \vec{\theta}=s} f(x) dx^\perp, \quad (2.40)$$

where dx^\perp denotes the 1-dimensional Lebesgue measure along the line defined by $\{x \in \mathbb{R}^2 : x \cdot \vec{\theta} = s\}$. We remark that the parametrization of tomographic data provided by formula (2.40) is related to the so-called parallel beam geometry used in the first-generation CT scanners in the 1970's.

Many variations in the data geometry are possible, such as limited angle data, local tomography data, exterior tomography, and combinations thereof. We refer the reader to the classical texts by Natterer [351] and Kak and Slaney [248]. See [351] for analytic inversion formulae, a thorough analysis of the mapping properties of \mathfrak{R} and its generalizations to higher dimensions. See [352] for another perspective on image reconstruction.

The Fourier transform and Radon transform are connected in a simple way. This result is known as the Central Slice Theorem. First, define the Fourier transform in one dimension by

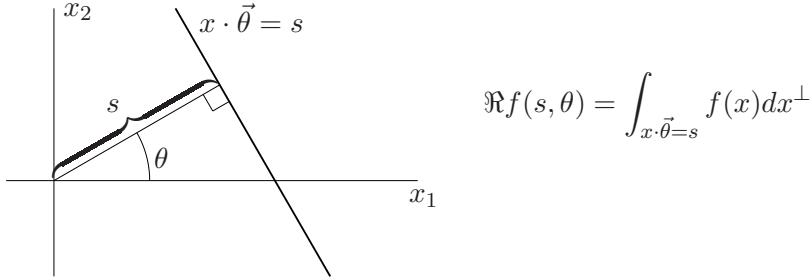


Figure 2.14: Illustration of the definition (2.40) of the Radon transform. The line $\{x \in \mathbb{R}^2 : x \cdot \vec{\theta} = s\}$ is drawn with a thick line.

Definition 2.3.1 *The Fourier Transform of a function defined on \mathbb{R} is given by*

$$\mathcal{F}(x)(\xi) = \hat{f}(\xi) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}} f(x) e^{-ix\xi} dx$$

Theorem 2.3.1 *Let f be an absolutely integrable function defined on the whole real line. For any real number r and unit vector $\vec{\theta}$, we have the identity*

$$\int_{-\infty}^{\infty} \Re f(s, \vec{\theta}) e^{-isr} ds = \hat{f}(r\vec{\theta}). \quad (2.41)$$

Proof: By the definition of the Radon transform

$$\begin{aligned} \int_{-\infty}^{\infty} \Re f(s, \vec{\theta}) e^{-isr} ds &= \int_{-\infty}^{\infty} \int_{x \cdot \vec{\theta}=s} f(x) e^{-isr} dx^\perp ds \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) e^{-ix \cdot (r\vec{\theta})} dx_1 dx_2 \\ &= \hat{f}(r\vec{\theta}). \end{aligned}$$

□

It will prove convenient to have a notation for the one-dimensional Fourier transform of a function in the scalar parameter as appears in the Central Slice Theorem. Let $\tilde{h}(s, \vec{\theta})$ denote such a Fourier transform:

$$\tilde{h}(s, \vec{\theta}) = \int_{-\infty}^{\infty} h(t, \vec{\theta}) e^{-its} dt. \quad (2.42)$$

Then the Central Slice Theorem says

$$\widetilde{\Re f}(r, \vec{\theta}) = \hat{f}(r\vec{\theta}).$$

The Radon inversion formula provides a way to obtain f from its Radon transform in the ideal case.

Theorem 2.3.2 *If f is an absolutely integrable function defined on the real line and \hat{f} is absolutely integrable, then*

$$f(x) = \frac{1}{(2\pi)^2} \int_0^\pi \int_{-\infty}^\infty e^{isx \cdot \vec{\theta}} \widetilde{\Re f}(s, \vec{\theta}) |s| ds d\theta. \quad (2.43)$$

Proof: First note that since the Radon transform satisfies $\Re f(-s, -\vec{\theta}) = \Re f(s, \vec{\theta})$,

$$\begin{aligned} \widetilde{\Re f}(-s, -\vec{\theta}) &= \int_{-\infty}^\infty \Re f(t, -\vec{\theta}) e^{-it(-s)} dt \\ &= \int_{-\infty}^\infty \Re f(t, -\vec{\theta}) e^{-i(-t)s} dt \\ &= \int_{-\infty}^\infty \Re f(-t, -\vec{\theta}) e^{-its} dt \\ &= \int_{-\infty}^\infty \Re f(t, \vec{\theta}) e^{-its} dt \\ &= \widetilde{\Re f}(s, \vec{\theta}). \end{aligned}$$

Now by the Fourier inversion formula, with $\xi = (r \cos \theta, r \sin \theta)$

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\xi) e^{ix \cdot \xi} d\xi \\ &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\infty \hat{f}(r\vec{\theta}) e^{irx \cdot \vec{\theta}} r dr d\theta \\ &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\infty \widetilde{\Re f}(r, \vec{\theta}) e^{irx \cdot \vec{\theta}} r dr d\theta \\ &= \frac{1}{(2\pi)^2} \int_0^\pi \int_{-\infty}^\infty \widetilde{\Re f}(r, \vec{\theta}) e^{irx \cdot \vec{\theta}} |r| dr d\theta, \end{aligned}$$

where the last equality follows from the fact that $\widetilde{\Re f}(-s, -\vec{\theta}) = \widetilde{\Re f}(s, \vec{\theta})$. \square

To summarize, this results in the following idealized reconstruction algorithm for x-ray CT imaging:

- Let f be the attenuation coefficient of a 2-D slice of a 3-D object. Then the intensity $I_{(s, \vec{\theta})}$ of the beam satisfies the differential equation

$$\frac{dI_{(s, \vec{\theta})}}{I_{(s, \vec{\theta})}} = -f(s, \vec{\theta}) ds$$

- We measure the Radon transform of f

$$\Re f(s, \vec{\theta}) = \log \left(\frac{I_0}{I_d} \right)$$

where I_0 is the intensity of the beam at the source, and I_d is the intensity of the beam at the detector.

- Reconstruct f from the Radon Inversion Formula (2.43).

For the Filtered Back-projection algorithm, we regard the radial integral in the Radon Inversion Formula as a filter. We denote the output of the filter by $\mathcal{G}\Re f(t, \vec{\theta})$ where

$$\mathcal{G}\Re f(t, \vec{\theta}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widetilde{\Re f}(r, \vec{\theta}) e^{irt} |r| dr.$$

Then, with $t = x \cdot \vec{\theta}$,

$$f(x) = \frac{1}{2\pi} \int_0^\pi \mathcal{G}\Re f(x \cdot \vec{\theta}, \vec{\theta}) d\theta.$$

Note that one sees from this formula that low-frequency components are suppressed by $|r|$ and high-frequency components are amplified. Let's look at the filter a little more carefully. Recall that the Fourier transform of $g'(t)$ is

$$\mathcal{F}(\partial_t g)(\xi) = i\xi \hat{g}(\xi).$$

Thus if we had r instead of $|r|$ in the Radon inversion formula, we would have had the

$$\text{"inversion formula"} = \frac{1}{2\pi i} \int_0^\pi \partial_r \Re f(r, \theta) d\theta$$

If f is real-valued, this quantity is purely imaginary! Thus, the $|r|$ is very important!

The MATLAB function `iradon.m` in the Image Processing Toolbox implements filtered back-projection. In the subsequent sections, we will be comparing the results of filtered back-projection implemented with `iradon.m` to other inversion techniques.

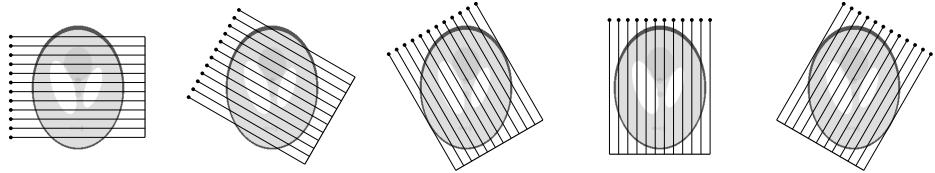


Figure 2.15: Parallel beam X-ray measurement geometry. Here $J = 5$ and $N = 11$. Black dots show the locations of the X-ray source at different times of measurement. The thick line represents the detector measuring the intensity of the X-rays after passing through the target. High attenuation is shown here as darker shade of gray and low attenuation as lighter shade.

2.3.4 Discrete tomographic data

We model practical tomographic X-ray data by a bounded set $\Omega \subset \mathbb{R}^2$, a non-negative attenuation coefficient f supported in $\overline{\Omega}$, and some finite collection $\{L_j\}_{j=1}^k$ of lines $L_j \subset \mathbb{R}^2$ intersecting Ω .

As a first example, we will use the following data set. It is an example of *parallel beam geometry* illustrated in Figure 2.15. The angular variable is sampled with equidistant steps over the half circle:

$$\theta_j = \theta_1 + \left(\frac{j-1}{J}\right)\pi, \quad 1 \leq j \leq J, \quad (2.44)$$

where $\theta_1 \in \mathbb{R}$ is an appropriate constant, a reference angle. The linear parameter s is also sampled uniformly over a suitable interval:

$$s_\nu = -S + 2\left(\frac{\nu-1}{N}\right)S, \quad 1 \leq \nu \leq N, \quad (2.45)$$

where $S > 0$.

Defining $k = JN$, the measurement (1.1) then takes the form

$$\mathbf{m} = \mathcal{A}f + \varepsilon = \begin{bmatrix} \int_{L_1} f(x_1, x_2) ds_1 \\ \vdots \\ \int_{L_k} f(x_1, x_2) ds_k \end{bmatrix} + \varepsilon, \quad (2.46)$$

where ds_j denotes the 1-dimensional Lebesgue measure along the line L_j . Each integral in (2.46) can be understood as a suitable rotation and scaling of formula (2.39).

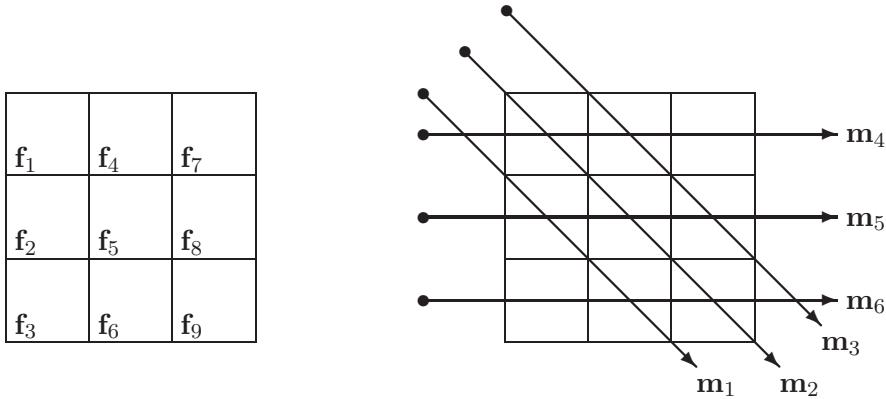
For the computational solution we need to build a finite-dimensional measurement model of the form (1.3). We discretize the tomographic problem by dividing the unknown area into n pixels and assume that attenuation values are constant within each pixel. We number the pixels from 1 to n and call the corresponding attenuation values $\mathbf{f}_j \geq 0$ for $j = 1, \dots, n$.

The measurement \mathbf{m}_i of the line integral of f over line L_i is approximated by

$$\mathbf{m}_i = \int_{L_i} f(x_1, x_2) ds \approx \sum_{j=1}^n a_{ij} \mathbf{f}_j, \quad (2.47)$$

where a_{ij} is the distance that L_i travels in the j th pixel. Note that only pixels that intersect the beam L_i are included in this sum. Further, if we have k measurements in the vector $\mathbf{m} \in \mathbb{R}^k$, then (2.47) yields a matrix equation $\mathbf{m} = A\mathbf{f}$, where the matrix is defined by $A = (a_{ij})$.

Consider the following discretization and measurements, where $J = 2$, $k = 6$, $N = 3$ and the total number of pixels is $N^2 = 9$:



Here we have divided the square-shaped domain $\Omega \subset \mathbb{R}^2$ into 9 pixels, denoted by thin lines. The length of the side of each pixel is 1. Inside the pixels there is a constant value \mathbf{f}_j of attenuation. The six arrows are X-rays used for probing the inner structure of Ω . Measurement data is the vector $\mathbf{m} = [\mathbf{m}_1, \dots, \mathbf{m}_6]^T$ modelled by (2.47). The resulting measurement model

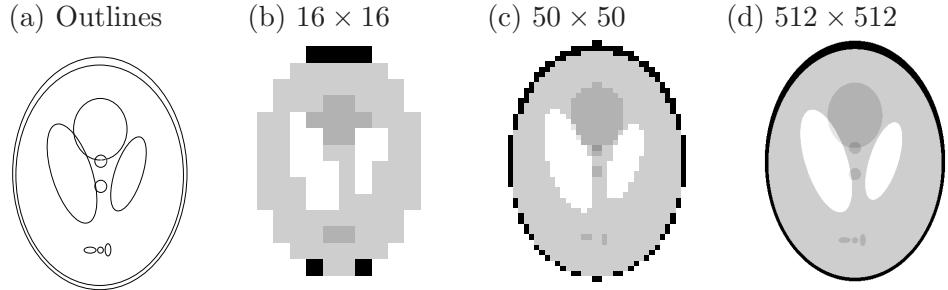


Figure 2.16: (a) Schematic illustration of the Shepp-Logan phantom. Areas of different attenuation values are bounded by various ellipses. (b)–(d) Plots of grayscale images of the Shepp-Logan phantom at different resolutions. High attenuation is shown here as darker shade of gray and low attenuation as lighter shade.

is

$$\begin{bmatrix} 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 \\ \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix}. \quad (2.48)$$

The model (2.48) is low-dimensional and simple. However, it already demonstrates one feature typical for inverse problems: nonuniqueness of solution. Namely, as can be seen in Exercise 2.3.5, there are several targets that produce exactly the same data. Thus the inverse problem cannot be uniquely solved using the measurement information alone.

Let us build a more realistic (higher-dimensional) data simulation model. We work with the so-called Shepp-Logan phantom, which is a piecewise constant model of a cross-section of a human head. The phantom is defined using ellipses and can be realized at any desired discrete resolution. See Figure 2.16 for pictures of the Shepp-Logan phantom at discretizations with $16 \times 16 = 256$ pixels, $50 \times 50 = 2500$ pixels and $512 \times 512 = 262144$ pixels.

Let us construct the measurement matrix A corresponding to the low-resolution case with 16×16 pixels and projection directions specified by

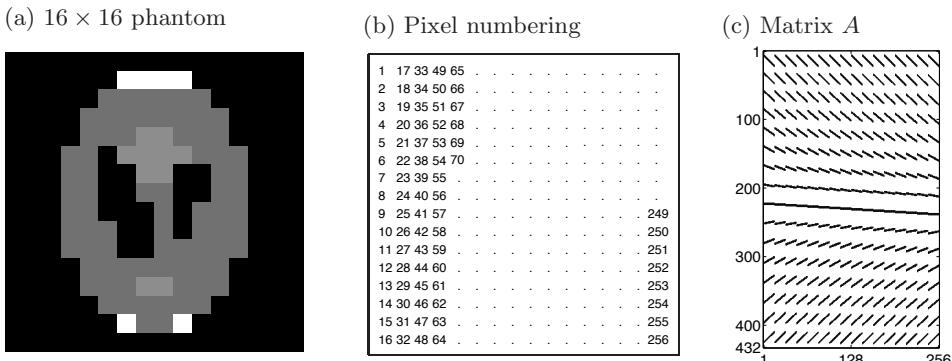


Figure 2.17: (a) Low-resolution Shepp-Logan phantom (16×16). Here black denotes zero attenuation, and white denotes maximum attenuation. (b) Numbering of the pixels in a 16×16 image when interpreted as a vector in \mathbb{R}^{256} . (c) Nonzero elements of the 432×256 tomographic measurement matrix.

taking $J = 16$ in (2.44). In this case the size of A is not too large and we can show a picture of the nonzero elements of A for observing its structure.

We use MATLAB's command `radon.m` to simulate parallel beam X-ray projection data from 16×16 pixel images with zero entries except one pixel with value 1. The pixel value 1 is first located in pixel 1 in the numeration shown in Figure 2.17(b), then in pixel 2, and so on. This way, column by column, we construct a measurement matrix A for a computational tomography model of the form (1.3).

What is the size of A ? Obviously, the number of columns must be 256, the number of pixels in a 16×16 image. The `radon.m` algorithm picked automatically the value $N = 27$ in formula (2.45), so A has $JN = 16 \cdot 27 = 432$ rows. We observe that there are 11086 nonzero elements out of the total 110592; this means that roughly 90% of the elements in A are zero.

Three-dimensional X-ray tomography problems can be approached similarly to the above explanation using voxelization instead of pixelization and by tracing the paths of X-rays through the voxels in a three-dimensional manner. However, for illustration and simplicity purposes we stick to the two-dimensional case in this book. In principle there is no essential difference between the two- and three-dimensional cases, only the computations will be significantly more demanding in 3D.

2.3.5 Naïve reconstruction

The 16×16 Shepp-Logan phantom used in Section 2.3.4 has too low resolution to really show the intended anatomic features properly. In the rest of the book we work with the 50×50 Shepp-Logan phantom when we need to construct the matrix A explicitly, and with the 512×512 phantom when we illustrate matrix-free large-scale methods in Section 9. See Figure 2.16.

Next we wish to experiment with naïve reconstructions of the 50×50 Shepp-Logan phantom shown in Figure 2.16(c). We choose the number of projection directions to be $J = 50$ in formula (2.44). We construct the measurement matrix A column by column as explained in Section 2.3.4. Matlab's `radon.m` algorithm picked automatically the value $N = 75$ in formula (2.45), so A has $JN = 50 \cdot 75 = 3750$ rows. We arrive at the following measurement model:

$$A \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_{2500} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_{3750} \end{bmatrix}, \quad (2.49)$$

where the elements of the 50×50 pixel image \mathbf{f} and the elements of the 75×50 sinogram \mathbf{m} are numbered similarly to Figure 2.17(b).

Once A is in place, we can try out naïve inversion, but not in the sense of (1.4) since A is not a square matrix. Instead we use least-squares naïve inversion defined as follows:

$$\mathbf{f} \approx (A^T A)^{-1} A^T \mathbf{m}. \quad (2.50)$$

Derivation and interpretation of formula (2.50) is postponed to Section 5.2. The result of applying (2.50) to ideal (non-noisy) tomographic data is shown in Figure 2.19(b), and it looks very good indeed. The relative error of this reconstruction is very small. Perhaps we can conclude that we succeeded in reconstructing the 50×50 phantom from indirect tomographic measurements?

Before jumping to such a (wrong) conclusion, let us see what a small amount of measurement noise does to the naïve reconstruction. We add white noise of relative amplitude 0.1% to the sinogram and try formula (2.50) again. The result is shown in 2.19(c), and it consists merely of numerical garbage. This shows that formula (2.50) is not practically useful since real measurements always contain noise.

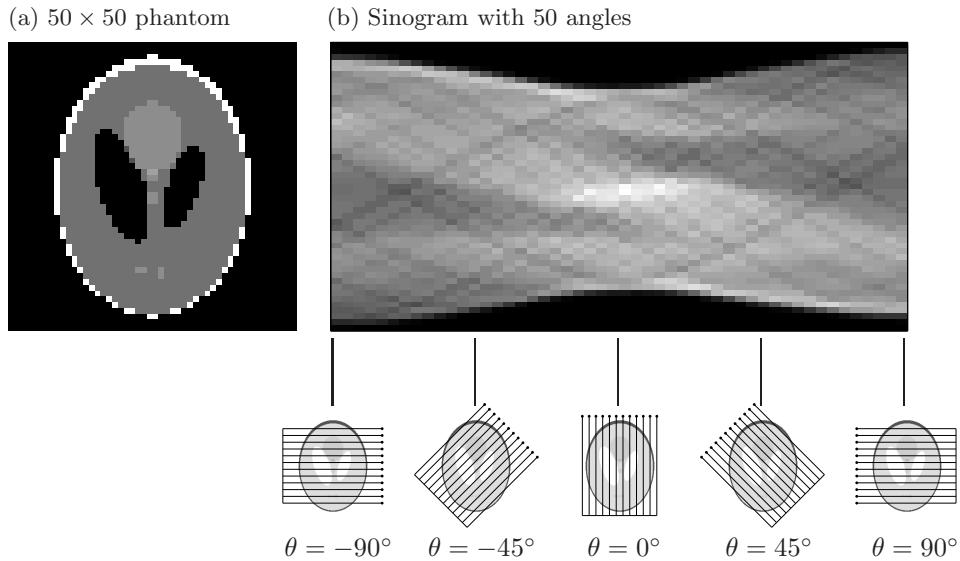


Figure 2.18: (a) Shepp-Logan phantom 50×50 . Here black denotes zero attenuation, and white denotes maximum attenuation. (b) Measured data (involving inverse crime) in sinogram form, where horizontal axis is the angle θ and vertical axis the variable s in (2.40). We have removed some purely zero rows from the top and bottom of the sinogram for clarity. Underneath the sinogram we show some of the projection directions to illustrate the structure of the sinogram.

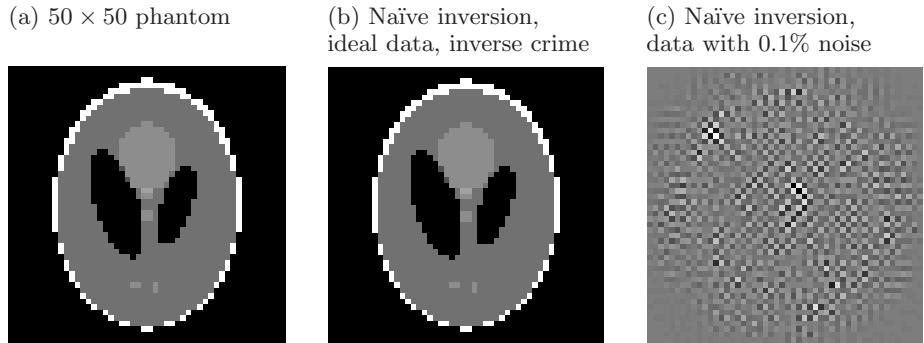


Figure 2.19: (a) Shepp-Logan phantom at resolution 50×50 . (b) Result of naïve inversion (2.50) from non-noisy data. The seemingly successful result is not to be trusted because an inverse crime was committed. (c) Result of naïve inversion (2.50) from data contaminated by 0.1% noise. The much worse performance of (c) compared to (b) indicates that naïve inversion is not stable with respect to noise.

2.3.6 Naïve reconstruction without inverse crime

We wish to avoid the inverse crime evident in Figure 2.19. To this end, we interpolate our data from tomographic data simulated using the Shepp-Logan phantom on a twice finer grid (100×100) than the grid used in the naïve reconstruction. This can be done conveniently as the phantom is defined analytically using ellipses, so it can be evaluated with arbitrary resolution. The measurement angles are the same. See Figure 2.20 for plots of the data.

Figure 2.21 shows the result of applying naïve reconstruction to the crime-free data. Now the result has unacceptable quality even when there is no added noise.

Exercise 2.3.1 Let $\mathbf{f} \in \mathbb{R}^8$ be a signal and $\mathbf{p} = [\mathbf{p}_{-1} \ \mathbf{p}_0 \ \mathbf{p}_1]^T$ a point spread function. Write down the 8×8 matrix A modeling the one-dimensional convolution (2.11) with the assumption that $\mathbf{f}_{j-\ell} = 0$ for the cases $j - \ell < 1$ and $j - \ell > 8$.

Exercise 2.3.2 Assume that an X-ray detector provides the proportional number Mc instead of the actual photon count c . Here $M > 0$ is a positive constant. Show that the calibration procedure described in Figure 2.12 works fine even if M is unknown.

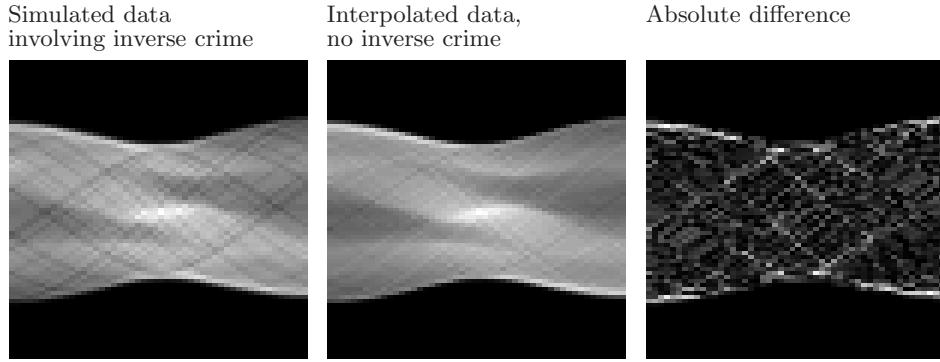


Figure 2.20: Tomographic data with and without inverse crime. Left: ideal data obtained by applying the measurement model matrix to the Shepp-logan phantom at the final reconstruction resolution of 50×50 pixels. Middle: tomographic data computed from 100×100 Shepp-logan phantom (at same measurement angles but finer arrangement of X-rays) and interpolated to lower resolution. Right: absolute difference between the two data sets.

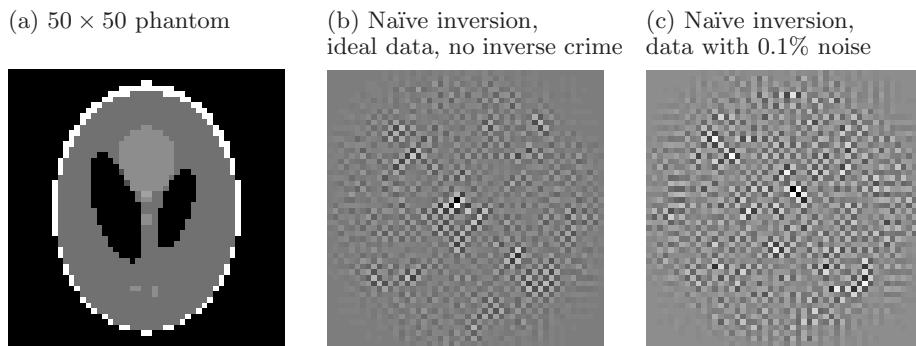


Figure 2.21: (a) Shepp-Logan phantom at resolution 50×50 . (b) Result of naïve inversion (2.50) from non-noisy data with no inverse crime. (c) Result of naïve inversion (2.50) from data contaminated by 0.1% noise. Compare to Figure 2.19.

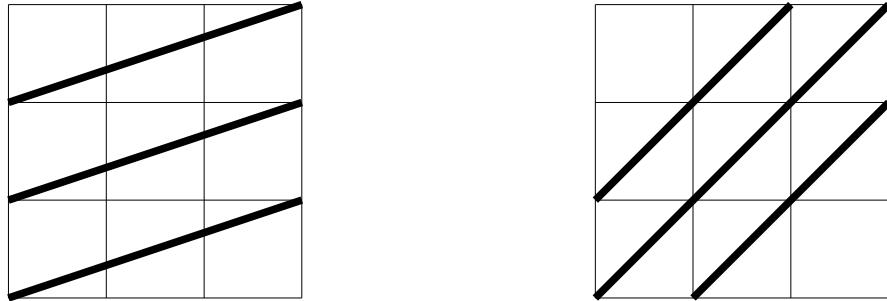


Figure 2.22: Tomographic X-ray measurement configuration related to Exercise 2.3.4. Thick lines depict X-rays, and the length of the side of a pixel is one.

Exercise 2.3.3 Define f and I_0 and I_1 appropriately in the context of the simple example shown in Figure 2.12. Furthermore, describe the measurement data in Figure 2.12 in terms of formula (2.38).

Exercise 2.3.4 In Figure 2.22, thin lines depict pixels and thick lines X-rays. Give a numbering to the nine pixels ($\mathbf{f} \in \mathbb{R}^9$) and to the six X-rays ($\mathbf{m} \in \mathbb{R}^6$), and construct the matrix A for the measurement model $\mathbf{m} = A\mathbf{f}$. The length of the side of a pixel is one.

Exercise 2.3.5 Show that the following targets produce exactly the same data in the measurement model (2.48):

4	4	5
1	3	4
1	0	2

5	6	2
1	5	2
4	0	-1

(a) What's wrong with the negative value -1 above? (b) Can you find more examples that produce the same data but only have non-negative entries?

Chapter 3

Ill-Posedness in Inverse Problems

Inverse problems are characterized by ill-posedness, in other words extreme sensitivity to measurement noise and modeling errors. In this chapter we will look at ill-posedness of the infinite-dimensional inverse problem, which requires some knowledge of operator theory, and ill-posedness of the finite-dimensional inverse problem which typically arises as a discretization of an infinite-dimensional inverse problem. For the necessary background on Banach spaces and compact linear operators, see Appendix B of this text or the references supplied there. We introduce the concept of regularization and the singular value decomposition.

3.1 Forward map and Hadamard's conditions

The examples in Sections 2.1–2.3 should convince the reader that there is something suspicious going on with the inversion of those three simple and indirect measurements. The reason for the observed instability is *ill-posedness*, which we discuss next.

The core of any inverse problem is the *forward map* $\mathcal{A} : \mathcal{D}(\mathcal{A}) \rightarrow Y$, a mathematical model of the corresponding direct problem. Here X and Y are suitable Hilbert spaces called *model space* and *data space*, respectively, and the subset $\mathcal{D}(\mathcal{A}) \subset X$ is the domain of definition of the bounded linear operator \mathcal{A} . The forward map is used as a mathematical model of the indirect measurement

$$m = \mathcal{A}f + \varepsilon. \quad (3.1)$$

Here $f \in \mathcal{D}(\mathcal{A}) \subset X$ is the quantity of interest, $m \in Y$ is measurement data,

and ε is noise satisfying $\|\varepsilon\|_Y \leq \delta$ with some known $\delta > 0$. (Here we denote the measurement by m instead of the vector notation \mathbf{m} because we want to cover more general data spaces than just $Y = \mathbb{R}^k$.)

Given a particular inverse problem, it is not necessarily straightforward to choose the spaces X and Y and the forward map \mathcal{A} . Constructing them must be considered as a nontrivial mathematical modeling task. The following aspects need to be modeled: physical processes involved, technical properties of the measurement device, geometry of the measurement, and possible limitations in data sets.

According to Hadamard, a solution method is called *well-posed* if the following three conditions are satisfied:

H_1 : **Existence.** There should be at least one solution.

H_2 : **Uniqueness.** There should be at most one solution.

H_3 : **Stability.** The solution must depend continuously on data.

We need to study the well-posedness of the naïve inversion $\mathcal{A}^{-1}m$ as a solution to the inverse problem “Given m , find f ”.

If the forward map is bijective from X to Y and allows a continuous inverse \mathcal{A}^{-1} , then naïve inversion satisfies all conditions H_1 – H_3 and we are dealing with a well-posed inverse problem.

This book is about *ill-posed inverse problems*, defined as the complement of well-posed problems: at least one of the conditions H_1 – H_3 must fail for naïve inversion $\mathcal{A}^{-1}m$. Condition H_1 is violated if the measured noisy data does not belong to the range of the forward map: $\mathcal{A}f + \varepsilon \notin \mathcal{A}(\mathcal{D}(\mathcal{A}))$. Condition H_2 fails if two quantities $f, g \in \mathcal{D}(\mathcal{A})$ give the same measurement: $\mathcal{A}f = \mathcal{A}g$, leading to non-uniqueness. The forward map does not always allow a continuous inverse, not even when restricted to the range $\mathcal{A}(\mathcal{D}(\mathcal{A}))$, so H_3 does not hold. See Figure 3.1 for a schematic illustration of the forward map and the related definitions.

3.2 Ill-posedness of the backward heat equation

The first effort related to the numerical solution of the backward heat equation was by Fritz John [239] for the problem in one spatial dimension on the whole real line. In this work he also showed that the problem is ill-posed in the sense that the solution does not depend continuously on the data. In [239] he considered the problem on $(-\infty, \infty)$ and expressed the time-dependent solution in terms of the fundamental solution. Taking the

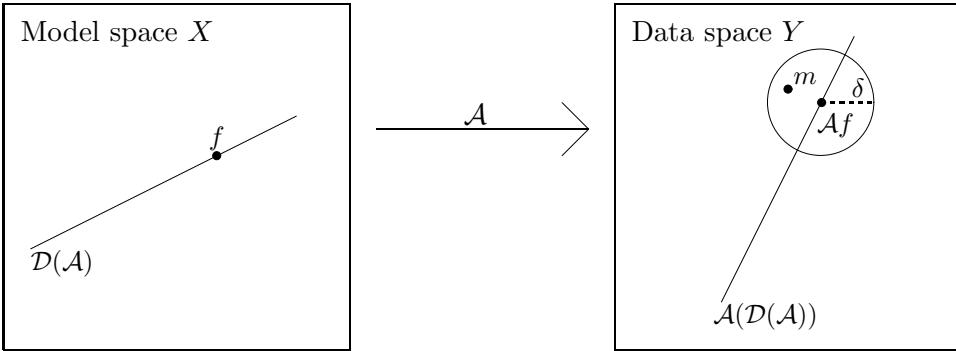


Figure 3.1: Schematic illustration of the linear forward map \mathcal{A} . The radius δ indicates the known maximum amplitude of measurement noise.

same approach for (2.20) leads to an analytic expression for the solution that can be used to solve both the forward and inverse problem. The solution is useful both for understanding the ill-posedness of the problem and as another method of either simulating data or computing a solution. Since one way to completely avoid an inverse crime is to use unrelated methods for reconstruction and data simulation, this is particularly useful here.

The forward problem considered here can be solved by the method of eigenfunction expansions, also known as separation of variables, to obtain an integral equation for the solution of the backward heat equation. This will then be discretized by a Galerkin method to obtain another formulation of $\mathcal{A}f = m$.

Consider the forward problem (2.20) on $(0, \pi)$ repeated below for convenience:

$$\begin{aligned} u_t - u_{xx} &= 0, \quad 0 < x < \pi, \quad t > 0 \\ u(0, t) &= 0, \quad t > 0 \\ u(\pi, t) &= 0, \quad t > 0 \\ u(x, 0) &= f(x), \quad 0 < x < \pi \end{aligned} \tag{3.2}$$

The technique of separation of variables can be found in nearly every introductory text on PDE's, and we will just include a brief outline of the method as applied to this problem, allowing the reader to go to a PDE's text for a more thorough treatment. Suppose $u(x, t)$ can be written as the product of a function depending on x and a function depending on t . That is, $u(x, t) = X(x)\mathcal{T}(t)$. Substituting this into the PDE results in

$X(x)\mathcal{T}'(t) = X''(x)\mathcal{T}(t)$ or

$$\frac{\mathcal{T}'(t)}{\mathcal{T}(t)} = \frac{X''(x)}{X(x)} = -\lambda^2 \quad (3.3)$$

where $-\lambda^2$ is the separation constant. The equation in the spatial variable is a Sturm-Liouville problem with boundary conditions $X(0) = X(\pi) = 0$, which results in eigenvalues n and eigenfunctions $\sin nx$. Thus, $-\lambda^2 = -n^2$ and $\mathcal{T}(t) = A_n e^{-n^2 t}$. By the superposition principle, we then have the general solution

$$u(x, t) = \sum_{n=1}^{\infty} c_n e^{-n^2 t} \sin nx, \quad c_n = \frac{2}{\pi} \int_0^{\pi} f(y) \sin ny dy. \quad (3.4)$$

The solution at time t is then given by

$$\begin{aligned} u(x, t) &= \sum_{n=1}^{\infty} \frac{2}{\pi} \left(\int_0^{\pi} f(y) \sin ny dy \right) e^{-n^2 t} \sin nx \\ &= \int_0^{\pi} k(x, y, t) f(y) dy \end{aligned}$$

where $k(x, y, t) = \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-n^2 t} \sin nx \sin ny$. Knowledge of the solution at a final time T results in a Fredholm integral equation of the first kind for the initial temperature distribution $f(x)$:

$$u(x, T) = \int_0^{\pi} k(x, y, T) f(y) dy. \quad (3.5)$$

It is clear from (3.5) that a unique solution to the backward problem exists, but a straightforward example illustrates that continuous dependence of the solution on the data is violated. Suppose $f(x) = M \sin mx$ where $M > 0$ and m is fixed, but arbitrary. Then $u(x, T) = M e^{-m^2 T} \sin mx$. Now $|u(x, T)|$ can be made arbitrarily small by choosing m sufficiently large, but $\|f(x)\|_{\infty} = M$. Exercise 3.2.1 is to write this in proof format.

To discretize equation (3.5), we will use a Galerkin method and define a partition of $[0, \pi]$ by

$$x_0 = 0, x_1 = \pi/K, \dots, x_j = j\pi/K, \dots, x_K = \pi.$$

Let ϕ_i , $i = 1, \dots, K$, be the orthonormal basis functions

$$\phi_i(x) = \begin{cases} 1, & x_{i-1} \leq x \leq x_i \\ 0, & \text{otherwise.} \end{cases}$$

Denote the final time data $u(x, T)$ by $g(x)$. Expand f and g in terms of these basis functions as follows:

$$f(x) = \sum_{j=1}^K f_j \phi_j(x), \quad g(x) = \sum_{j=1}^K b_j \phi_j(x)$$

where

$$f_j = \int_0^\pi f(x) \phi_j(x) dx, \quad b_j = \int_0^\pi g(x) \phi_j(x) dx.$$

Then

$$\sum_{j=1}^K b_j \phi_j(x) = \int_0^\pi k(x, y, T) \sum_{j=1}^K f_j \phi_j(y) dy.$$

Multiply by $\phi_i(x)$ and integrate from 0 to π :

$$\int_0^\pi \sum_{j=1}^K b_j \phi_j(x) \phi_i(x) dx = \int_0^\pi \int_0^\pi k(x, y, T) \sum_{j=1}^K f_j \phi_j(y) \phi_i(x) dy dx.$$

By the orthogonality of the ϕ_i this becomes

$$b_i \int_{x_{i-1}}^{x_i} 1 dx = \Delta x \quad b_i = \sum_{j=1}^K f_j \int_{x_{i-1}}^{x_i} \int_{x_{j-1}}^{x_j} k(x, y, T) \phi_j(y) \phi_i(x) dy dx.$$

Letting $i = 1, \dots, K$, we will write this as a linear system

$$\begin{bmatrix} & \\ A_{ij} & \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_K \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_K \end{bmatrix}$$

where the elements for the A matrix are given by

$$A_{ij} = \frac{1}{\Delta x} \int_{x_{i-1}}^{x_i} \int_{x_{j-1}}^{x_j} k(x, y, T) \phi_j(y) \phi_i(x) dy dx. \quad (3.6)$$

Approximating the series for $k(x, y, T)$ by a finite sum with N terms results in

$$\begin{aligned} A_{ij} &\approx \frac{2}{\pi} \sum_{n=1}^N e^{-n^2 T} \int_{x_{i-1}}^{x_i} \int_{x_{j-1}}^{x_j} \sin(nx) \sin(ny) dy dx \\ &= \frac{2}{\pi} \sum_{n=1}^N e^{-n^2 T} \int_{x_{i-1}}^{x_i} \sin(nx) dx \int_{x_{j-1}}^{x_j} \sin(ny) dy. \\ &= \frac{2}{\pi} \sum_{n=1}^N \frac{e^{-n^2 T}}{n^2} (\cos(nx_i) - \cos(nx_{i-1}))((\cos(ny_j) - \cos(ny_{j-1})) \end{aligned}$$

Note that no time-stepping is necessary with this method since we have a closed form of the solution. The accuracy of the method on the example of Section 2.2.3 is demonstrated in Table 3.2. One sees that the accuracy improves linearly with the discretization, and it requires quite a fine discretization of $[0, \pi]$ (512 discretization points) to attain an accuracy on the nearly same order as that of the finite difference method of Section 2.2.3.

Time T	Δx	$\ u(x, T)\ _\infty$	Relative sup-norm error
0.1	0.0245	6.67	0.0242
	0.0123	6.67	0.0122
	0.0061	6.67	0.0061
0.2	0.0245	4.49	0.0245
	0.0123	4.49	0.0123
	0.0061	4.49	0.0061
0.3	0.0245	3.01	0.0255
	0.0123	3.01	0.0127
	0.0061	3.01	0.0061
0.4	0.0245	2.01	0.0286
	0.0123	2.01	0.0143
	0.0061	2.01	0.0071

Table 3.1: Accuracy of the Galerkin method forward solver on the backward heat equation (2.32) – (2.35).

Exercise 3.2.1 Give an ϵ, δ proof that the choice $u(x, 0) = M \sin mx$ and $u(x, T) = M e^{-m^2 T} \sin mx$ for $M > 0$ and $m \in \mathbb{N}$ fixed but arbitrary demonstrates that the solution to the backwards heat equation does not depend continuously on the data.

Exercise 3.2.2 Add a column to Table 3.2 displaying the condition number of A for each time T and Δx .

Exercise 3.2.3 Study the numerical forward solution of the heat equation by adding noise to the initial data $u_0(x) = 10 \sin 2x$ in the MATLAB program `InverseHeatCondDataSimulator.m` and computing the solution at time $T =$

0.4. Plot the difference between the computed $u(x, T)$ from a noisy and non-noisy initial condition. How does this differ from what we see in Figures 2.9 and 2.37?

3.3 Ill-posedness in the continuous case

This section relies on some knowledge of operator theory, and the reader may need to first cover some or all of the material in the appendices before covering this section. Recall from Appendix B that $\mathcal{L}(X, Y)$ denotes the space of bounded linear mappings from a normed linear space X to a normed linear space Y .

In terms of the linear operator $\mathcal{A} : U \rightarrow V$ with U and V being subsets of normed spaces X and Y , the conditions of well-posedness imply

- \mathcal{A} is surjective (onto)
- \mathcal{A} is injective (one-to-one)
- \mathcal{A}^{-1} is continuous

These three properties are not independent. The following theorem is a consequence of the closed graph theorem, which implies that the range of a continuous operator from a Banach space to a Banach space is closed.

Theorem 3.3.1 (*Theorem 7.72 [395]*) *Let X and Y be Banach spaces. If $\mathcal{A} \in \mathcal{L}(X, Y)$ is bijective, then $\mathcal{A}^{-1} \in \mathcal{L}(Y, X)$.*

By Theorem B.0.2 of Appendix B, boundedness implies continuity, so the problem $\mathcal{A}f = m$ is well-posed. From the following theorem, we see that an ill-posed problem is guaranteed to result when a compact linear operator acts on an infinite-dimensional Banach space.

Theorem 3.3.2 *Let X and Y be Banach spaces. Suppose $\mathcal{A} : U \subset X \rightarrow Y$ is a compact linear operator, and $\dim U$ is not finite. Then the problem $\mathcal{A}f = m$ is ill-posed.*

Proof: Suppose $\dim U = \infty$ and $\mathcal{A}^{-1} \in \mathcal{L}(Y, U)$, that is, the inverse of \mathcal{A} exists and is continuous. Then $I_U = \mathcal{A}^{-1}\mathcal{A}$ is the composition of a compact and a continuous operator and thus compact. However, the identity map I_U on an infinite-dimensional Banach space is not compact. Thus, $\dim(U)$ is finite. \square

Two straightforward consequences of this theorem are given in the following two corollaries. Their proofs are Exercises 3.3.1 and 3.3.2.

Corollary 3.3.1 *If \mathcal{A} is a compact, linear operator from H to H , where H is an infinite-dimensional Hilbert space, and \mathcal{A}^{-1} exists, then \mathcal{A}^{-1} is unbounded.*

Corollary 3.3.2 *If $\mathcal{A} : H \rightarrow H$ is a compact, linear operator and \mathcal{A}^{-1} exists, then $\dim(H)$ is finite.*

The following theorem provides one way to prove that an operator is compact. Theorems 3.3.3 and 3.3.4 are very important because they explain why the ill-posed behavior of a large class of linear systems, $\mathcal{A}f = m$ cannot be rigorously detected by examining a single approximation matrix A_k to the operator \mathcal{A} , and why one needs to examine the sequence of approximations $\{A_k\}$ for large k . This leads to the test defined by condition (3.18).

Theorem 3.3.3 *Let $\mathcal{A} : H \rightarrow H$ be a linear operator and $K_n \in \mathcal{L}(H, H)$ a sequence of compact operators. If $K_n \rightarrow \mathcal{A}$ in the operator norm, then \mathcal{A} is a compact operator.*

The proof can be found in [395]; it uses a diagonal sequence argument.

If K is a Hilbert-Schmidt kernel of an integral operator, the resulting integral equation, a Fredholm integral equation of the first kind when $\mathcal{A}f = m$, is an ill-posed problem, as we see from the following theorem.

Theorem 3.3.4 *Let*

$$(\mathcal{A}f)(x) = \int_{\Omega} K(x, y) f(y) dy$$

with kernel $K \in L^2(\Omega \times \Omega)$ (i.e., K is a Hilbert-Schmidt kernel from $\Omega \times \Omega \rightarrow \mathbb{R}$). Then $\mathcal{A} \in \mathcal{L}(L^2(\Omega), L^2(\Omega))$ is compact.

Proof: First note that if Φ_i is an orthonormal basis for $L^2(\Omega)$, then $\Phi_i(x) \Phi_j(y)$ is an orthonormal basis for $L^2(\Omega \times \Omega)$. Express K in this basis:

$$K(x, y) = \sum_{i,j=1}^{\infty} k_{i,j} \Phi_i(x) \Phi_j(y)$$

with

$$k_{i,j} = \int_{\Omega} \int_{\Omega} K(x, y) \Phi_i(x) \Phi_j(y) dx dy.$$

The convergence is in L^2 , and we find

$$\|K\|_{L^2}^2 = \sum_{i,j=1}^{\infty} k_{i,j}^2.$$

Define

$$K_n(x, y) = \sum_{i,j=1}^n k_{i,j} \Phi_i(x) \Phi_j(y).$$

Then

$$(\mathcal{A}_n f)(x) = \int_{\Omega} K_n(x, y) f(y) dy = \sum_{i,j=1}^n k_{i,j} \int_{\Omega} \Phi_i(x) \Phi_j(y) f(y) dy.$$

Thus, \mathcal{A}_n maps from $L^2(\Omega)$ to a finite-dimensional subspace of $L^2(\Omega)$, which we will denote by $\tilde{L}^2(\Omega)$. By the converse of Corollary 3.3.1, due to the fact that the range(\mathcal{A}_n) is finite-dimensional, \mathcal{A}_n is compact. Now we see

$$\begin{aligned} \|(\mathcal{A} - \mathcal{A}_n) f\|_{\tilde{L}^2(\Omega)}^2 &= \left\| \int_{\Omega} (K(x, y) - K_n(x, y)) f(y) dy \right\|_{\tilde{L}^2(\Omega)}^2 \\ &= \int_{\Omega} \left| \int_{\Omega} (K(x, y) - K_n(x, y)) f(y) dy \right|^2 dx \\ &\leq \int_{\Omega} \left(\int_{\Omega} |K(x, y) - K_n(x, y)| |f(y)| dy \right)^2 dx \\ &\quad \text{by Cauchy-Schwarz} \\ &\leq \int_{\Omega} \left(\int_{\Omega} |K(x, y) - K_n(x, y)|^2 dx \int_{\Omega} |f(y)|^2 dy \right) dx \\ &= \left(\int_{\Omega} \int_{\Omega} |K(x, y) - K_n(x, y)|^2 dx dy \right) \int_{\Omega} |f(y)|^2 dy. \end{aligned}$$

Therefore,

$$\begin{aligned} \|\mathcal{A} - \mathcal{A}_n\|^2 &\leq \int_{\Omega} \int_{\Omega} |K(x, y) - K_n(x, y)|^2 dx dy \\ &= \sum_{i,j=n+1}^{\infty} |k_{i,j}|^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Thus, \mathcal{A}_n converges to \mathcal{A} in the operator norm, and \mathcal{A}_n is compact. We conclude that \mathcal{A} is compact. \square

As we saw in Section 3.2, the 1-D heat equation (3.2) on $[0, \pi]$ is solved by

$$u(x, t) = \int_0^{\pi} k(x, y; t) f(y) dy$$

where

$$k(x, y; t) = \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-n^2 t} \sin(n x) \sin(n y). \quad (3.7)$$

Using an appropriate approximation k_n to the kernel k , one can show that the operator is compact. (See Exercise 3.3.4.) Thus, the inverse problem of determining the initial temperature $u(x, 0)$ given $u(x, T)$ is an ill-posed problem. It is also a Fredholm integral equation of the first kind.

Definition 3.3.1 Let $G \subset \mathbb{R}^n$ be a bounded, open set and $x, y \in G$. A kernel K is **weakly singular** if and only if K is smooth for $x \neq y$ and there exist constants $b > 0$ and $\nu < n$ such that

$$|K(x, y)| \leq b |x - y|^{-\nu}.$$

Under certain conditions (see Appendix B or [448]) a weakly singular kernel K results in a compact operator \mathcal{A} defined by

$$(\mathcal{A}f)(x) = \int_{\Omega} K(x, y) f(y) dy.$$

The condition $\nu < n$ for weakly singular kernels K guarantees that K is absolutely integrable (which is why the singularity is called “weak”). To see this, suppose $|K(x, y)| \leq b |x - y|^{-\nu}$ and τ_n is the area of the unit sphere in \mathbb{R}^n . Let G be an open bounded subset and $y \in G$ with $\mathbb{B}_y(R) \subset G$. Then

$$\begin{aligned} \int_G |K(x, y)| dx &\leq C + \int_{\mathbb{B}_y(R)} |K(x, y)| dx \leq C + \int_{\mathbb{B}_y(R)} \frac{b}{|x - y|^{\nu}} dx \\ &\leq C + \int_{|z| \leq R} \frac{b}{|z|^{\nu}} dz = C + \int_0^R \frac{b \sigma_n r^{n-1}}{r^{\nu}} dr \\ &= C + \int_0^R b \sigma_n r^{n-1-\nu} dr = C + \frac{b \sigma_n R^{n-\nu}}{n - \nu}, \end{aligned}$$

which is finite.

Returning to the example of X-ray tomography, recall from (2.39) the data can be modeled as

$$\log \frac{I(x)}{I_0} = - \int_{-y(x)}^{+y(x)} \beta(x, y) dy,$$

where $\beta(x, y)$ is the attenuation coefficient and x is in a circle of radius R . If β is radially symmetric, that is, $\beta(x, y) = \beta(\sqrt{x^2 + y^2}) = \beta(r)$, a change

of variables gives

$$\begin{aligned}
p(x) = \log \frac{I(x)}{I_0} &= -2 \int_0^{\sqrt{R^2-x^2}} \beta(\sqrt{x^2+y^2}) dy \\
&\quad r = \sqrt{x^2+y^2}, \quad dr = \frac{y}{r} dy \\
&= -2 \int_x^R \frac{r}{\sqrt{r^2-x^2}} \beta(r) dr \\
&\quad z = R^2 - r^2, \quad dz = -2r dr \\
&= \int_{R^2-x^2}^0 \frac{1}{\sqrt{R^2-x^2-z}} \beta(\sqrt{R^2-z}) dz.
\end{aligned}$$

With the notations $\tau = R^2 - x^2$ and $\Phi(z) = \beta(\sqrt{R^2-z})$, we find

$$P(\tau) = p(\pm\sqrt{R^2-\tau}) = - \int_0^\tau \frac{\Phi(z)}{\sqrt{\tau-z}} dz.$$

Since $k(\tau, z) = |\tau - z|^{-1/2}$, this kernel is weakly singular.

3.3.1 The Singular Value Expansion (SVE)

A very thorough treatment of the Singular Value Expansion of a kernel can be found in, for example, the text by Hansen [193]. We include some of that material here to provide a better understanding of the underlying continuous problem and also to motivate the SVD. In this section we will consider the Fredholm integral equation of the first kind, in which we wish to find $f(t)$ such that

$$\int_0^1 K(s, t)f(t)dt = g(s), \quad 0 \leq s \leq 1, \quad \|K\|_{L^2([0,1] \times [0,1])}^2 \leq C. \quad (3.8)$$

The SVE Theorem states that any kernel K with $\|K\|_{L_2} < \infty$ can be written in the form (where the upper limit of the sum is finite in the case of a degenerate kernel)

$$K(s, t) = \sum_{i=1}^{\infty} \mu_i u_i(s)v_i(t),$$

where u_i and v_i are the **singular functions** and the μ_i are the **singular values** of K . The singular functions are orthonormal and the singular values satisfy

$$\mu_1 \geq \mu_2 \geq \dots \geq 0 \quad \text{and} \quad \sum_{i=1}^{\infty} \mu_i^2 = \|K\|_{L_2}.$$

Also

$$\int_0^1 K(s, t) v_i(t) = \mu_i u_i(s).$$

Multiplying (3.8) by $\mu_i u_i(s)$, then integrating with respect to s and using the fact that the u_i are orthonormal implies

$$\begin{aligned} \int_0^1 \mu_i u_i(s) g(s) ds &= \int_0^1 \mu_i u_i(s) \int_0^1 \left(\sum_{j=1}^{\infty} \mu_j u_j(s) v_j(t) f(t) \right) dt ds \\ &= \sum_{j=1}^{\infty} \int_0^1 v_j(t) f(t) \mu_i \mu_j \left(\int_0^1 u_i(s) u_j(s) ds \right) dt \\ &= \int_0^1 v_j(t) f(t) \mu_i^2 dt. \end{aligned}$$

Thus, $\langle v_i, f \rangle v_i(t) = \frac{1}{\mu_i} \langle u_i, g \rangle v_i(t)$, and we have the following expression for the solution of (3.8)

$$f(t) = \sum_{i=1}^{\infty} \langle v_i, f \rangle v_i(t) = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \langle u_i, g \rangle v_i(t). \quad (3.9)$$

The behavior of the singular values and singular functions depends on the properties of the kernel K . As observed in [193], the smoother the kernel K , the faster the singular values μ_i decay. The proof is due to Smithies [420]. The smaller the μ_i , the more oscillatory the functions u_i and v_i will be. There is no known proof of this result in general, but it has been frequently observed. Also, noteworthy is that the factor $\frac{1}{\mu_i}$ amplifies highly oscillatory contributions in g .

A useful characterization of ill-posedness in terms of the singular values was introduced in [212].

Definition 3.3.2 *If there exists a real number $\alpha > 0$ such that the singular values satisfy $\mu_n = O(n^{-\alpha})$ then α is the **degree of ill-posedness**.*

1. *If $0 < \alpha \leq 1$, the problem is **mildly ill-posed**.*
2. *If $\alpha > 1$, the problem is **moderately ill-posed**.*
3. *If $\mu_n = \mathcal{O}(e^{-\alpha n})$, the problem is **severely ill-posed**.*

For the problem (3.8) to have a solution f given by (3.9), g must satisfy

$$\sum_{i=1}^{\infty} \left(\frac{\langle u_i, g \rangle}{\mu_i} \right)^2 < \infty.$$

This is known as the **Picard condition**. This is a stronger condition than $g \in L^2(0, 1)$ since $g \in L^2$ requires $|\langle u_n, g \rangle| \leq \frac{1}{\sqrt{n}}$ while Picard's condition requires $|\langle u_n, g \rangle| \leq \frac{\mu_n}{\sqrt{n}}$. The Picard condition is equivalent to $g \in \text{Range}(K)$ since

$$\sum_{i=1}^{\infty} \mu_i \langle v_i, f \rangle u_i(t) = \sum_{i=1}^{\infty} \langle u_i, g \rangle u_i(t)$$

implies the right hand side is a projection of g onto $\text{span}\{u_i\}_{i \in \mathbb{N}}$. Let g_k be the approximation of g obtained by truncating the SVE after k terms:

$$g_k(t) = \sum_{i=1}^k \langle u_i, g \rangle u_i(t)$$

Clearly, the Picard condition is satisfied and the solution is given by

$$f_k(t) = \sum_{i=1}^k \frac{\langle u_i, g \rangle}{s_i} v_i(t)$$

If $g \notin \text{range}(K)$, then $g_k \rightarrow g$ as $k \rightarrow \infty$, but $\|f_k\|_{L^2}^2 = \langle f_k, f_k \rangle \rightarrow \infty$. This shows the lack of stability.

In practice g will contain measurement errors, and can be expressed as $g = g^{\text{exact}} + \eta$, where $g^{\text{exact}} \in \text{range}(K)$ but $\eta \notin \text{range}(K)$.

3.3.2 Discretization

We introduce two discretizations of (3.8).

- **Quadrature method:**

A discretization of (3.8) using a quadrature method based on the points $0 \leq t_1 < t_2 < t_3 \dots < t_n \leq 1$ in which

$$\int_0^1 \Phi(t) dt \approx \sum_{j=1}^n w_j \Phi(t_j)$$

can be applied at M discrete points $0 \leq s_1 < s_2 < s_3 \dots < s_M \leq 1$ leading to M linear equations

$$\sum_{j=1}^n w_j K(s_i, t_j) f(t_j) = g(s_i).$$

Defining a matrix A with entries $a_{ij} \equiv w_j K(s_i, t_j)$, a vector \vec{b} with entries $b_i \equiv g(s_i)$, and a vector \vec{f} with entries $f_j \equiv f(t_j)$, the system of equations becomes $A\mathbf{f} = \mathbf{g}$. The number of unknowns n does not have to coincide with the number of equations M .

• **Galerkin method:**

Let ϕ_i and ψ_j be orthonormal basis functions in $L^2(0, 1)$ and define vectors \mathbf{b}, ξ and matrix A by

$$\begin{aligned} b_i &= \int_0^1 g(s) \phi_i(s) ds, \\ \xi_j &= \int_0^1 f(t) \psi_j(t) dt, \\ a_{ij} &= \int_0^1 \int_0^1 K(s, t) \phi_i(s) \psi_j(t) ds dt, \end{aligned}$$

and write

$$K(s, t) = \sum_i \sum_j a_{ij} \phi_i(s) \psi_j(t) \quad \text{and} \quad g(s) = \sum_i b_i \phi_i(s).$$

Then the solution ξ of $A\xi = \mathbf{b}$ generates an approximate solution $f(t) = \sum_j \xi_j \psi_j(t)$. If $\psi_i = \phi_i$ and the kernel K is symmetric ($K(t, s) = K(s, t)$), then the resulting matrix A will be symmetric, and this is known as the *Rayleigh-Ritz method*.

Exercise 3.3.1 Prove Corollary 3.3.1.

Exercise 3.3.2 Prove Corollary 3.3.2.

Exercise 3.3.3 If $K \in \mathcal{L}(H, H)$ is compact and $\{e_n\}$ is an orthonormal basis for H , then $K e_n \rightarrow 0$ as $n \rightarrow \infty$.

Exercise 3.3.4 Using an appropriate approximation k_n to the kernel k (3.7), show that the operator \mathcal{A} defined by

$$(\mathcal{A}f)(x, t) \equiv \int_0^\pi k(x, y; t) f(y) dy$$

is compact.

Exercise 3.3.5 Let $\{e_n\}$ be an orthonormal sequence (not necessarily a basis) in a Hilbert space H . Let $\{\lambda_n\}_{n=1}^{\infty} \subset \mathbb{C} \setminus \{0\}$ with $\sum_{n=1}^{\infty} |\lambda_n|^2 < \infty$. Define $K : H \rightarrow H$ by

$$Kx = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n.$$

Prove that K is compact and linear.

Exercise 3.3.6 Let $H = L^2(0, \pi)$ and let

$$k(x, y) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sin((n+1)x) \sin(ny), \quad x, y \in (0, \pi).$$

show that the operator \mathcal{A} defined by

$$(\mathcal{A}f)(x, t) \equiv \int_0^{\pi} k(x, y; t) f(y) dy$$

is compact by showing that k is a Hilbert-Schmidt kernel.

3.4 Regularized inversion

Having established that many inverse problems are ill-posed, we now introduce an indispensable technique known as *regularization* or *regularized inversion* to deal with the ill-posedness.

The basic inverse problem related to the indirect measurement (3.1) is this:

- (IP₁) **Inverse Problem:** Let $m = \mathcal{A}f + \varepsilon$ as in (3.1). Given m and $\delta > 0$ with $\|m - \mathcal{A}f\|_Y \leq \delta$, recover f approximately.

In ill-posed inverse problems there does not exist any continuous function from Y to X that would map $\mathcal{A}f \in Y$ to $f \in X$. This can be viewed as extreme sensitivity to perturbations in $\mathcal{A}f$, which is inevitable because of measurement noise. Consequently, it is not straightforward to design a computational method that would map $m = \mathcal{A}f + \varepsilon$ to some point in X near f .

The naïve way of approaching the inverse problem IP₁ would be to approximate f by $\mathcal{A}^{-1}m$. Because m and $\mathcal{A}f$ are close to each other, the point $f = \mathcal{A}^{-1}\mathcal{A}f$ must be close to $\mathcal{A}^{-1}m$. This approach is fine for well-posed problems but does not work for ill-posed inverse problems since \mathcal{A} may fail to be either injective or surjective, and even if \mathcal{A} is invertible, its inverse may not be continuous. So what can be done?

Sometimes it helps to consider a restricted problem, such as

(IP₂) **Restricted Inverse Problem:** Let $m = \mathcal{A}f + \varepsilon$ as in (3.1).

Given m and $\delta > 0$ with $\|m - \mathcal{A}f\|_Y \leq \delta$,
extract any information about f .

For example, one might look for the locations of inclusions in known background material. In any case, the most important property of an inversion method is robustness against noise.

Let us define the notions of *regularization strategy* and *admissible choice of regularization parameter*. We need to assume that $\text{Ker}(\mathcal{A}) = \{0\}$; however, this is not a serious lack of generality since we can always consider the restriction of \mathcal{A} to $(\text{Ker}(\mathcal{A}))^\perp$ by working in the linear space of equivalence classes $[f + \text{Ker}(\mathcal{A})]$.

Definition 3.4.1 Let X and Y be Hilbert spaces. Let $\mathcal{A} : X \rightarrow Y$ be an injective bounded linear operator. Consider the measurement $m = \mathcal{A}f + \varepsilon$. A family of linear maps $\mathcal{R}_\alpha : Y \rightarrow X$ parameterized by $0 < \alpha < \infty$ is called a regularization strategy if

$$\lim_{\alpha \rightarrow 0} \mathcal{R}_\alpha \mathcal{A}f = f \quad (3.10)$$

for every $f \in X$.

Further, assume we are given a noise level $\delta > 0$ so that $\|m - \mathcal{A}f\|_Y \leq \delta$. A choice of regularization parameter $\alpha = \alpha(\delta)$ as function of δ is called admissible if

$$\begin{aligned} \alpha(\delta) &\rightarrow 0 \text{ as } \delta \rightarrow 0, \text{ and} \\ \sup_m \left\{ \|\mathcal{R}_{\alpha(\delta)} m - f\| : \|\mathcal{A}f - m\| \leq \delta \right\} &\rightarrow 0 \text{ as } \delta \rightarrow 0 \text{ for every } f \in X. \end{aligned} \quad (3.11)$$

Figure 3.2 shows a schematic illustration of regularized inversion. See [131, 268] for more details about regularization.

This book studies computational inversion methods that apply to practical inverse problems and allow precise mathematical analysis. Developing such methods typically involves considering the following series of questions:

- Q₁: **Uniqueness.** Is the forward map \mathcal{A} injective on $\mathcal{D}(\mathcal{A})$? In other words, are there two different objects $f \neq \tilde{f}$ producing exactly the same infinite-precision measurement data: $\mathcal{A}f = \mathcal{A}\tilde{f}$?
- Q₂: **Reconstruction from ideal data.** Assume that \mathcal{A} is injective on $\mathcal{D}(\mathcal{A})$. Given infinite-precision data $\mathcal{A}f$, how to calculate f ?

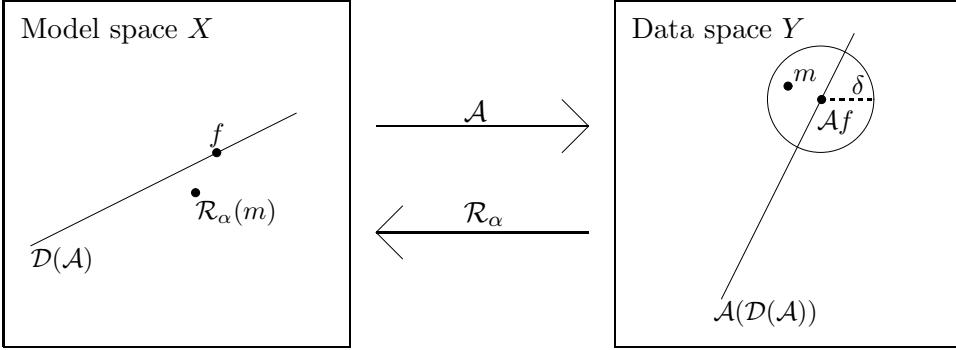


Figure 3.2: Schematic illustration of regularization for linear inverse problems. The linear forward map \mathcal{A} does not necessarily have a continuous inverse. The regularized approximate inverses $\mathcal{R}_\alpha : Y \rightarrow X$ are continuous for any choice of regularization parameter $0 < \alpha < \infty$.

Q₃: **Conditional stability.** Assume that \mathcal{A} is injective on $\mathcal{D}(\mathcal{A})$. Is it possible to derive a formula of the type $\|f - \tilde{f}\|_X \leq g(\|\mathcal{A}f - \mathcal{A}\tilde{f}\|_Y)$, where $g : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function satisfying $g(0) = 0$?

Q₄: **Characterization of the range.** How to decide whether a given $m \in Y$ belongs to $\mathcal{A}(\mathcal{D}(\mathcal{A}))$, in other words, whether $m = \mathcal{A}f$ for some $f \in X$?

Q₅: **Reconstruction from practical data.** Given the noisy data $m = \mathcal{A}f + \varepsilon$, how to approximate f in a noise-robust fashion?

3.5 The singular value decomposition for matrices

In practice the continuum measurement model (1.1) of the form $\mathbf{m} = \mathcal{A}\mathbf{f} + \varepsilon$ needs to be approximated by a discrete model of the form $\mathbf{m} = \mathbf{Af} + \varepsilon$, where A is a matrix, $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{m} \in \mathbb{R}^k$. Let us now discuss a tool that allows explicit analysis of Hadamard's conditions in this finite-dimensional setting, namely *singular value decomposition* of A .

We know from matrix algebra that any matrix $A \in \mathbb{R}^{k \times n}$ can be written in the form

$$A = UDV^T, \quad (3.12)$$

where $U \in \mathbb{R}^{k \times k}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, that is,

$$U^T U = UU^T = I, \quad V^T V = VV^T = I,$$

and $D \in \mathbb{R}^{k \times n}$ is a diagonal matrix. The right side of (3.12) is called the singular value decomposition (SVD) of matrix A , and the diagonal elements d_j are the *singular values* of A . The properties of d_j , and the columns u_i of U , and the columns V_i of V correspond to those of the SVE.

In the case $k = n$ the matrix D is square-shaped: $D = \text{diag}(d_1, \dots, d_k)$. If $k > n$ then

$$D = \begin{bmatrix} \text{diag}(d_1, \dots, d_n) \\ \mathbf{0}_{(k-n) \times n} \end{bmatrix} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & d_n \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \quad (3.13)$$

and in the case $k < n$ the matrix D takes the form

$$\begin{aligned} D &= [\text{diag}(d_1, \dots, d_k), \mathbf{0}_{k \times (n-k)}] \\ &= \begin{bmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & & & \vdots & \vdots & \vdots \\ \vdots & & \ddots & & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & d_k & 0 & \cdots & 0 \end{bmatrix}. \end{aligned} \quad (3.14)$$

The diagonal elements d_j are nonnegative and in decreasing order:

$$d_1 \geq d_2 \geq \dots \geq d_{\min(k,n)} \geq 0. \quad (3.15)$$

Note that some or all of the d_j can be equal to zero.

Recall the definitions of the following linear subspaces related to the matrix A :

$$\begin{aligned} \text{Ker}(A) &= \{\mathbf{f} \in \mathbb{R}^n : A\mathbf{f} = 0\}, \\ \text{Range}(A) &= \{\mathbf{m} \in \mathbb{R}^k : \text{there exists } \mathbf{f} \in \mathbb{R}^n \text{ such that } A\mathbf{f} = \mathbf{m}\}, \\ \text{Coker}(A) &= (\text{Range}(A))^\perp \subset \mathbb{R}^k. \end{aligned}$$

See Figure 3.3 for a diagram illustrating these concepts.

Failure of Hadamard's existence and uniqueness conditions can now be detected from the matrix D . If $k > n$ then $\dim(\text{Range}(A)) < k$ and we can choose a nonzero $\mathbf{m}_0 \in \text{Coker}(A)$ as shown in Figure 3.3. Even in the case

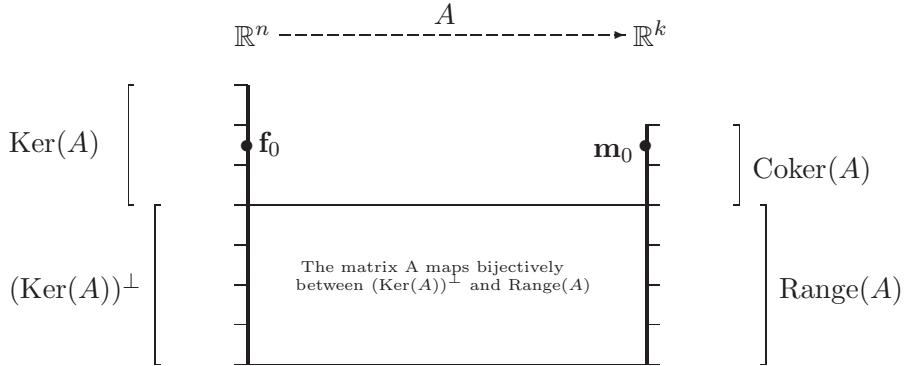


Figure 3.3: This diagram illustrates various linear subspaces related to a matrix mapping \mathbb{R}^n to \mathbb{R}^k . The two thick vertical lines represent the linear spaces \mathbb{R}^n and \mathbb{R}^k ; in this schematic picture we have $n = 7$ and $k = 6$. Furthermore, $\dim(\text{Ker}(A)) = 3$ and $\dim(\text{Range}(A)) = 4$ and $\dim(\text{Coker}(A)) = 2$. Note that the 4-dimensional orthogonal complement of $\text{Ker}(A)$ in \mathbb{R}^n is mapped in a bijective manner to $\text{Range}(A)$. The points $\mathbf{f}_0 \in \text{Ker}(A)$ and $\mathbf{m}_0 \in \text{Coker}(A)$ are used in the text.

$\varepsilon = 0$ we have problems since there does not exist any $\mathbf{f} \in \mathbb{R}^n$ satisfying $A\mathbf{f} = \mathbf{m}_0$, and consequently the existence condition H₁ fails since the output $A^{-1}\mathbf{m}_0$ is not defined for the input \mathbf{m}_0 . In case of nonzero random noise the situation is even worse since even though $A\mathbf{f} \in \text{Range}(A)$, it might happen that $A\mathbf{f} + \varepsilon \notin \text{Range}(A)$. If $k < n$ then $\dim(\text{Ker}(A)) > 0$ and we can choose a nonzero $\mathbf{f}_0 \in \text{Ker}(A)$ as shown in Figure 3.3. Then even in the case $\varepsilon = 0$ we have a problem of defining $A^{-1}\mathbf{m}$ uniquely since both $A^{-1}\mathbf{m}$ and $A^{-1}\mathbf{m} + \mathbf{f}_0$ satisfy $A(A^{-1}\mathbf{m}) = \mathbf{m} = A(A^{-1}\mathbf{m} + \mathbf{f}_0)$. Thus the uniqueness condition H₂ fails unless we specify an explicit way of dealing with the null-space of A . Note that if $d_{\min(k,n)} = 0$ then both conditions H₁ and H₂ fail.

The above problems with existence and uniqueness are quite clear since they are related to integer-valued dimensions. In contrast, ill-posedness related to the continuity condition H₃ is more tricky in our finite-dimensional context. Consider the case $n = k$ so A is a square matrix, and assume that A is invertible. In that case we can write

$$A^{-1}\mathbf{m} = A^{-1}(A\mathbf{f} + \varepsilon) = \mathbf{f} + A^{-1}\varepsilon, \quad (3.16)$$

where the error $A^{-1}\varepsilon$ can be bounded by

$$\|A^{-1}\varepsilon\| \leq \|A^{-1}\| \|\varepsilon\|.$$

Now if $\|\varepsilon\|$ is small and $\|A^{-1}\|$ has reasonable size then the error $A^{-1}\varepsilon$ is small. However, if $\|A^{-1}\|$ is large, then the error $A^{-1}\varepsilon$ can be huge even when ε is small. This is the kind of amplification of noise we see in Figures 2.5, 2.9 and 2.19.

Note that if $\varepsilon = 0$ in (3.16) then we do have $A^{-1}\mathbf{m} = \mathbf{f}$ even if $\|A^{-1}\|$ is large. However, in practice the measurement data always has some noise, and even computer simulated data is corrupted with round-off errors. Those inevitable perturbations prevent using $A^{-1}\mathbf{m}$ as a reconstruction method for an ill-posed problem.

To define ill-posedness related to the continuity condition H_3 rigorously, we must consider the relative sizes of the singular values. Consider the case $n = k$ and $d_n > 0$, when we do not have the above problems with existence or uniqueness. It seems that nothing is wrong since we can invert the matrix A as

$$A^{-1} = VD^{-1}U^T, \quad D^{-1} = \text{diag}\left(\frac{1}{d_1}, \dots, \frac{1}{d_k}\right),$$

and define $\mathcal{R}(\mathbf{m}) = A^{-1}\mathbf{m}$ for any $\mathbf{m} \in \mathbb{R}^k$. The problem comes from the *condition number*

$$\text{Cond}(A) := \frac{d_1}{d_k} \tag{3.17}$$

being large. Namely, if d_1 is several orders of magnitude greater than d_k , then numerical inversion of A becomes difficult since the diagonal inverse matrix D^{-1} contains floating point numbers of very different sizes. This in turn leads to uncontrollable amplification of truncation errors.

Strictly mathematically speaking, though, A is an invertible matrix even in the case of large condition number. For a rigorous definition, we must return to the continuum problem approximated by the matrix model. Suppose that we model the continuum measurement by a sequence of matrices A_k having size $k \times k$ for $k = k_0, k_0 + 1, k_0 + 2, \dots$ so that the approximation to the forward problem becomes better as k grows. Then we say that condition H_3 fails if

$$\lim_{k \rightarrow \infty} \text{Cond}(A_k) = \infty. \tag{3.18}$$

Thus, the ill-posedness cannot be rigorously detected from one approximation matrix A_k but only from the sequence $\{A_k\}_{k=k_0}^\infty$. Theorem 3.3.3 tells us further that the ill-posedness of the problem is not evident from a single approximation matrix A_k to the operator \mathcal{A} , but only from a sequence of approximations.

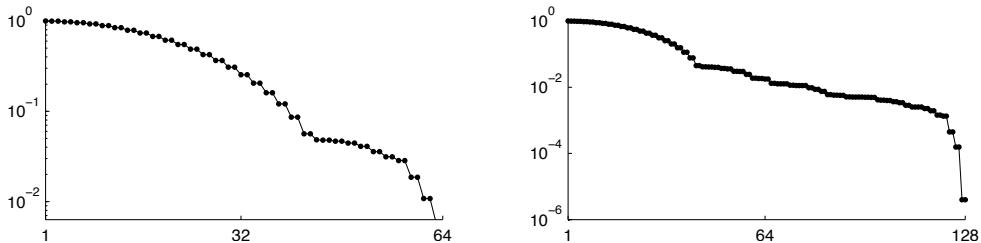


Figure 3.4: Left: plot of the singular values of the 64×64 convolution matrix in logarithmic scale. Right: plot of the singular values of the 128×128 convolution matrix in logarithmic scale.

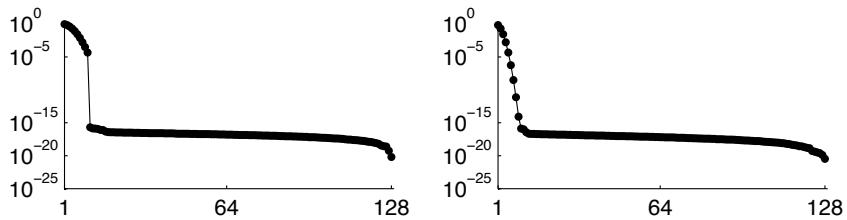


Figure 3.5: Singular values of the matrix A defined by (3.6) related to the backward heat problem. Left: final time $T = 0.1$. Right: final time $T = 0.4$.

3.6 SVD for the guiding examples

We start with the one-dimensional convolution example introduced in Section 2.1. We compute the singular values of measurement matrices for the two resolutions $k = n = 64$ and $k = n = 128$. See Figure 3.4 for a logarithmic plot of the singular values. The singular values decrease very fast towards zero but nevertheless stay positive; this is a sign of ill-posedness.

Next we consider the heat propagation model discussed in Section 2.2. Singular values of the matrix A defined by (3.6) for $T = 0.1$ and $T = 0.4$ are shown in Figure 3.5. The distribution of the singular values does not change significantly with Δx , but does have some dependence on the final time T . It is an exercise to compute the condition number of A .

Finally, Figure 3.6 shows singular values of a measurement matrix related to the X-ray tomography problem. The matrix is the one constructed in Section 2.3.5 for the resolution 50×50 and with 50 projection directions.

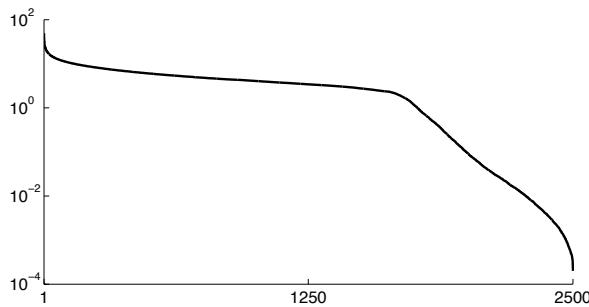


Figure 3.6: Singular values of the X-ray tomography measurement matrix corresponding to the 50×50 resolution and 50 uniformly distributed projection directions.

Exercise 3.6.1 Use the singular values of the SVD of the guiding examples to classify the problems as mildly ill-posed, moderately ill-posed, or severely ill-posed.

Exercise 3.6.2 Calculate the condition number of the matrix A defined by (3.6) for an $n \times n$ matrix with $n = 8, 16, 32, 64$ and 128 . Make a table and compare your results when $T = 0.1$ and $T = 0.4$.

Chapter 4

Truncated singular value decomposition

In this chapter we introduce the concept of least-squares inversion, arguably the most-used method for solving inverse problems and the method we recommend in general for linear problems. For ill-posed problems the minimization of the discrete least squares cost functional will ultimately result in an ill-posed linear system that must be regularized. Truncated SVD is the first regularized inversion method discussed in this book. It also provides a way to get a quick estimate on the degree of ill-posedness of the inverse problem. It has the virtues of being straightforward to implement and is popular starting point for regularized inversion.

4.1 Minimum norm solution

Let us define the minimum norm solution of the matrix equation $A\mathbf{f} = \mathbf{m}$, where $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{m} \in \mathbb{R}^k$ and A has size $k \times n$.

Definition 4.1.1 A vector $\mathcal{L}(\mathbf{m}) \in \mathbb{R}^n$ is called a least-squares solution of the equation $A\mathbf{f} = \mathbf{m}$ if

$$\|A\mathcal{L}(\mathbf{m}) - \mathbf{m}\| = \min_{\mathbf{z} \in \mathbb{R}^n} \|A\mathbf{z} - \mathbf{m}\|. \quad (4.1)$$

Furthermore, $\mathcal{L}(\mathbf{m})$ is called the minimum norm solution if

$$\|\mathcal{L}(\mathbf{m})\| = \inf\{\|\mathbf{z}\| : \mathbf{z} \text{ is a least-squares solution of } A\mathbf{f} = \mathbf{m}\}. \quad (4.2)$$

We need a method for computing minimum norm solutions. For this, write A in the form of its SVD $A = UDV^T$ as explained in Section 3.5. Recall that

the singular values are ordered from largest to smallest as shown in (3.15), and let r be the largest index for which the corresponding singular value is nonzero:

$$r = \max\{j \mid 1 \leq j \leq \min(k, n), d_j > 0\}. \quad (4.3)$$

The definition of index r is essential in the following analysis, so we will be extra-specific:

$$d_1 > 0, \quad d_2 > 0, \quad \dots \quad d_r > 0, \quad d_{r+1} = 0, \quad \dots \quad d_{\min(k,n)} = 0.$$

Of course, it is also possible that all singular values are zero, in which case r is not defined and A is the zero matrix, or none of the singular values may be zero.

The next result gives a method to determine the minimum norm solution.

Theorem 4.1.1 *Let A be a $k \times n$ matrix and denote by $A = UDV^T$ the singular value decomposition of A . The minimum norm solution of the equation $A\mathbf{f} = \mathbf{m}$ is given by $A^+\mathbf{m}$ where*

$$A^+\mathbf{m} = VD^+U^T\mathbf{m},$$

and where

$$D^+ = \begin{bmatrix} 1/d_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1/d_2 & & & & \vdots \\ \vdots & & \ddots & & & \\ & & & 1/d_r & & \\ & & & & 0 & \\ \vdots & & & & & \ddots & \vdots \\ 0 & \cdots & & & & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times k}.$$

Proof: Write the singular matrix V in the form $V = [V_1 \ V_2 \ \cdots \ V_n]$ and note that the column vectors V_1, \dots, V_n form an orthogonal basis for \mathbb{R}^n . We write $\mathbf{f} \in \mathbb{R}^n$ as a linear combination $\mathbf{f} = \sum_{j=1}^n a_j V_j = V\mathbf{a}$, and our goal is to find such coefficients a_1, \dots, a_n that \mathbf{f} becomes a minimum norm solution.

Set $\mathbf{m}' = U^T\mathbf{m} \in \mathbb{R}^k$ and compute

$$\begin{aligned} \|A\mathbf{f} - \mathbf{m}\|^2 &= \|UDV^T V\mathbf{a} - U\mathbf{m}'\|^2 \\ &= \|D\mathbf{a} - \mathbf{m}'\|^2 \\ &= \sum_{j=1}^r (d_j a_j - \mathbf{m}'_j)^2 + \sum_{j=r+1}^k (\mathbf{m}'_j)^2, \end{aligned} \quad (4.4)$$

where we used the orthogonality of U (namely, $\|U\mathbf{z}\| = \|\mathbf{z}\|$ for any vector $\mathbf{z} \in \mathbb{R}^k$). Now since d_j and \mathbf{m}'_j are given and fixed, the expression (4.4) attains its minimum when $a_j = \mathbf{m}'_j/d_j$ for $j = 1, \dots, r$. So any \mathbf{f} of the form

$$\mathbf{f} = V \begin{bmatrix} d_1^{-1}\mathbf{m}'_1 \\ \vdots \\ d_r^{-1}\mathbf{m}'_r \\ a_{r+1} \\ \vdots \\ a_n \end{bmatrix}$$

is a least-squares solution. The smallest norm $\|\mathbf{f}\|$ is clearly given by the choice $a_j = 0$ for $r < j \leq n$, so the minimum norm solution is uniquely determined by the formula $\mathbf{a} = D^+\mathbf{m}'$. \square

Definition 4.1.2 *The matrix A^+ is called the pseudoinverse, or the Moore-Penrose inverse of A .*

How does the pseudoinverse take care of Hadamard's existence and uniqueness conditions H₁ and H₂? First of all, in the case of nonexistence, $\text{Coker}(A)$ is nontrivial, and any vector $\mathbf{m} \in \mathbb{R}^k$ can be written as the sum $\mathbf{m} = \mathbf{m}_A + (\mathbf{m}_A)^\perp$, where $\mathbf{m}_A \in \text{range}(A)$ and $(\mathbf{m}_A)^\perp \in \text{Coker}(A)$ and $\mathbf{m}_A \cdot (\mathbf{m}_A)^\perp = 0$. Then A^+ simply maps $(\mathbf{m}_A)^\perp$ to zero. Second, in the case of non-uniqueness, $\text{Ker}(A)$ is nontrivial, and we need to choose the reconstructed vector from a whole linear subspace of candidates. Using A^+ chooses the candidate with smallest norm.

4.2 Truncated SVD

After the analysis in Section 4.1 it remains to discuss Hadamard's continuity condition H₃. Recall from Section 3.5 that we may run into problems if d_r is much smaller than d_1 . In that case even the use of the pseudoinverse results in numerical instability because the diagonal element d_r^{-1} appearing in D^+ is much larger than d_1^{-1} . We can overcome this by using the truncated SVD.

Definition 4.2.1 *For any $\alpha > 0$ define the truncated SVD (TSVD) by*

$A_\alpha^+ = VD_\alpha^+U^T$ where

$$D_\alpha^+ = \begin{bmatrix} 1/d_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1/d_2 & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ & & & 1/d_{r_\alpha} & & 0 \\ & & & & & 0 \\ \vdots & & & & & \ddots & \vdots \\ 0 & \cdots & & \cdots & & 0 \end{bmatrix} \in \mathbb{R}^{n \times k}$$

and

$$r_\alpha = \min \left\{ r, \max \{j \mid 1 \leq j \leq \min(k, n), d_j > \alpha\} \right\}. \quad (4.5)$$

We can then define a reconstruction function \mathcal{L}_α by the formula

$$\mathcal{L}_\alpha(m) = VD_\alpha^+U^T m. \quad (4.6)$$

Then all of Hadamard's conditions hold: $\mathcal{L}_\alpha : \mathbb{R}^k \rightarrow \mathbb{R}^n$ is a well-defined, single-valued linear mapping with norm

$$\|\mathcal{L}_\alpha\| = \|VD_\alpha^+U^T\| \leq \|V\| \|D_\alpha^+\| \|U^T\| = \|D_\alpha^+\| = d_{r_\alpha}^{-1},$$

implying continuity. To be specific, while of course $\|\mathcal{L}_\alpha\| = d_{r_\alpha}^{-1} < \infty$ implies the linear mapping is continuous in the mathematical sense, equation (3.16) now takes the form

$$\mathcal{L}_\alpha(m) = VD_\alpha^+U^T(A\mathbf{f} + \varepsilon) = VD_\alpha^+DV^T\mathbf{f} + VD_\alpha^+U^T\varepsilon. \quad (4.7)$$

The vector $VD_\alpha^+DV^T\mathbf{f}$ is an approximation to \mathbf{f} , and the error term can be estimated as follows:

$$\|VD_\alpha^+U^T\varepsilon\| \leq \|VD_\alpha^+U^T\| \|\varepsilon\| = \|D_\alpha^+\| \|\varepsilon\| = d_{r_\alpha}^{-1} \|\varepsilon\|. \quad (4.8)$$

By the ordering (3.15) of singular values we have

$$d_1^{-1} \leq d_2^{-1} \leq \cdots \leq d_r^{-1},$$

and by (4.8) the noise gets amplified in the inversion less and less if we keep fewer singular values (or, equivalently, increase α).

We see from definition (4.6) and by denoting $\mathbf{a} := D_\alpha^+U^T m$ that the reconstruction is a linear combination of the columns V_1, \dots, V_n of the matrix $V = [V_1 \ V_2 \ \cdots \ V_n]$:

$$\mathcal{L}_\alpha(m) = V\mathbf{a} = a_1V_1 + \cdots + a_nV_n.$$

Thus the columns V_1, \dots, V_n , called *singular vectors*, are the building blocks of any reconstruction using truncated SVD.

4.3 Measuring the quality of reconstructions

It is important to have a quantitative measure available for the closeness of a reconstruction to the actual measured object. Of course, such a measure makes sense only with simulated data, when the original object is known.

Denote the quantity of interest by $\mathbf{f} \in \mathbb{R}^n$ and assume that a regularized reconstruction $T(\mathbf{m}) \in \mathbb{R}^n$ is computed from noisy data $\mathbf{m} = A\mathbf{f} + \varepsilon \in \mathbb{R}^k$. We will use throughout the book the following definition of relative error:

$$\frac{\|\mathbf{f} - T(\mathbf{f})\|_2}{\|\mathbf{f}\|_2} \cdot 100\%, \quad (4.9)$$

where $\|\mathbf{f}\|_2^2 = \sum_{j=1}^n |\mathbf{f}_j|^2$.

Sometimes it is more informative to study relative errors in other norms than the Euclidean norm. For example we may consider

$$\frac{\|\mathbf{f} - T(\mathbf{f})\|_p}{\|\mathbf{f}\|_p} \cdot 100\%, \quad (4.10)$$

where the p -norm is defined for $1 < p < \infty$ by

$$\|\mathbf{f}\|_p^p = \left(\sum_{j=1}^n |\mathbf{f}_j|^p \right)^{1/p}$$

and for $p = \infty$ by

$$\|\mathbf{f}\|_\infty = \max_{j=1,\dots,n} |\mathbf{f}_j|.$$

It is important to realize that the relative error measures (4.9) and (4.10) do not necessarily coincide with the visual quality of the reconstruction as judged by a human observer. The measures can be used for detecting and quantifying gross deviations from the desired outcome, but if one reconstruction has relative error 20% and another one 22%, the latter is not necessarily worse. Defining what is a good reconstruction depends on the application. However, it is always valuable to calculate a quantitative error such as (4.9) or (4.10) so that gross errors will not go undetected.

4.4 TSVD for the guiding examples

4.4.1 TSVD for deconvolution

Let us study truncated SVD in the case of one-dimensional deconvolution. Figure 4.2 shows reconstructions computed from data corrupted by random noise with standard deviation $\sigma = 0.05 \cdot \max |f(x)|$.

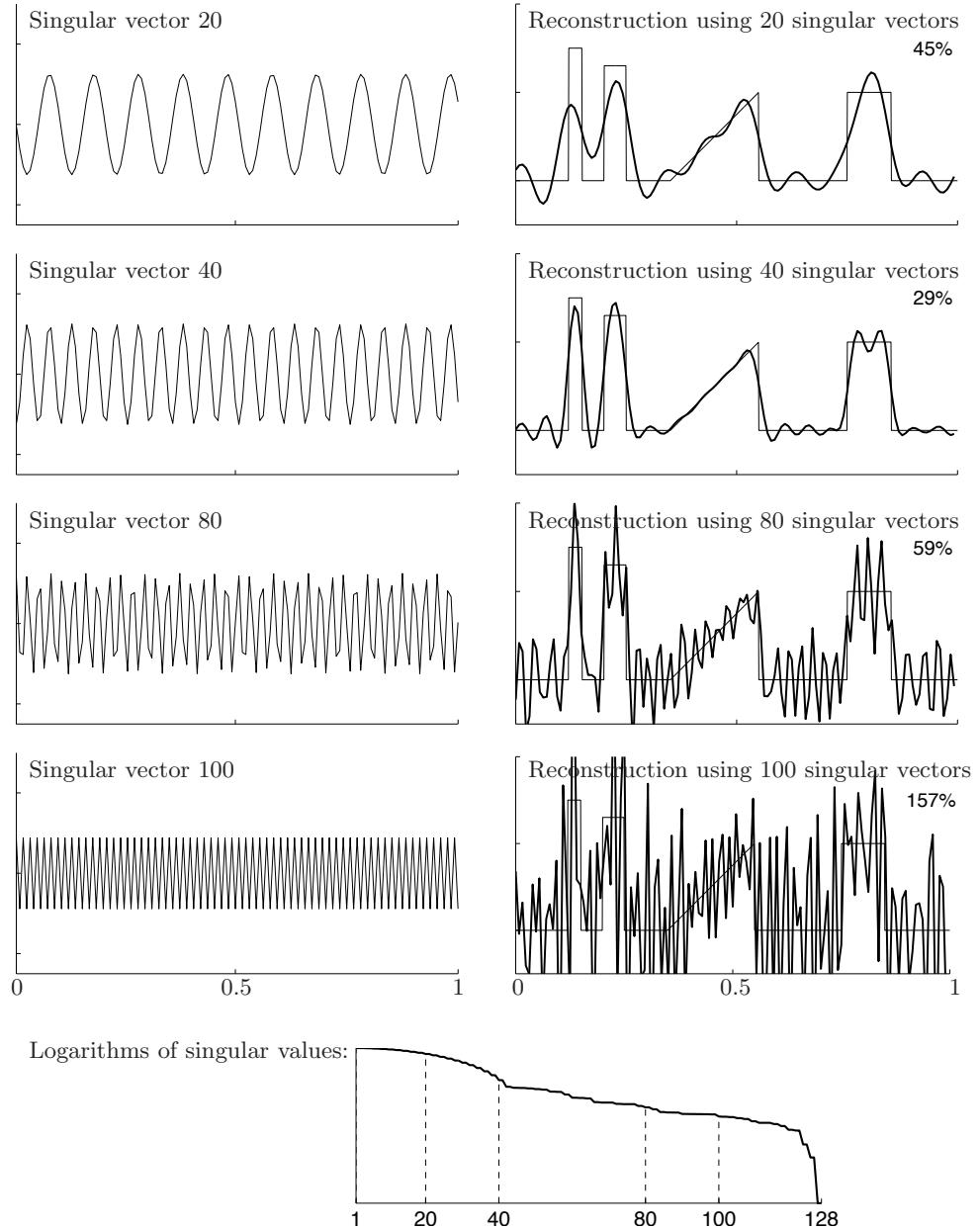


Figure 4.1: Left column: singular vectors related to the one-dimensional 128×128 convolution matrix. Right column: reconstructions (thick line) using all singular vectors up to the number shown on the left in the truncated SVD. Original signal is drawn with thin line for comparison. The percentages shown are relative errors of reconstructions. Bottom: logarithmic plot of the singular values.

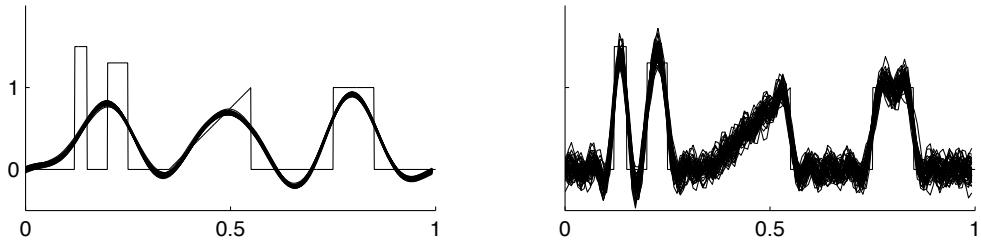


Figure 4.2: Study of noise-robustness of truncated singular value decomposition reconstructions. Here $\sigma = 0.05 \cdot \max |f(x)|$, and 50 independent realizations of random white noise was used for both of the two plots. Left: 50 reconstructions using the first 10 singular vectors. Right: 50 reconstructions using the first 40 singular vectors. The axis limits and scales are the same in both plots.

We next study the noise-robustness of truncated SVD reconstructions. In Figure 4.2 we used 50 realizations of random noise for computing truncated SVD reconstructions using the first 10 singular vectors, and another 50 realizations for computing reconstructions using the first 40 singular vectors. We see in the figure that there is much more variation in the set of reconstructions using 40 singular vectors, implying that the reconstructions with only 10 singular vectors are more robust against noise. However, the reconstructions from 10 singular vectors are also more erroneous, so there is a trade-off between accuracy and noise-robustness. This is a typical feature of solutions to inverse problems.

4.4.2 TSVD for backward heat propagation

To study the method of TSVD for the backward heat propagation problem (2.21)–(2.22), we apply the TSVD to the matrix defined by (3.6), and to avoid an inverse crime, we create data by the finite-difference solution of the forward problem presented in Section 2.2.2. The results of this method applied to the example with solution $u(x, 0) = 10 \sin 2x$, $0 < x < \pi$, are shown in Figure 4.3 using noise-free data and noisy data with $\sigma = 0.10 \cdot \max |u(x, T)|$. Notice that this method is considerably more robust than the method in Section 2.2.3, and we are able to use a much higher noise level and still obtain good reconstructions of the initial temperature. The reconstructions in Figure 4.3 are from data measured at $T = 0.1$. It is an exercise to reconstruct the initial temperature from later times such as $T = 0.4$ (see Exercise 4.4.1).

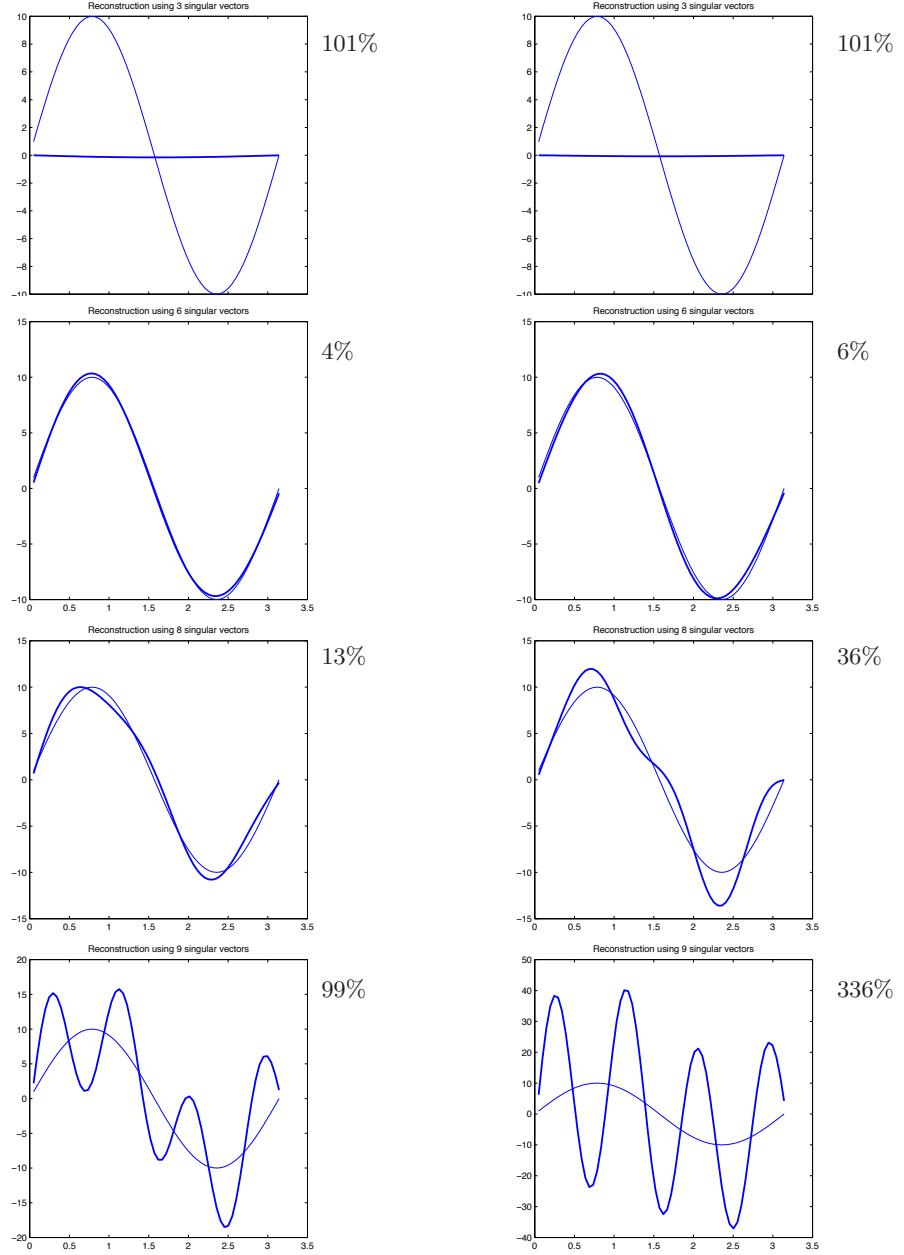


Figure 4.3: Left column: reconstructions (thick line) from noise-free data using all singular vectors up to the number shown. The actual solution is drawn with a thin line for comparison. The percentages shown are relative sup norm errors of the reconstructions. Right column: reconstructions (thick line) from noisy data with $\sigma = 0.10 \cdot \max |u(x, T)|$ using all singular vectors up to the number shown. The actual solution is drawn with a thin line for comparison. The percentages shown are relative sup norm errors of the reconstructions.

4.4.3 TSVD for X-ray tomography

We consider the tomographic measurement matrix A constructed in Section 2.3.5 for the resolution 50×50 and with 50 projection directions. We compute numerically the singular value decomposition of A , which is already quite intensive computationally as the size of A is 3750×2500 . The singular values are plotted in Figures 3.6 and 4.4.

We use formula (4.6) to compute reconstructions of the 50×50 Shepp-Logan phantom from noisy data with no inverse crime. We choose the truncation indices 200, 300, 1500 and 2000; the first two were chosen because the singular vectors 200 and 300 are especially beautiful. The reason for including number 1500 is that it is one of the choices yielding the minimal relative error of 48% using formula (4.9). Namely, all choices between 1440 and 1680 yield relative error of 48%, and any other choice leads to a larger error. Number 2000 was included to show what happens to the reconstruction when we start to use the really small singular values. Figure 4.4 shows the reconstructions and some of the singular vectors.

Exercise 4.4.1 Compute reconstructions using TSVD for the example of Section 4.4.2 from data measured at time $T = 0.4$. Use the finite-difference method to create the data and compute reconstructions from noise-free data and data corrupted with noise with $\sigma = 0.10 \cdot \max |f(x)|$. Produce plots as in Figure 4.3.

Exercise 4.4.2 Using the kernel

$$f(x) = \begin{cases} 5x - 0.5 & \text{for } 0.1 \leq x \leq 0.2, \\ -5x + 1.5 & \text{for } 0.2 < x \leq 0.3, \\ 1 & \text{for } 0.5 \leq x \leq 0.6, \\ 0.2 & \text{for } 0.6 < x < 0.7, \\ 1.3 & \text{for } 0.7 \leq x \leq 0.8, \\ 0 & \text{otherwise,} \end{cases}$$

add noise with amplitude $\sigma = 0.05$ and solve the 1-D deconvolution problem, reproducing plots as in Figure 4.1.

Exercise 4.4.3 Using the noisy kernel from Exercise 4.4.2 study the noise robustness of the TSVD reconstructions.

Exercise 4.4.4 Using data at resolution 50×50 and only 12 projection directions reproduce Figure 4.4 with the corresponding results.

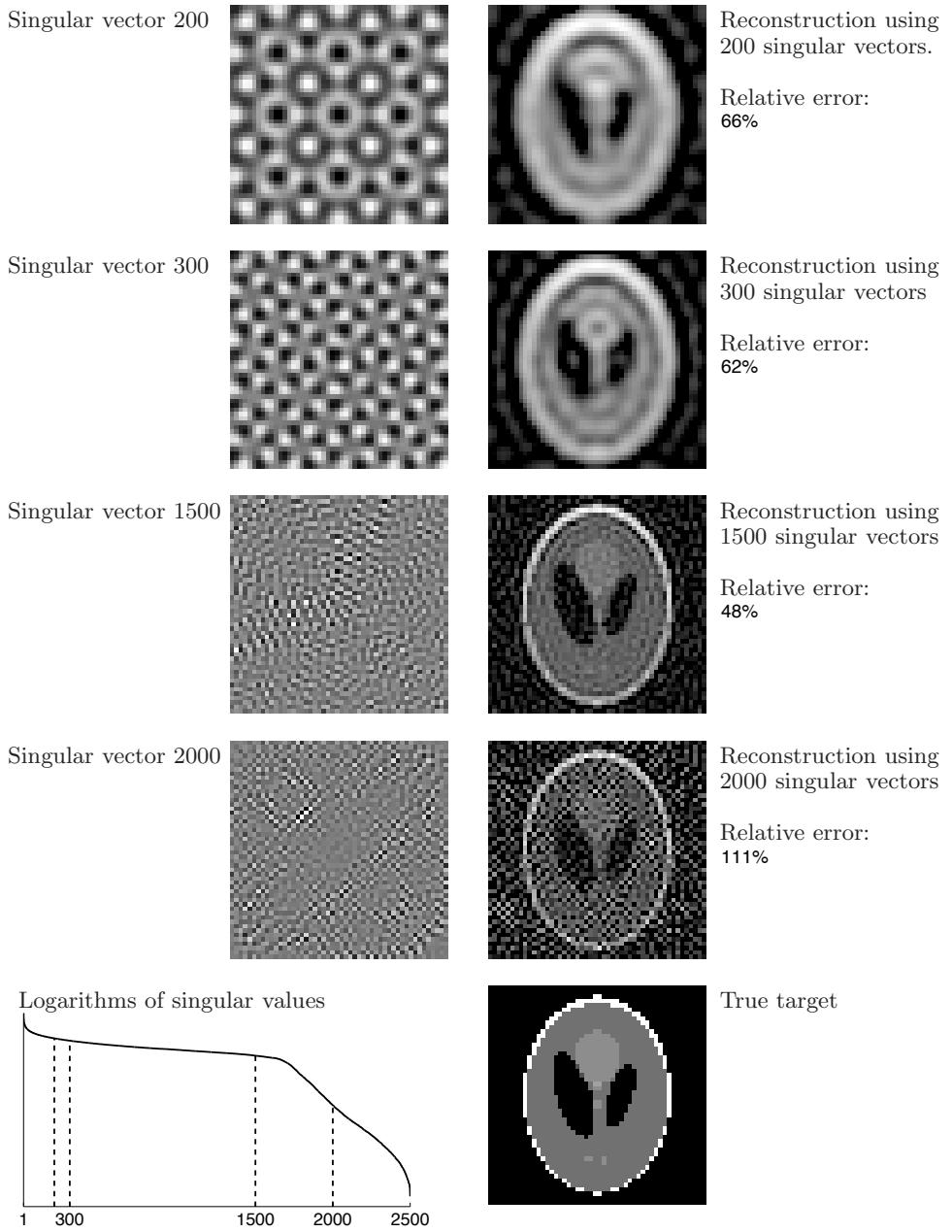


Figure 4.4: Study of truncated singular value decomposition for X-ray tomography at resolution 50×50 and 50 projection directions. Left column: singular vectors related to tomography matrix. Right column: reconstructions using all singular vectors up to the number shown on the left in the truncated SVD. The percentages shown are relative square norm errors of reconstructions.

Chapter 5

Tikhonov Regularization

Tikhonov regularization is typically the method of first choice for linear problems. It provides some smoothing, and generalized Tikhonov regularization provides an opportunity to incorporate known properties of the solution into the solution method. As we shall see in this chapter, it is simple to implement, but introduces the classic question, “How can I choose the regularization parameter?” Two popular, but not always reliable methods for selecting the regularization parameter, the Morozov discrepancy principle and the L-curve method, are explained in Section 5.4. In fact, there is no known method for choosing the regularization parameter that results in an optimal solution, but it is important to be aware of these widely-used methods. The results of using Tikhonov regularization on our three guiding examples is demonstrated in this chapter and through the exercises.

5.1 Classical Tikhonov regularization

The Tikhonov regularized solution of equation $\mathbf{m} = A\mathbf{f} + \varepsilon$ is the vector $T_\alpha(\mathbf{m}) \in \mathbb{R}^n$ that minimizes the expression

$$\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|^2 + \alpha\|T_\alpha(\mathbf{m})\|^2,$$

where $\alpha > 0$ is called a regularization parameter. We denote

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \|A\mathbf{z} - \mathbf{m}\|^2 + \alpha\|\mathbf{z}\|^2 \right\}. \quad (5.1)$$

Tikhonov regularization can be understood as a balance between two requirements:

- (i) $T_\alpha(\mathbf{m})$ should give a small residual $AT_\alpha(\mathbf{m}) - \mathbf{m}$,

(ii) $T_\alpha(\mathbf{m})$ should be small in L^2 norm.

The regularization parameter $\alpha > 0$ can be used to “tune” the balance.

Note that in inverse problems there are typically infinitely many choices of $T_\alpha(\mathbf{m})$ satisfying (i), and one of the roles of (ii) is to make the solution unique.

Theorem 5.1.1 *Let A be a $k \times n$ matrix. The Tikhonov regularized solution for equation $\mathbf{m} = A\mathbf{f} + \varepsilon$ is given by*

$$T_\alpha(\mathbf{m}) = V\mathcal{D}_\alpha^+ U^T \mathbf{m}, \quad (5.2)$$

where $A = UDV^T$ is the singular value decomposition, and

$$\mathcal{D}_\alpha^+ = \text{diag} \left(\frac{d_1}{d_1^2 + \alpha}, \dots, \frac{d_{\min(k,n)}}{d_{\min(k,n)}^2 + \alpha} \right) \in \mathbb{R}^{n \times k}. \quad (5.3)$$

Proof: Write $T_\alpha(\mathbf{m}) \in \mathbb{R}^n$ as linear combination of column vectors of the matrix V : $T_\alpha(\mathbf{m}) = \sum_{j=1}^n a_j V_j = V\mathbf{a}$. Set $\mathbf{m}' = U^T \mathbf{m}$ and compute

$$\begin{aligned} & \|AT_\alpha(\mathbf{m}) - \mathbf{m}\|^2 + \alpha\|T_\alpha(\mathbf{m})\|^2 \\ &= \|UDV^T V\mathbf{a} - UU^T \mathbf{m}' + \alpha\|V\mathbf{a}\|^2 \\ &= \|D\mathbf{a} - \mathbf{m}'\|^2 + \alpha\|\mathbf{a}\|^2 \\ &= \sum_{j=1}^r (d_j a_j - \mathbf{m}'_j)^2 + \sum_{j=r+1}^k (\mathbf{m}'_j)^2 + \alpha \sum_{j=1}^n a_j^2 \\ &= \sum_{j=1}^r (d_j^2 + \alpha) \left(a_j^2 - 2 \frac{d_j \mathbf{m}'_j}{d_j^2 + \alpha} a_j \right) + \alpha \sum_{j=r+1}^n a_j^2 + \sum_{j=1}^k (\mathbf{m}'_j)^2 \quad (5.4) \end{aligned}$$

$$\begin{aligned} &= \sum_{j=1}^r \left(d_j^2 + \alpha \right) \left(a_j - \frac{d_j \mathbf{m}'_j}{d_j^2 + \alpha} \right)^2 + \alpha \sum_{j=r+1}^n a_j^2 \\ &\quad - \sum_{j=1}^r \frac{(d_j \mathbf{m}'_j)^2}{d_j^2 + \alpha} + \sum_{j=1}^k (\mathbf{m}'_j)^2, \quad (5.5) \end{aligned}$$

where completing the square in the leftmost term in (5.4) yields (5.5). Our task is to choose such values for the parameters a_1, \dots, a_n that (5.5) attains its minimum. Clearly the correct choice is

$$a_j = \begin{cases} \frac{d_j}{d_j^2 + \alpha} \mathbf{m}'_j, & 1 \leq j \leq r, \\ 0, & r + 1 \leq j \leq n, \end{cases}$$

or in short $\mathbf{a} = \mathcal{D}_\alpha^+ \mathbf{m}'$. □

Recall from Section 4.2 that the TSVD solution is given by

$$\mathcal{L}_\alpha(\mathbf{m}) = V D_\alpha^+ U^T \mathbf{m} = \sum_{i=1}^{r_\alpha} \frac{\mathbf{u}_i^T \mathbf{m}}{d_i} \mathbf{v}_i \quad (5.6)$$

where the \mathbf{u}_i and \mathbf{v}_i are the columns of the matrices U and V , respectively, while

$$T_\alpha(\mathbf{m}) = V \mathcal{D}_\alpha^+ U^T \mathbf{m} = \sum_{i=1}^r \left(\frac{d_i^2}{d_i^2 + \alpha} \right) \frac{\mathbf{u}_i^T \mathbf{m}}{d_i} \mathbf{v}_i = \sum_{i=1}^r \left(\frac{d_i}{d_i^2 + \alpha} \right) (\mathbf{u}_i^T \mathbf{m}) \mathbf{v}_i. \quad (5.7)$$

Recall that $r_\alpha \leq r$. Thus, we see that the entries of \mathcal{D}_α^+ effectively weight the contributions of the vectors in the SVD, and if the regularization parameter α is sufficiently small (smaller than the smallest singular value), the Tikhonov regularized solution is essentially the same as the solution obtained by the SVD. By increasing α , less weight is placed on the small singular values, which also correspond to the highly oscillatory right singular vectors.

5.1.1 Tikhonov regularization for the deconvolution problem

Let us apply Tikhonov regularization to our basic test problem of one-dimensional deconvolution. In Figure 5.1 we see the Tikhonov regularized solutions corresponding to four different choices of regularization parameter. To investigate the noise-robustness of the Tikhonov regularized solutions, reconstructions from 50 data sets with independent realizations of random white noise with $\sigma = 0.05$ were computed. The plots are superimposed in Figure 5.2.

5.1.2 Tikhonov regularization for backward heat propagation

We apply Tikhonov regularization to the backward heat propagation problem with final time data at $T = 0.1$, computing data \mathbf{m} from the finite difference method described in Section 4.4.2 with noise added with standard deviation $\sigma = 0.10 \cdot \max |u(x, T)|$. The matrix A was constructed as in (3.6), and the minimization problem (5.1) was solved for regularization parameter α ranging from 10^{-5} to 10^{-2} . The results for a selection of regularization parameters are given in Figure 5.3. It is clear from these figures that a choice of α that is too large overdamps the solution while a small

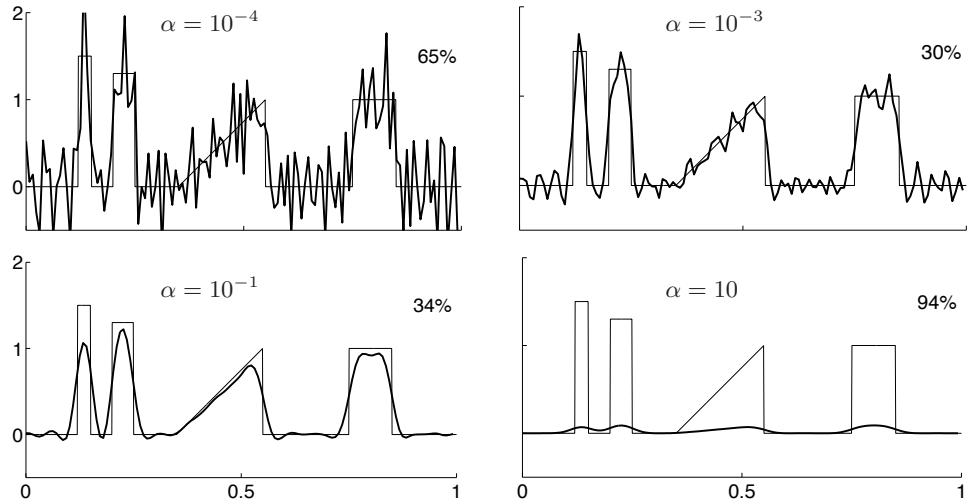


Figure 5.1: Tikhonov regularized reconstructions. The percentages shown are relative errors of reconstructions.

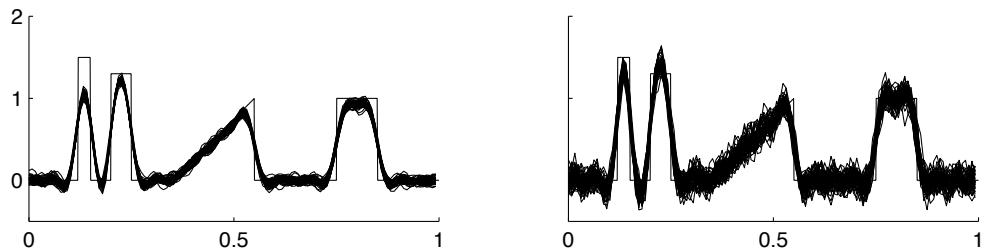


Figure 5.2: Study of noise-robustness of Tikhonov regularized reconstructions. Here the standard deviation of the noise level is $\sigma = 0.05 \cdot \max |f(x)|$, and 50 independent realizations of random white noise was used for both of the two plots. Left: regularization parameter $\alpha = 10^{-1}$. Right: regularization parameter $\alpha = 10^{-2}$.

choice that is too small results in wild oscillations. However, here the choice $\alpha = 10^{-2}$ is just right and results in a good reconstruction. Exercises applying Tikhonov regularization to the problem of backward heat propagation are found in Section 5.4 after we have discussed methods of choosing the regularization parameter.

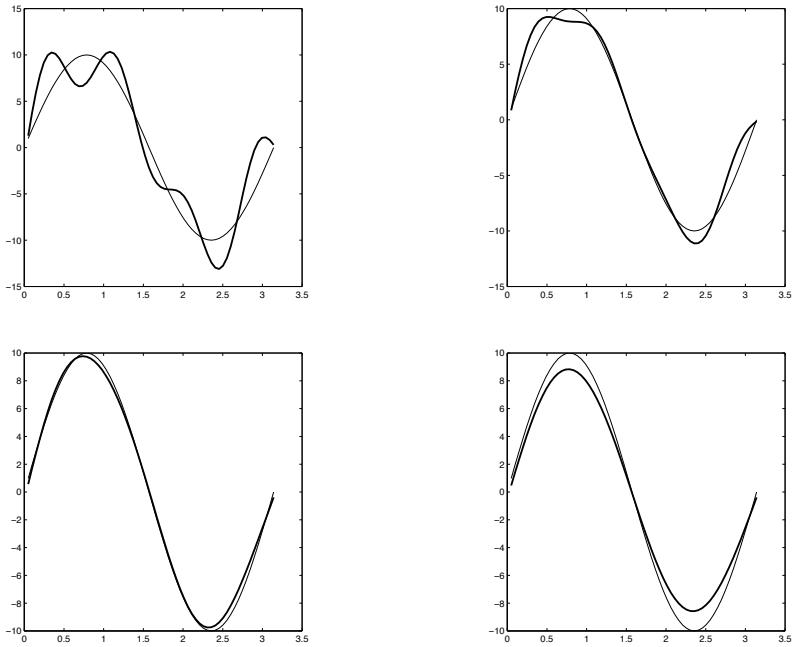


Figure 5.3: Tikhonov regularized reconstructions (dark line) with the actual solution superimposed (thin line) for the backward heat propagation problem of Section 2.2. The regularization parameter is $\alpha = 9 \times 10^{-5}$ (upper left), $\alpha = 4 \times 10^{-4}$ (upper right), $\alpha = 1 \times 10^{-2}$ (lower left), $\alpha = 7 \times 10^{-2}$ (lower right). The L^2 -norm relative errors of reconstructions are (in the same order) 44%, 21%, 7%, 14%.

Exercise 5.1.1 Show that the matrix $A^T A + \alpha I$ is always invertible when $\alpha > 0$ and A is an arbitrary $k \times n$ matrix. Hint: use SVD.

Exercise 5.1.2 Using the kernel from Exercise 4.4.2 with added noise of amplitude $\sigma = 0.05$, compute deconvolutions using classical Tikhonov regu-

larization with regularization parameter ranging from $\alpha = 10^0$ to $\alpha = 10^{-5}$ in increments of powers of 10. Create plots as in Figure 5.1.

5.2 Normal equations and stacked form

Consider the quadratic functional $Q_\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by

$$Q_\alpha(\mathbf{f}) = \|A\mathbf{f} - \mathbf{m}\|^2 + \alpha\|\mathbf{f}\|^2.$$

It can be proven that Q_α has a unique minimum for any $\alpha > 0$. The minimizer $T_\alpha(\mathbf{m})$ (i.e. the Tikhonov regularized solution of $\mathbf{m} = A\mathbf{f} + \varepsilon$) satisfies

$$0 = \frac{d}{dt} \left\{ \|A(T_\alpha(\mathbf{m}) + t\mathbf{w}) - \mathbf{m}\|^2 + \alpha\|T_\alpha(\mathbf{m}) + t\mathbf{w}\|^2 \right\} \Big|_{t=0}$$

for any $\mathbf{w} \in \mathbb{R}^n$.

Compute

$$\begin{aligned} & \frac{d}{dt} \|A(T_\alpha(\mathbf{m}) + t\mathbf{w}) - \mathbf{m}\|^2 \Big|_{t=0} \\ &= \frac{d}{dt} \langle AT_\alpha(\mathbf{m}) + tA\mathbf{w} - \mathbf{m}, AT_\alpha(\mathbf{m}) + tA\mathbf{w} - \mathbf{m} \rangle \Big|_{t=0} \\ &= \frac{d}{dt} \left\{ \|AT_\alpha(\mathbf{m})\|^2 + 2t\langle AT_\alpha(\mathbf{m}), A\mathbf{w} \rangle + t^2\|A\mathbf{w}\|^2 \right. \\ &\quad \left. - 2t\langle \mathbf{m}, A\mathbf{w} \rangle - 2\langle AT_\alpha(\mathbf{m}), \mathbf{m} \rangle + \|\mathbf{m}\|^2 \right\} \Big|_{t=0} \\ &= 2\langle AT_\alpha(\mathbf{m}), A\mathbf{w} \rangle - 2\langle \mathbf{m}, A\mathbf{w} \rangle, \end{aligned}$$

and

$$\begin{aligned} & \frac{d}{dt} \alpha \langle T_\alpha(\mathbf{m}) + t\mathbf{w}, T_\alpha(\mathbf{m}) + t\mathbf{w} \rangle \Big|_{t=0} \\ &= \alpha \frac{d}{dt} \left\{ \|T_\alpha(\mathbf{m})\|^2 + 2t\langle T_\alpha(\mathbf{m}), \mathbf{w} \rangle + t^2\|\mathbf{w}\|^2 \right\} \Big|_{t=0} \\ &= 2\alpha \langle T_\alpha(\mathbf{m}), \mathbf{w} \rangle. \end{aligned}$$

Thus, we have $\langle AT_\alpha(\mathbf{m}) - \mathbf{m}, A\mathbf{w} \rangle + \alpha \langle T_\alpha(\mathbf{m}), \mathbf{w} \rangle = 0$, and by taking the transpose,

$$\langle A^T AT_\alpha(\mathbf{m}) - A^T \mathbf{m}, \mathbf{w} \rangle + \alpha \langle T_\alpha(\mathbf{m}), \mathbf{w} \rangle = 0.$$

This results in the variational form

$$\langle (A^T A + \alpha I) T_\alpha(\mathbf{m}) - A^T \mathbf{m}, \mathbf{w} \rangle = 0. \quad (5.8)$$

Since (5.8) holds for any nonzero $\mathbf{w} \in \mathbb{R}^n$, we necessarily have $(A^T A + \alpha I) T_\alpha(\mathbf{m}) = A^T \mathbf{m}$. So the Tikhonov regularized solution $T_\alpha(\mathbf{m})$ satisfies

$$T_\alpha(\mathbf{m}) = (A^T A + \alpha I)^{-1} A^T \mathbf{m}, \quad (5.9)$$

and actually (5.9) can be used for computing $T_\alpha(\mathbf{m})$ defined in the basic situation (5.1).

In the generalized case of (5.14) we get by a similar computation

$$T_\alpha(\mathbf{m}) = (A^T A + \alpha L^T L)^{-1} A^T \mathbf{m}. \quad (5.10)$$

Next we will derive a computationally attractive *stacked form* version of (5.2).

We rethink problem (5.2) so that we have two measurements on \mathbf{f} that we minimize simultaneously in the least squares sense. Namely, we consider both equations $A\mathbf{f} = \mathbf{m}$ and $\mathbf{f} = 0$ as independent measurements of the same object \mathbf{f} , where $A \in \mathbb{R}^{k \times n}$. Now we stack the matrices and right hand sides so that the regularization parameter $\alpha > 0$ is involved correctly:

$$\begin{bmatrix} A \\ \sqrt{\alpha} \end{bmatrix} \mathbf{f} = \begin{bmatrix} \mathbf{m} \\ 0 \end{bmatrix}. \quad (5.11)$$

We write (5.11) as $\tilde{A}\mathbf{f} = \tilde{\mathbf{m}}$ and solve for $T_\alpha(\mathbf{m})$ defined in (5.10) in MATLAB by

$$\mathbf{f} = \tilde{A} \setminus \tilde{\mathbf{m}}, \quad (5.12)$$

where \setminus stands for finding the least squares solution. This is a good method for medium-dimensional inverse problems, where n and k are of the order $\sim 10^3$. Formula (5.12) is applicable to higher-dimensional problems than formula (5.2) since there is no need to compute the SVD for (5.12).

Why would (5.12) be equivalent to (5.9)? In general, a computation similar to the above shows that a vector \mathbf{z}_0 , defined as the minimizer

$$\mathbf{z}_0 = \arg \min_{\mathbf{z}} \|B\mathbf{z} - \mathbf{b}\|^2,$$

satisfies the normal equations $B^T B \mathbf{z}_0 = B^T \mathbf{b}$. In this case the minimizing \mathbf{z}_0 is called the least squares solution to equation $B\mathbf{z} = \mathbf{b}$. In the context of

our stacked form formalism, the least squares solution of (5.11) satisfies the normal equations

$$\tilde{A}^T \tilde{A} \mathbf{f} = \tilde{A}^T \tilde{\mathbf{m}}.$$

However,

$$\tilde{A}^T \tilde{A} = \begin{bmatrix} A^T & \sqrt{\alpha}I \end{bmatrix} \begin{bmatrix} A \\ \sqrt{\alpha}I \end{bmatrix} = A^T A + \alpha I,$$

and

$$\tilde{A}^T \tilde{\mathbf{m}} = \begin{bmatrix} A^T & \sqrt{\alpha}I \end{bmatrix} \begin{bmatrix} \mathbf{m} \\ 0 \end{bmatrix} = A^T \mathbf{m},$$

so it follows that $(A^T A + \alpha I) \mathbf{f} = A^T \mathbf{m}$.

As an example, consider the tomographic measurement matrix A constructed in Section 2.3.5 for the resolution 50×50 and with 50 projection directions. We can use the stacked form approach (5.11) and (5.12) to avoid the expensive computation of the singular value decomposition of A . The results are found in Figure 5.4.

5.3 Generalized Tikhonov regularization

Sometimes we have *a priori* information about the solution of the inverse problem. For example, we may know that \mathbf{f} is close to a vector $\mathbf{f}_* \in \mathbb{R}^n$; then we minimize

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \|A\mathbf{z} - \mathbf{m}\|^2 + \alpha \|\mathbf{z} - \mathbf{f}_*\|^2 \right\}. \quad (5.13)$$

Another typical situation is that \mathbf{f} is known to be smooth. Then we minimize

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \|A\mathbf{z} - \mathbf{m}\|^2 + \alpha \|L\mathbf{z}\|^2 \right\}. \quad (5.14)$$

or

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \|A\mathbf{z} - \mathbf{m}\|^2 + \alpha \|L(\mathbf{z} - \mathbf{f}_*)\|^2 \right\}. \quad (5.15)$$

where L is a discretized differential operator.

For example in dimension 1, representing the vector \mathbf{f} as a continuous function f with $f(s_j) = \mathbf{f}_j$, we can discretize the derivative of the continuum by the difference quotient

$$\frac{df}{ds}(s_j) \approx \frac{f(s_{j+1}) - f(s_j)}{\Delta s} = \frac{\mathbf{f}_{j+1} - \mathbf{f}_j}{\Delta s}.$$

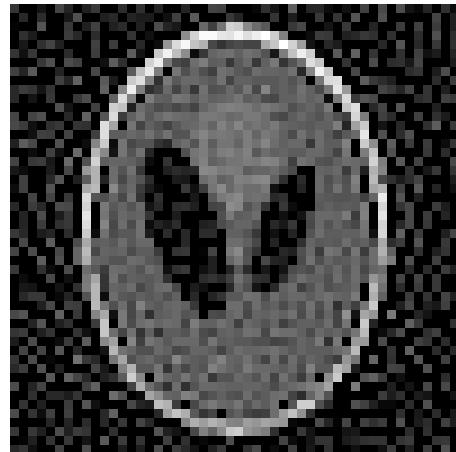
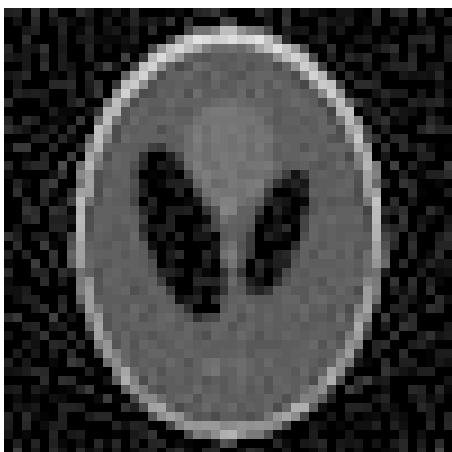
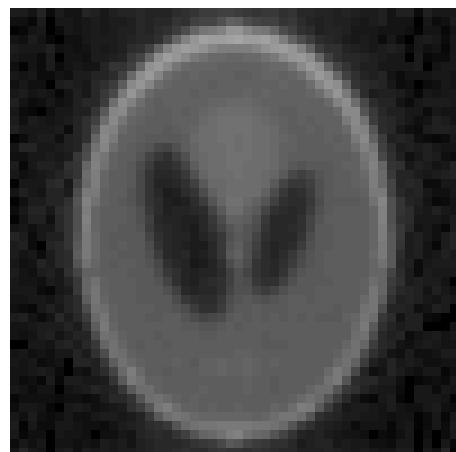
Original: 50×50 Shepp-Logan phantomToo small parameter $\alpha = 0.1$, error 56%Medium parameter $\alpha = 10$, error 43%Too large parameter $\alpha = 100$, error 59%

Figure 5.4: Tikhonov regularized tomographic reconstructions with $L = I$. The resolution is 50×50 , and 50 projection directions were used evenly distributed along a full circle. The percentages shown are relative errors of reconstructions computed using formula (4.9).

This leads to the discrete differentiation matrix

$$L = \frac{1}{\Delta s} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \ddots & \\ & & & & & & \ddots \\ 0 & \cdots & & 0 & -1 & 1 & 0 \\ 0 & \cdots & & 0 & 0 & -1 & 1 \\ 1 & \cdots & & 0 & 0 & 0 & -1 \end{bmatrix}. \quad (5.16)$$

5.3.1 Generalized Tikhonov regularization for the deconvolution problem

Let us apply generalized Tikhonov regularization to our basic test problem of one-dimensional deconvolution. We take L as in formula (5.16) and use stacked form (5.12) for computing reconstructions with various choices of regularization parameter α . See Figure 5.5.

Compare Figure 5.5 to the classical Tikhonov regularization case $L = I$ shown in Figure 5.1. For large α the case $L = I$ leads to the reconstruction being close to zero, whereas in the case of the discrete derivative L given by (5.16) we see that large α takes the reconstruction close to a constant function, not necessarily zero. Also, using (5.16) seems to promote smoothness in the reconstructions, as expected.

Also, we study numerically the noise-robustness of the reconstructions.

Exercise 5.3.1 Show that the variational form corresponding to the minimization problem

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \{\|\mathbf{A}\mathbf{z} - \mathbf{m}\|^2 + \alpha\|L\mathbf{z}\|^2\}$$

is given by

$$\langle (\mathbf{A}^T \mathbf{A} + \alpha L^T L) T_\alpha(\mathbf{m}) - \mathbf{A}^T \mathbf{m}, \mathbf{w} \rangle = 0 \quad \text{for all } \mathbf{w} \in \mathbb{R}^n.$$

Exercise 5.3.2 Write the generalized Tikhonov problem

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \{\|\mathbf{A}\mathbf{z} - \mathbf{m}\|^2 + \alpha\|L(\mathbf{z} - \mathbf{f}_*)\|^2\}$$

in stacked form.

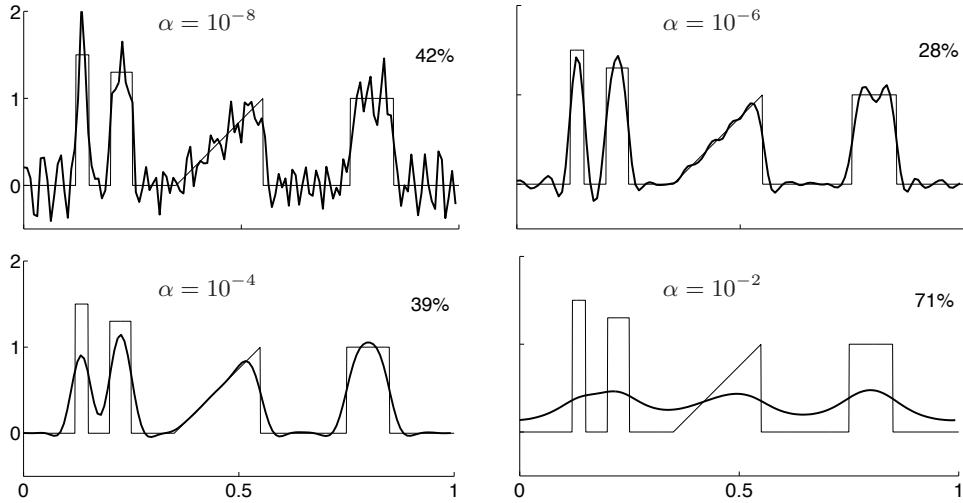


Figure 5.5: Generalized Tikhonov regularized reconstructions. The percentages shown are relative errors of reconstructions.

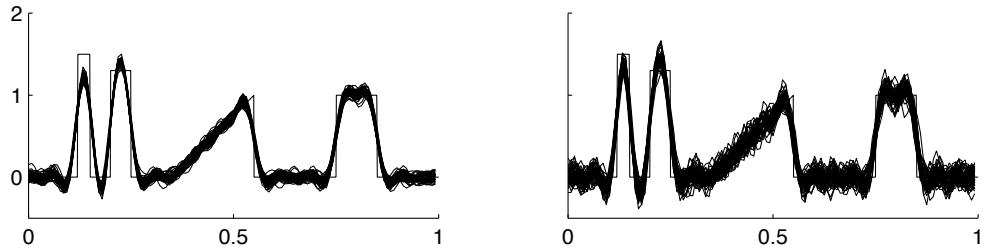


Figure 5.6: Study of noise-robustness of generalized Tikhonov regularized reconstructions. Here 50 independent realizations of random white noise was used for both of the two plots. The standard deviation of independent noise components was $\sigma = 0.05$. Left: regularization parameter $\alpha = 0.1$. Right: regularization parameter $\alpha = 0.01$.

Exercise 5.3.3 Experiment with generalized Tikhonov regularization on the problem of backward heat propagation.

Exercise 5.3.4 Choose L to be the Laplace operator and repeat the x-ray tomography example with generalized Tikhonov regularization and this choice of L . Note that in matrix form

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (5.17)$$

5.4 Choosing the regularization parameter

How does one choose the regularization parameter $\alpha > 0$ optimally? This is a difficult question and in general unsolved. There are some methods for choosing α that are based on the noise level in the data, for example Morozov's discrepancy principle: If we have an estimate on the magnitude of error in the data, then any solution that produces a measurement with error of the same magnitude is acceptable. Other methods, such as the L-curve method, are based on striking a balance between the norm of a penalty term and the norm of the residual. We discuss both methods here.

Some methods, such as the generalized cross-validation method [164] and recent methods with inexact knowledge of the noise level [187], are not discussed here due to space restrictions. We remark that a very recent sparsity-based choice rule is explained below in Section 6.3.

5.4.1 Morozov's discrepancy principle

Assume that $\mathbf{m} = A\mathbf{f} + \varepsilon$ and that we know the size of noise: $\|\varepsilon\| = \delta > 0$. Then $T_\alpha(\mathbf{m})$ is an acceptable reconstruction if

$$\|AT_\alpha(\mathbf{m}) - \mathbf{m}\| \leq \delta.$$

For example, if the elements of the noise vector $\varepsilon \in \mathbb{R}^k$ satisfy $\varepsilon_j \sim N(0, \sigma^2)$, then we can take $\delta = \sqrt{k}\sigma$ since the expectation of the size is $E(\|\varepsilon\|) = \sqrt{k}\sigma$.

The idea of Morozov discrepancy principle is to choose $\alpha > 0$ such that

$$\|AT_\alpha(\mathbf{m}) - \mathbf{m}\| = \delta.$$

Theorem 5.4.1 *The Morozov discrepancy principle gives a unique choice for $\alpha > 0$ if and only if δ satisfies*

$$\|P\mathbf{m}\| \leq \delta \leq \|\mathbf{m}\|,$$

where P is orthogonal projection to the subspace $\text{Coker}(A)$.

Proof: From the proof of Theorem 5.1.1 we find the equation

$$AT_\alpha(\mathbf{m}) = UDV^T V\mathcal{D}_\alpha^+ U^T \mathbf{m} = UDD_\alpha^+ \mathbf{m}',$$

so we have

$$\begin{aligned}\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|^2 &= \|D\mathcal{D}_\alpha^+ \mathbf{m}' - \mathbf{m}'\|^2 \\ &= \sum_{j=1}^{\min(k,n)} \left(\frac{d_j^2}{d_j^2 + \alpha} - 1 \right)^2 (\mathbf{m}'_j)^2 + \sum_{j=\min(k,n)+1}^k (\mathbf{m}'_j)^2 \\ &= \sum_{j=1}^r \left(\frac{\alpha}{d_j^2 + \alpha} \right)^2 (\mathbf{m}'_j)^2 + \sum_{j=r+1}^k (\mathbf{m}'_j)^2.\end{aligned}$$

From this expression we see that the mapping

$$\alpha \mapsto \|AT_\alpha(\mathbf{m}) - \mathbf{m}\|^2$$

is monotonically increasing and thus, noting the formal identity

$$\sum_{j=r+1}^k (\mathbf{m}'_j)^2 = \|AT_0(\mathbf{m}) - \mathbf{m}\|^2$$

we get

$$\sum_{j=r+1}^k (\mathbf{m}'_j)^2 \leq \|AT_\alpha(\mathbf{m}) - \mathbf{m}\|^2 \leq \lim_{\alpha \rightarrow \infty} \|AT_\alpha(\mathbf{m}) - \mathbf{m}\|^2 = \sum_{j=1}^k (\mathbf{m}'_j)^2,$$

and the claim follows from orthogonality of U . \square

Numerical implementation of Morozov's method is now simple. Just find the optimal α as the unique zero of the function

$$f(\alpha) = \sum_{j=1}^r \left(\frac{\alpha}{d_j^2 + \alpha} \right)^2 (\mathbf{m}'_j)^2 + \sum_{j=r+1}^k (\mathbf{m}'_j)^2 - \delta^2. \quad (5.18)$$

Let us try Morozov's method for our one-dimensional deconvolution model problem. We take $n = 128 = k$ and simulate convolution data avoiding inverse crime as explained in Section 2.1.4. Next we add a noise vector $\varepsilon \in \mathbb{R}^k$ whose elements satisfy $\varepsilon_j \sim N(0, \sigma^2)$ with $\sigma = 0.01$. Then $\delta = \sqrt{k}\sigma \approx 0.113$. See Figure 5.7 for the resulting reconstruction.

5.4.2 The L-curve method

As the method of Morozov does not apply to the generalized regularization formulas (5.13)–(5.15), we need to discuss alternative approaches. One possibility is to use the so-called L-curve method.

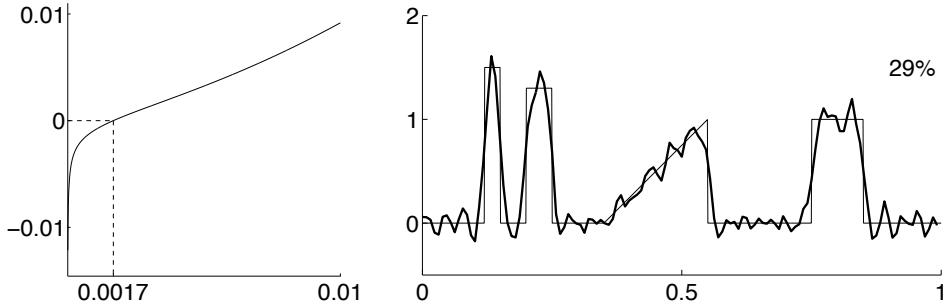


Figure 5.7: Morozov discrepancy principle and Tikhonov regularization for the one-dimensional deconvolution problem. Left: plot of the function $f(\alpha)$ defined in (5.18). Right: Tikhonov regularized deconvolution using parameter $\alpha = 0.0017$. Relative square norm error is 29%.

The idea of the L-curve method is to choose a collection of candidates for regularization parameter:

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_M < \infty,$$

and compute $T_{\alpha_j}(\mathbf{m})$ for each $1 \leq j \leq M$. Then the points

$$(\log \|AT_\alpha(\mathbf{m}) - \mathbf{m}\|, \log \|LT_\alpha(\mathbf{m})\|) \in \mathbb{R}^2 \quad (5.19)$$

are plotted in the plane, forming approximately a smooth curve. This curve has typically the shape of the letter L with smooth corner. The optimal value of α is thought to be found as near the corner as possible. See Figure 5.8 for an illustration of the L-curve.

This method is applicable to generalized Tikhonov regularization as well as the case where $L = I$. To understand the L-curve method in a bit more detail, we introduce the generalized singular value decomposition (GSVD) of the matrix pair (A, L) .

Definition 5.4.1 Let A be an $m \times n$ matrix and L a $p \times n$ matrix with $m \geq n \geq p$ where $\ker(A) \cap \ker(L) = \{0\}$ and L has full row rank ($\text{rank}(L) = p$). The generalized singular value decomposition (GSVD) of the matrix pair

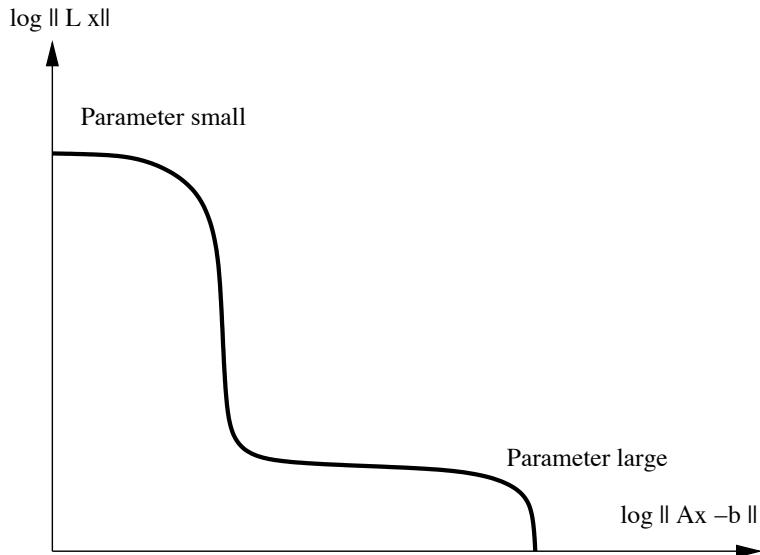


Figure 5.8: An idealized illustration of an L-curve formed by plotting a continuum of points defined by (5.19).

(A, L) is

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n-p} \end{bmatrix} W^{-1} \quad \text{and} \quad L = V \begin{bmatrix} M & 0 \end{bmatrix} W^{-1}$$

where U , $m \times n$, and V , $p \times p$, are orthonormal matrices, W is an $n \times n$ nonsingular matrix, $\Sigma = \text{diag}(\sigma_1, \sigma_2 \dots, \sigma_p)$ with $0 \leq \sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_p \leq 1$, $M = \text{diag}(\mu_1, \mu_2 \dots, \mu_p)$ with $1 \geq \mu_1 \geq \mu_2 \geq \dots \geq \mu_p \geq 0$ and $\sigma_i^2 + \mu_i^2 = 1$. The generalized singular values are defined as the ratios $\gamma_i = \frac{\sigma_i}{\mu_i}$. Since the σ_i are increasing and μ_i is decreasing we find

$$\gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \dots \leq \gamma_p.$$

Similarly to formula (5.7) the generalized Tikhonov regularized solution can be expressed in terms of the generalized singular values by

$$T_\alpha(\mathbf{m}) = \sum_{i=1}^r \frac{\gamma_i^2}{\gamma_i^2 + \alpha} \frac{\mathbf{u}_i^T \mathbf{m}}{\gamma_i} \mathbf{v}_i. \quad (5.20)$$

This allows us to write down an explicit expression for $\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|_2^2$ and

$\|LT_\alpha(\mathbf{m})\|^2$. That is,

$$\begin{aligned}\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|_2^2 &= \left\| \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n-p} \end{bmatrix} W^{-1} T_\alpha(\mathbf{m}) - U^T \mathbf{m} \right\|^2 - \|U^T \mathbf{m}\|^2 + \|\mathbf{m}\|^2 \\ &= \sum_{i=1}^p \left(\frac{\gamma_i^2}{\gamma_i^2 + \alpha^2} - 1 \right)^2 (\mathbf{u}_i^T \mathbf{m})^2 - \|U^T \mathbf{m}\|^2 + \|\mathbf{m}\|^2 \\ &= \sum_{i=1}^p \left(\frac{\alpha^2}{\gamma_i^2 + \delta^2} \right)^2 (\mathbf{u}_i^T \mathbf{m})^2 - \|U^T \mathbf{m}\|^2 + \|\mathbf{m}\|^2,\end{aligned}\quad (5.21)$$

and

$$\begin{aligned}\|LT_\alpha(\mathbf{m})\|^2 &= \left\| \begin{bmatrix} M & 0 \end{bmatrix} W^{-1} T_\alpha(\mathbf{m}) \right\|^2 \\ &= \|[M0] \begin{bmatrix} D & 0 \\ 0 & I_{n-p} \end{bmatrix} U^T \mathbf{m}\|^2 \\ &= \sum_{i=1}^p \left(\frac{\gamma_i}{\gamma_i^2 + \alpha^2} \right)^2 (\mathbf{u}_i^T \mathbf{m})^2.\end{aligned}\quad (5.22)$$

The above formula shows that $\|LT_\alpha(\mathbf{m})\|$ is a monotonically decreasing function and

$$\begin{aligned}\|LT_0(\mathbf{m})\|^2 &= \sum_{i=1}^p \frac{1}{\gamma_i^2} (\mathbf{u}_i^T \mathbf{m})^2 = \|L\mathcal{L}_\alpha(\mathbf{m})\|^2 \\ \|LT_\alpha(\mathbf{m})\|^2 &\rightarrow 0 \text{ as } \alpha \rightarrow \infty\end{aligned}$$

This explains the ordering of points along the L-curve: as the regularization parameter increases, the points move downward and to the right. Also, the L-curve is bounded from above and bounded to the left and right.

From [193] any point (a, b) on the curve $(\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|, \|LT_\alpha(\mathbf{m})\|)$ is a solution of the two constrained least square problems:

$$\begin{aligned}a &= \min\{\|A\mathbf{f} - \mathbf{m}\|\} \quad \text{subject to} \quad \|L\mathbf{f}\| \leq b, \quad 0 \leq b \leq \|LT_0(\mathbf{m})\| \\ b &= \min\{\|L\mathbf{f}\|\} \quad \text{subject to} \quad \|A\mathbf{f} - \mathbf{m}\| \leq a, \quad \|P\mathbf{m}\| \leq a \leq \|\mathbf{m}\|\end{aligned}$$

As a consequence, all approximate solutions of the regularization problem by other computations will generate a point in Figure 5.8 above the L-curve.

To understand the shape of the L-curve, assumptions are needed on the Fourier coefficients $c_i = u_i^T \mathbf{m}$. One can show that the faster the c_i decay to 0, the sharper the L-shaped corner. It can also be shown that a log-log

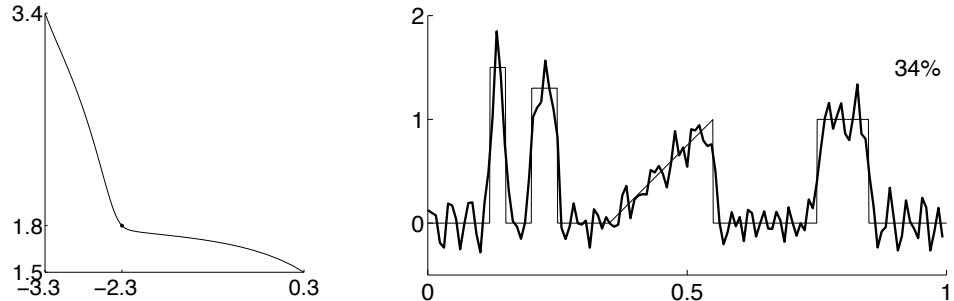


Figure 5.9: L-curve for the one-dimensional deconvolution problem.

scale emphasizes the L-shape. The short horizontal section of the L-curve in Figure 5.8 arises because when α is extremely small,

$$\begin{aligned}\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|_2^2 &\approx \sum_{i=1}^p \left(\frac{\alpha^2}{\gamma_i^2} \right)^2 (\mathbf{u}_i^T \mathbf{m})^2 - \|U^T \mathbf{m}\|^2 + \|\mathbf{m}\|^2 \\ \|LT_\alpha(\mathbf{m})\|^2 &\approx \|LT_0(\mathbf{m})\|^2 = \|L\mathcal{L}_\alpha(\mathbf{m})\|^2\end{aligned}$$

and so $\|LT_\alpha(\mathbf{m})\|^2$ is constant while $\|AT_\alpha(\mathbf{m}) - \mathbf{m}\|_2^2$ is increasing. For α extremely large we find a vertical line downward at $\|AT_\alpha(\mathbf{m}) - \mathbf{m}\| \approx \|\mathbf{m}\|$.

For more information about the L-curve method, see the book by Hansen [193] and references therein.

The L-curve method applied to our guiding examples can be found in Figures 5.9, 5.10, and 5.11 for the problems, respectively, of one-dimensional deconvolution, the problem of backward heat propagation from data at $T = 0.1$, and x-ray tomography for the resolution 50×50 and with 50 projection directions. It is clear from the results that the regularization parameter α corresponding to the corner of the L does not always result in the optimal reconstruction. This is explored further in the exercises, and the reader is referred to the articles [463, 189] for analysis of the nonconvergence of the L-curve method.

Exercise 5.4.1 Consider the regularized solution $T_\alpha(\mathbf{m})$ of equation $\mathbf{m} = \mathbf{Ax} + \varepsilon$ using truncated SVD with truncation index $p(\alpha)$ for $\alpha > 0$. Assume that the noise level is $\delta = \|\varepsilon\|$. Show that the discrepancy condition

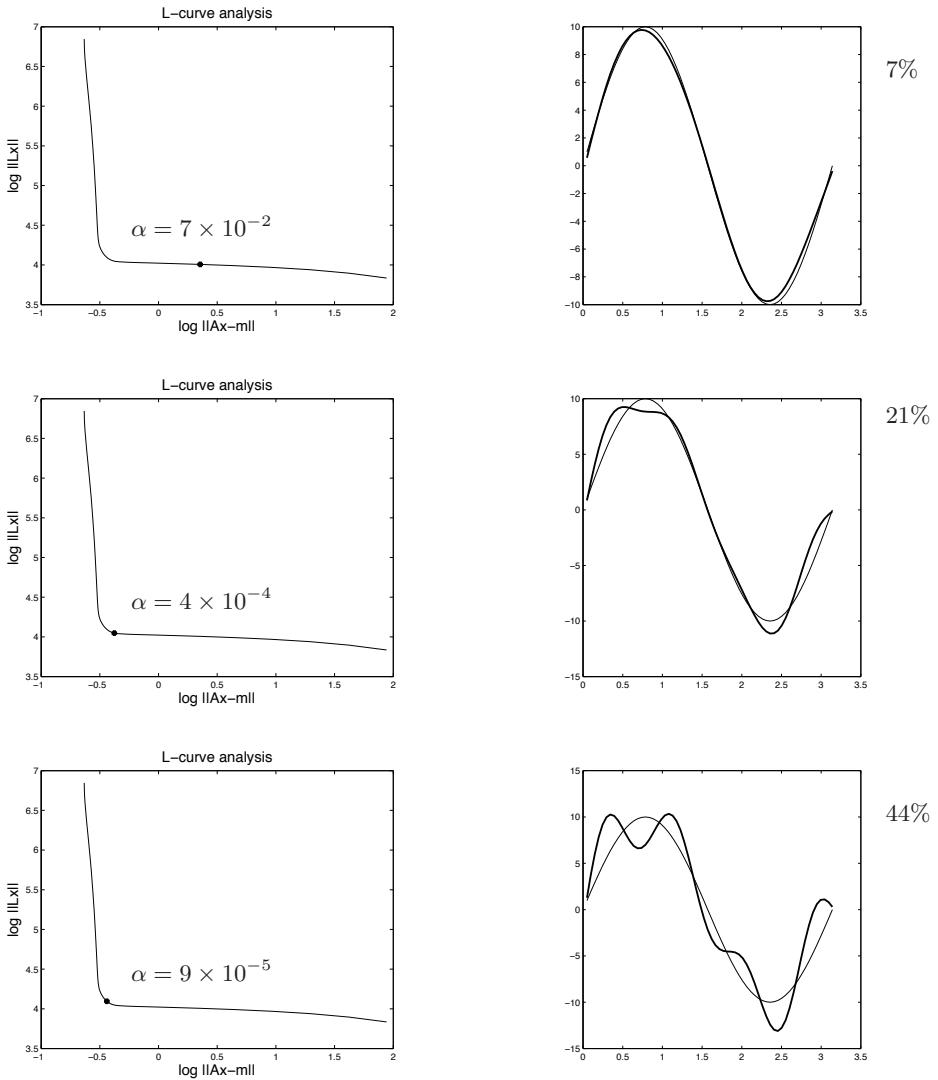


Figure 5.10: L-curves and the corresponding reconstructions for the backward heat problem. Notice that this is an example of non-convergence of the L-curve method. The best reconstruction does not correspond to a regularization parameter α coinciding with the corner of the L. The percentages represent the percent relative error.

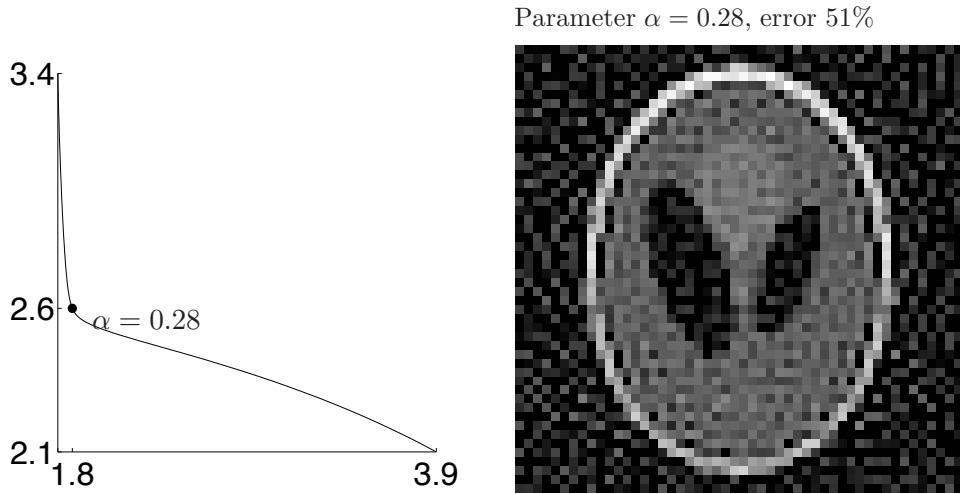


Figure 5.11: L-curve method for Tikhonov regularized tomographic reconstructions with $L = I$. The resolution is 50×50 , and 50 projection directions were used evenly distributed along a full circle. Compare to Figure 5.4.

$\|AT_\alpha(\mathbf{m}) - \mathbf{m}\| \leq \delta$ can be written in the form

$$\sum_{j=p(\alpha)+1}^m (y'_j)^2 \leq \delta^2.$$

(This is the equivalent of Morozov's discrepancy condition for truncated SVD.)

Exercise 5.4.2 Use the Morozov discrepancy principle for the backward heat propagation problem from $T = 0.1$ with added noise of 1%. Does the MDP predict an optimal parameter?

Exercise 5.4.3 Use Tikhonov regularization for the backward heat propagation problem from $T = 0.4$. Plot the L-curves and determine whether the corner corresponds to the best reconstruction in this case. What regularization parameter gives you the best reconstruction?

Exercise 5.4.4 Find the parameter for the one-dimensional deconvolution problem in Figure 5.9 that corresponds to the corner of the L. Find the parameter that results in the best reconstruction and plot its location on the

L-curve. What does its position tell you about the balance between a small norm residual and a small penalty term?

Exercise 5.4.5 Repeat Exercise 5.4.4 using the kernel defined in Exercise 4.4.2.

Exercise 5.4.6 Use instead the Morozov discrepancy principle to choose a regularization parameter in Exercise 5.4.5. Does it predict an optimal regularization parameter?

Exercise 5.4.7 Determine whether the parameter $\alpha = 0.28$ results in the best reconstruction for x-ray tomographic reconstructions of the Shepp-Logan phantom with resolution 50×50 and data from 50 projection directions, as shown in Figure 5.4. Clarify whether you are using a norm criterion to decide which reconstruction is best or whether you are choosing it visually. If $\alpha = 0.28$ is not best, determine what value of α is and where it lies on the L-curve.

5.5 Large-scale implementation

The formulation (5.10) of Tikhonov regularization is remarkable because it allows matrix-free implementation. Namely, assume that we have available computational routines called **Amult** and **Lmult** that take an arbitrary vector $\mathbf{z} \in \mathbb{R}^n$ as argument and return

$$\text{Amult}(\mathbf{z}) = A\mathbf{z} \in \mathbb{R}^k, \quad \text{Lmult}(\mathbf{z}) = L\mathbf{z} \in \mathbb{R}^{k'},$$

respectively. Further, since the transposes $A^T : \mathbb{R}^k \rightarrow \mathbb{R}^n$ and $L^T : \mathbb{R}^{k'} \rightarrow \mathbb{R}^n$ appear in (5.10) as well, we need computational routines called **ATmult** and **LTmult** that take vectors $\mathbf{v} \in \mathbb{R}^k$ and $\mathbf{w} \in \mathbb{R}^{k'}$ as arguments and return

$$\text{ATmult}(\mathbf{v}) = A^T\mathbf{v} \in \mathbb{R}^n, \quad \text{LTmult}(\mathbf{w}) = L^T\mathbf{w} \in \mathbb{R}^n.$$

Now we can solve the linear equation $(A^T A + \alpha L^T L)\mathbf{f} = A^T \mathbf{m}$ without actually constructing any of the matrices A, A^T, L or L^T ! The trick is to use an iterative solution strategy, such as the conjugate gradient method.

We will demonstrate large-scale inversion in Chapter 9 in the context of X-ray tomography with realistic resolution.

5.5.1 Conjugate direction methods

The conjugate gradient method belongs to a class of methods known as conjugate direction methods. These were invented for the quadratic optimization problem

$$\text{minimize } \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T \mathbf{x}, \quad (5.23)$$

where Q is an $n \times n$ symmetric positive definite matrix.

Definition 5.5.1 Given a symmetric matrix Q , then two vectors \mathbf{d}_1 and \mathbf{d}_2 are said to be Q -orthogonal, or conjugate with respect to Q , if $\mathbf{d}_1^T Q \mathbf{d}_2 = 0$. A finite set of vectors is said to be a Q -orthogonal set if $\mathbf{d}_i^T Q \mathbf{d}_j = 0$ for all $i \neq j$.

Theorem 5.5.1 If Q is positive definite and the set of nonzero vectors $\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k$ are Q -orthogonal, then these vectors are linearly independent.

Proof: Suppose there are constants α_i , $i = 0, 1, \dots, k$ such that $\alpha_0 \mathbf{d}_0 + \dots + \alpha_k \mathbf{d}_k = 0$. Then, multiplying by Q and \mathbf{d}_i^T ,

$$\alpha_0 \mathbf{d}_0^T Q \mathbf{d}_0 + \dots + \alpha_i \mathbf{d}_i^T Q \mathbf{d}_i + \dots + \alpha_k \mathbf{d}_k^T Q \mathbf{d}_k = 0.$$

By Q -orthogonality, all terms are 0 except the i th, so we must also have $\alpha_i \mathbf{d}_i^T Q \mathbf{d}_i = 0$. But this means $\alpha_i = 0$ since Q is positive definite. Since i was arbitrary, we can then conclude that $\alpha_i = 0$ for $i = 0, 1, \dots, k$. \square

Let us consider why Q -orthogonality is so useful. Corresponding to the $n \times n$ matrix Q , let $\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{n-1}$ be n nonzero Q -orthogonal vectors. Then they are linearly independent, so the solution \mathbf{x}^* of (5.23) can be written

$$\mathbf{x}^* = \sum_{i=0}^{n-1} \alpha_i \mathbf{d}_i.$$

Then

$$Q\mathbf{x}^* = \sum_{i=0}^{n-1} \alpha_i Q\mathbf{d}_i$$

and

$$\mathbf{d}_j^T Q \mathbf{x}^* = \sum_{i=0}^{n-1} \alpha_i \mathbf{d}_j^T Q \mathbf{d}_i = \alpha_j \mathbf{d}_j^T Q \mathbf{d}_j.$$

Recall that the solution to (5.23) satisfies $Q\mathbf{x}^* = \mathbf{b}$. Solving for α_j ,

$$\alpha_j = \frac{\mathbf{d}_j^T Q \mathbf{x}^*}{\mathbf{d}_j^T Q \mathbf{d}_j} = \frac{\mathbf{d}_j^T \mathbf{b}}{\mathbf{d}_j^T Q \mathbf{d}_j}.$$

So

$$\mathbf{x}^* = \sum_{i=0}^{n-1} \frac{\mathbf{d}_i^T \mathbf{b}}{\mathbf{d}_i^T Q \mathbf{d}_i} \mathbf{d}_i.$$

The Q -orthogonal set allowed us to find an expression for \mathbf{x}^* only in terms of Q , \mathbf{b} and the \mathbf{d}_i 's. The expansion for \mathbf{x}^* can be viewed as the result of an iterative process of n steps where at the i th step $\alpha_i \mathbf{d}_i$ is added. This gives us the basic conjugate direction method:

Theorem 5.5.2 Conjugate Direction Theorem *Let $\{\mathbf{d}_i\}_{i=0}^{n-1}$ be a set of nonzero Q -orthogonal vectors. For any $\mathbf{x}_0 \in \mathbb{R}^n$ the sequence $\{\mathbf{x}_k\}$ generated by*

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \quad k \geq 0 \quad \text{with} \tag{5.24}$$

$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \quad \text{and} \quad \mathbf{g}_k = Q\mathbf{x}_k - \mathbf{b} \tag{5.25}$$

converges to the unique solution \mathbf{x}^* of $Q\mathbf{x} = \mathbf{b}$ after n steps. That is, $\mathbf{x}_n = \mathbf{x}^*$.

Proof: Since the \mathbf{d}_i 's are linearly independent, we can write $\mathbf{x}^* - \mathbf{x}_0 = \alpha_0 \mathbf{d}_0 + \alpha_1 \mathbf{d}_1 + \cdots + \alpha_{n-1} \mathbf{d}_{n-1}$, multiply by Q and then \mathbf{d}_k^T to get

$$\alpha_k = \frac{\mathbf{d}_k^T Q(\mathbf{x}^* - \mathbf{x}_0)}{\mathbf{d}_k^T Q \mathbf{d}_k}. \quad (5.26)$$

Following the iterative process (5.24):

$$\begin{aligned}\mathbf{x}_1 &= \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 \\ \mathbf{x}_2 &= \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 + \alpha_1 \mathbf{d}_1 \\ \mathbf{x}_k &= \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 + \cdots + \alpha_{k-1} \mathbf{d}_{k-1}\end{aligned}$$

we see that

$$\mathbf{d}_k^T Q(\mathbf{x}_k - \mathbf{x}_0) = \sum_{i=0}^{k-1} \alpha_i \mathbf{d}_k^T Q \mathbf{d}_i.$$

which implies, by the Q -orthogonality of the \mathbf{d}_i 's that $\mathbf{d}_k^T Q \mathbf{x}_k = \mathbf{d}_k^T Q \mathbf{x}_0$. Substituting this into (5.26) gives

$$\begin{aligned}\alpha_k &= \frac{\mathbf{d}_k^T Q \mathbf{x}^* - \mathbf{d}_k^T Q \mathbf{x}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} = \frac{\mathbf{d}_k^T \mathbf{b} - \mathbf{d}_k^T Q \mathbf{x}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \\ &= \frac{-\mathbf{d}_k^T (Q \mathbf{x}_k - \mathbf{b})}{\mathbf{d}_k^T Q \mathbf{d}_k} = \frac{\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k}\end{aligned}$$

□

5.5.2 Conjugate gradient

In the conjugate gradient method the directions are not specified beforehand, but rather are determined at each step in the iteration. At step k , one evaluates the current negative gradient vector and adds to it a linear combination of the previous direction vectors to obtain a new conjugate direction vector along which to move. Advantages of the conjugate gradient method include:

1. Unless the solution is attained in less than n steps, the gradient is always nonzero and linearly independent of all previous direction vectors. In fact, the gradient \mathbf{g}_k is orthogonal to the subspace generated by $\mathbf{d}_0, \dots, \mathbf{d}_{k-1}$.

2. The conjugate gradient method has a simple formula for determining the new direction vectors.
3. Since the directions are based on gradients, the method makes good progress toward the solution at every step. Note that in arbitrary sequences of conjugate directions, progress may be poor until the final few steps. While this is not significant for quadratic problems, it is for the generalization to non-quadratic problems.

Note that with roundoff error, the conjugate gradient method does not terminate in a finite number of steps.

The Conjugate Gradient Algorithm

For any $\mathbf{x}_0 \in \mathbb{R}^n$, define $\mathbf{d}_0 = -\mathbf{g}_0 = \mathbf{b} - Q\mathbf{x}_0$ and

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{d}_k \\ \alpha_k &= \frac{-\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \\ \mathbf{d}_{k+1} &= -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k \\ \beta_k &= \frac{-\mathbf{g}_{k+1}^T Q \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k}\end{aligned}$$

where $\mathbf{g}_k = Q\mathbf{x}_k - \mathbf{b}$. Note that the first step is a Steepest Descent step.

5.5.3 Preconditioning

Consider the unregularized minimization problem

$$\text{minimize} \|A\mathbf{f} - \mathbf{m}\|_2^2 \quad (5.27)$$

or the regularized problem

$$\text{minimize} \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \|\mathbf{f}\|_2^2. \quad (5.28)$$

Each is a quadratic minimization problem with a unique minimum. We showed that the Tikhonov regularized solution $T_\alpha(\mathbf{m})$ to (5.28) satisfies $T_\alpha(\mathbf{m}) = (A^T A + \alpha I)^{-1} A^T \mathbf{m}$ or the *normal equations*

$$(A^T A + \alpha I) T_\alpha(\mathbf{m}) = A^T \mathbf{m}.$$

Preconditioning is a technique to speed up potentially slow convergence of the minimization problem solved by an iterative method. The preconditioner

is a matrix S chosen so that AS^{-1} has a more favorable spectrum than A and in place of the unregularized minimization problem (5.27), we instead minimize

$$\text{minimize} \| (AS^{-1})\mathbf{y} - \mathbf{m} \|_2^2, \quad S\mathbf{f} = \mathbf{y}. \quad (5.29)$$

This is an equivalent problem since

$$AS^{-1}\mathbf{y} - \mathbf{m} = AS^{-1}S\mathbf{f} - \mathbf{m} = A\mathbf{f} - \mathbf{m}.$$

Thus, we have the normal equations

$$(S^{-1})^T A^T AS^{-1}\mathbf{y} = (S^{-1})^T A\mathbf{m}.$$

The desired (and partially contradictory) properties of S are

- AS^{-1} should be better conditioned than A and/or have only a few distinct singular values
- S should have about the same number of nonzero entries as A
- It should be cheap to solve equations with S and S^T

Note that there is an extra cost associated with preconditioning and that is solving $S\mathbf{f} = \mathbf{y}$. One example of a cheap and simple preconditioner is a diagonal scaling of the columns of A :

$$S = \text{diag}(s_1^{1/2}, \dots, s_n^{1/2}), \quad s_j = \|a_j\|_2^2.$$

For further study on conjugate direction methods, the conjugate gradient method, and preconditioning, we recommend going to a text on optimization methods. While such references abound, we referred to [384, 118, 262, 323] in writing these sections.

Exercise 5.5.1 Explain how the conjugate gradient method can be used to solve the linear equation $(A^T A + \alpha L^T L)\mathbf{f} = A^T \mathbf{m}$ without actually constructing any of the matrices A, A^T, L or L^T .

Exercise 5.5.2 Implement the conjugate gradient method can be used to solve the linear equation $(A^T A + \alpha L^T L)\mathbf{f} = A^T \mathbf{m}$ for the deconvolution problem.

Exercise 5.5.3 Implement the conjugate gradient method can be used to solve the linear equation $(A^T A + \alpha L^T L)\mathbf{f} = A^T \mathbf{m}$ for the backward heat equation.

Chapter 6

Total variation regularization

Many applications of computational inversion methods call for reconstructions with sharp features. For example, natural scenes contain edges between areas of different colors. Therefore, a method recovering ideal sharp photographs from misfocused noisy snapshots should be *edge-preserving*.

Using truncated SVD or Tikhonov regularization, especially with derivative penalty, generally results in smooth reconstructions. An alternative method was introduced [370, 398] with edge-preservation in mind. The basic idea is to replace the 2-norm by 1-norm in the penalty term of generalized Tikhonov regularization: find a vector $\mathbf{f} \in \mathbb{R}^n$ that minimizes the expression

$$\|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \sum_{j=1}^n |(L\mathbf{f})_j|, \quad (6.1)$$

where the finite difference matrix L is given by (5.16). This resulted in a rich literature of fruitful applications of this so-called *total variation regularization*. In statistical literature, this type of methods go by the name *lasso*, see [436].

In this book we concentrate on two computational approaches for minimizing the expression (6.1): medium-scale constrained quadratic programming [303, 209, 162, 282] and large-scale gradient-based minimization methods for an approximate version of (6.1) with smoothed absolute-value function similarly to [288, 103, 236, 201]. These methods have the advantage that non-negativity constraints can be easily implemented.

Many other computational methods for minimizing the expression (6.4) have been introduced as well, including a lagged diffusivity method [125], Lagrange multiplier methods [72], frame-based thresholding methods [111,

98, 76], domain decomposition methods [146, 145], Bregman distance methods [367, 484, 163, 60, 369, 493], primal-dual methods [74, 73, 134, 354], finite element methods [141, 201], and other methods [311, 126, 462, 457, 314, 182, 469, 183, 71]. Further treatments of total variation regularization and related methods can be found in the books [461, 368, 75, 404, 194].

6.1 What is total variation?

The total variation of a function in one dimension on an interval $[a, b]$ is defined by [397] as follows.

Definition 6.1.1 *Let f be a real-valued function defined on the interval $[a, b]$. The total variation of f , denoted by $TV(f)$, is defined to be*

$$TV(f) = \sup \sum_{i=1}^k |f(x_i) - f(x_{i-1})|$$

where the supremum is over all partitions $a = x_0 < x_1 < \dots < x_k = b$ of $[a, b]$.

Note that if f is piecewise constant with a finite number of jump discontinuities, then the total variation is the sum of the magnitude of the jumps. If f is differentiable, letting $\Delta x_i = x_i - x_{i-1}$,

$$TV(f) = \sup \sum_{i=1}^k \frac{|f(x_i) - f(x_{i-1})|}{|x_i - x_{i-1}|} |x_i - x_{i-1}|,$$

and taking the limit as $\Delta x_i \rightarrow 0$ results in

$$TV(f) = \int_a^b |f'(x)| dx. \quad (6.2)$$

In higher dimensions this generalizes to ($\Omega = [a, b]^n$)

$$TV(f) = \int_{\Omega} |\nabla f(x)| dx. \quad (6.3)$$

By constraining our least-squares minimization problem to minimize the total variation of the solution (in the discrete sense), we expect the solution to have fewer oscillations. We also expect it to be “blockier” since we are not minimizing the function amplitude or derivative directly.

One way to generalize Tikhonov regularization for the indirect measurement $\mathbf{m} = A\mathbf{f} + \varepsilon$ is to consider the minimization problem

$$T_\alpha(\mathbf{m}) = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \|A\mathbf{z} - \mathbf{m}\|_2^2 + \alpha \|L\mathbf{z}\|_p^p \right\}, \quad (6.4)$$

where the finite difference matrix L is given by (5.16). The parameter $1 \leq p < \infty$ in (6.4) is related to the so-called ℓ^p -norm defined for vectors $\mathbf{h} \in \mathbb{R}^n$ by

$$\|\mathbf{h}\|_p := \left(\sum_{j=1}^n |\mathbf{h}_j|^p \right)^{1/p}. \quad (6.5)$$

Note that taking $p = 2$ in formula (6.4) leads to generalized Tikhonov regularization as discussed in Section 5.3.

An important special case is $p = 1$; then (6.5) simplifies to the form

$$\|\mathbf{h}\|_1 := |\mathbf{h}_1| + \cdots + |\mathbf{h}_n|. \quad (6.6)$$

Inversion based on formula (6.4) with $p = 1$ is referred to as *total variation regularization*.

Total variation regularization can be understood as a balance between two requirements:

- (i) $T_\alpha(\mathbf{m})$ should give a small residual $AT_\alpha(\mathbf{m}) - \mathbf{m}$,
- (ii) $L T_\alpha(\mathbf{m})$ should be small in ℓ^1 norm.

Taking $p = 2$ in formula (6.4) strongly favors smooth reconstructions over discontinuous ones, whereas $p = 1$ allows piecewise constant reconstructions.

Let us give an intuitive reason why taking $p = 1$ in formula (6.4) is so different from taking $p = 2$. We define two functions in the interval $0 \leq x \leq 1$. The first one is the linear function $f(x) = ax$ with $a = 20/9$, and the second one is the discontinuous function

$$h(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq 1/2, \\ 2 & \text{for } 1/2 < x \leq 1. \end{cases}$$

We choose $n = 10$ and discretize the functions using the points

$$x_1 = 0, \quad x_2 = \frac{1}{10}, \quad \dots, \quad x_{10} = \frac{9}{10},$$

and setting $\mathbf{f} = [f(x_1), f(x_2), \dots, f(x_{10})]^T$ and $\mathbf{h} = [h(x_1), h(x_2), \dots, h(x_{10})]^T$. Furthermore, denote $\Delta x := x_2 - x_1 = 1/10$.

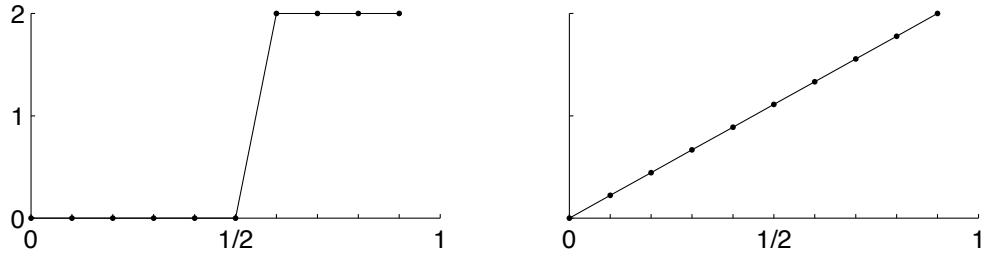


Figure 6.1: Two functions with interesting differences in their 1- and 2-norms. Left: \mathbf{h} . Right: \mathbf{f} .

Note that both functions climb up from zero to two in the same interval ($f(0) = 0 = h(0)$ and $f(9/10) = 2 = h(9/10)$), but f does so smoothly and h with a single jump discontinuity.

Let L_0 be the 9×10 matrix achieved by removing the last row in matrix L given by formula (5.16). Then

$$L_0 \mathbf{h} = 10 \begin{bmatrix} \mathbf{h}_2 - \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{10} - \mathbf{h}_9 \end{bmatrix},$$

and similarly for $L_0 \mathbf{f}$.

A simple calculation shows that

$$\|L_0 \mathbf{f}\|_2^2 = 44.44 \dots, \quad \|L_0 \mathbf{f}\|_1 = 20,$$

and

$$\|L_0 \mathbf{h}\|_2^2 = 400, \quad \|L_0 \mathbf{h}\|_1 = 20.$$

The point is that the ℓ^2 -norm penalizes the jump discontinuity much more than a smooth change, whereas the ℓ^1 -norm gives exactly the same penalty for both.

Exercise 6.1.1 Consider the function $f(x) = \sin 2\pi x$ on the interval $0 \leq x \leq 1$ and define $h(x)$ to be the “square wave”

$$h(x) = \begin{cases} 0 & \text{for } x = 0, \\ 1 & \text{for } 0 < x < 0.5, \\ 0 & \text{for } x = 0.5, \\ -1 & \text{for } 0.5 < x < 1 \\ 0 & \text{for } x = 1. \end{cases}$$

Choose $n = 11$ and discretize the functions using the points

$$x_1 = 0, x_2 = \frac{1}{10}, \dots, x_{10} = \frac{9}{10}, x_{11} = 1.$$

Using this discretization, compute $\|L_0\mathbf{f}\|_2^2, \|L_0\mathbf{f}\|_1, \|L_0\mathbf{h}\|_2^2, \|L_0\mathbf{h}\|_1$. How does this compare to the results of the example of this section in terms of the net change of the function? Does the result change if you use 21 discretization points?

6.2 Quadratic programming

We want to find a vector $\mathbf{f} \in \mathbb{R}^n$ that solves (6.4) with $p = 1$. We write the vector $L\mathbf{f} \in \mathbb{R}^n$ in the form

$$\mathbf{v}_+ - \mathbf{v}_- = L\mathbf{f},$$

where \mathbf{v}_\pm are nonnegative vectors: $\mathbf{v}_\pm \in \mathbb{R}_+^n$, or $(\mathbf{v}_\pm)_j \geq 0$ for all $j = 1, \dots, n$. Now minimizing (6.4) with $p = 1$ is equivalent to minimizing

$$\|A\mathbf{f}\|_2^2 - 2\mathbf{m}^T A\mathbf{f} + \alpha\mathbf{1}^T \mathbf{v}_+ + \alpha\mathbf{1}^T \mathbf{v}_-,$$

where $\mathbf{1}$ is the vector with all elements equal to one: $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T \in \mathbb{R}^n$, and the minimization is taken over $y \in \mathbb{R}^{3n}$ defined by

$$\mathbf{y} = \begin{bmatrix} \mathbf{f} \\ \mathbf{v}_+ \\ \mathbf{v}_- \end{bmatrix}, \quad \text{where} \quad \mathbf{v}_+ \in \mathbb{R}_+^n, \quad \mathbf{v}_- \in \mathbb{R}_+^n.$$

Note the identity $\|A\mathbf{f}\|_2^2 = \mathbf{f}^T A^T A \mathbf{f}$ and write

$$H = \begin{bmatrix} 2A^T A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} -2A^T \mathbf{m} \\ \alpha\mathbf{1} \\ \alpha\mathbf{1} \end{bmatrix}.$$

We then have the quadratic optimization problem in standard form

$$\arg \min_{\mathbf{y}} \left\{ \frac{1}{2} \mathbf{y}^T H \mathbf{y} + \mathbf{h}^T \mathbf{y} \right\} \quad (6.7)$$

with the constraints

$$L \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{2n} \end{bmatrix} - \begin{bmatrix} y_{2n+1} \\ \vdots \\ y_{3n} \end{bmatrix} \quad (6.8)$$

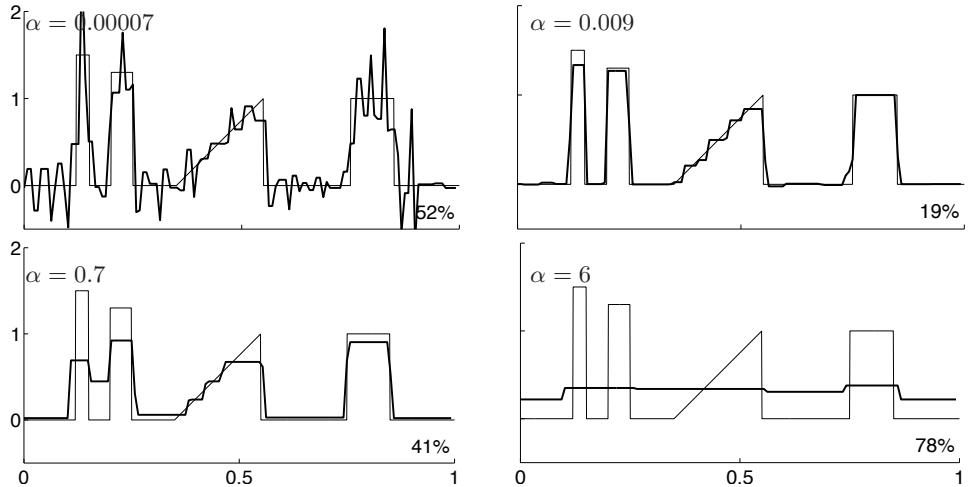


Figure 6.2: Total variation regularized reconstructions. The percentages shown are relative errors of reconstructions. Note the staircasing effect in the linear ramp part of the signal; this is a typical artefact of total variation inversion. Here $n = 128$.

and

$$y_j \geq 0 \text{ for } j = n + 1, \dots, 3n. \quad (6.9)$$

Several software packages (such as `quadprog.m` routine in MATLAB's Optimization Toolbox) exist that can deal with a problem of the form (6.7) with constraints of type (6.8).

One downside of the above approach is that the optimization problem (6.7) has $3n$ degrees of freedom, whereas the original problem (6.4) has only n . Numerical optimization becomes harder in higher dimensions. However, the advantage is that (6.7) is in a well-understood standard form.

Let us apply the above method to our one-dimensional deconvolution test problem. See Figures 6.2 and 6.3. In the figures the blockiness, or edge-preserving nature, of the reconstructions is evident. One also sees that too large a choice of α damps out the amplitude of the reconstructions, while too small of a choice results in a highly oscillatory reconstruction. The challenge remains to find just the right choice. The noise-robustness of the reconstruction is investigated in Figure 6.3. The robustness diminishes with the size of α .

The two-dimensional case is slightly more complicated since we need to

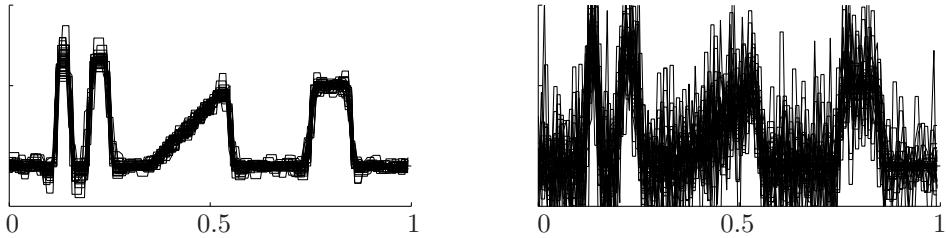


Figure 6.3: Study of noise-robustness of total variation regularized reconstructions. Here the noise level is $\sigma = 0.05 \cdot \max |f(x)|$, and 50 independent realizations of random white noise was used for both of the two plots. Left: regularization parameter $\alpha = 0.02$. Right: regularization parameter $\alpha = 0.005$.

discretize the gradient of the unknown with respect to two directions. One possibility is to write horizontal and vertical difference quotients in the form of two matrices L_H and L_V and look for a vector $\mathbf{f} \in \mathbb{R}^n$ that minimizes the expression

$$\|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \sum_{j=1}^n |(L_H \mathbf{f})_j| + \alpha \sum_{j=1}^n |(L_V \mathbf{f})_j|. \quad (6.10)$$

We remark that modifying the formulation (6.10) as above leads to an optimization problem with $5n$ degrees of freedom.

Exercise 6.2.1 *Experiment with total variation regularization on the deconvolution problem with kernel defined in Exercise 4.4.2. Consider noise levels with standard deviation $\sigma = 0.01 \cdot \max |f(x)|$, $\sigma = 0.05 \cdot \max |f(x)|$, and $\sigma = 0.1 \cdot \max |f(x)|$. Try to find an optimal regularization parameter.*

Exercise 6.2.2 *Experiment with total variation regularization using the kernel*

$$f(x) = \begin{cases} 1 & \text{for } 0.1 \leq x \leq 0.2, \\ -2.5x + 1.55 & \text{for } 0.3 \leq x \leq 0.5, \\ 3x - 1.2 & \text{for } 0.5 < x < 0.6, \\ 1.3 & \text{for } 0.8 \leq x \leq 0.9, \\ 0 & \text{otherwise} \end{cases}$$

for the deconvolution problem, and reproduce plots as in Figure 6.2.

6.3 Sparsity-based parameter choice

Total variation regularization can be seen as an example of the more general concept of *compressed sensing*, whose study started with the articles [68,

[127]. Compressed sensing is an efficient framework for recovering sparse signals from noisy and corrupted data. The term “sparse” means here that a continuous signal can be represented by a finite number of coefficients in a suitable basis, for example using Fourier or wavelet transform coefficients.

A new sparsity-based method for choosing the regularization parameter was introduced in [282]. Suppose that we know *a priori* the finite number of nonzero coefficients in the true signal. We then compute reconstructions from the noisy data using a large collection of regularization parameters (similarly to the L-curve method of Section 5.4.2), count the number of nonzero transform coefficients of each reconstruction, and choose the parameter so that the corresponding reconstruction has approximately the correct number of nonzero transform coefficients. This parameter choice rule is called *the S-curve method*.

Why would the above method work in general? The thinking goes as follows. When the regularization parameter tends to infinity, the reconstruction converges typically to a constant, often zero. Thus there are no nonzero coefficients (or at most a small and known maximum number of them) in the limit case. On the other hand, when the parameter tends to zero, there is no regularization and the reconstruction is typically very erratic and oscillatory. Most of the transform coefficients are then needed to represent the reconstruction. Thus there is a more or less continuous transition from most coefficients nonzero (small regularization parameter) to no nonzero coefficients (large regularization parameter). Roughly speaking, somewhere in between those extremes there is a parameter value yielding approximately the desired level of sparsity.

Before turning to a numerical example, let us comment on the possibility of having the above kind of *a priori* information. How would we know the number of nonzero coefficients in the unknown target? There are at least two typical sources of such information. First, we may know some technical properties of the true signal, as in the recovery of corrupted dual-tone multi-frequency signals used for telecommunications over analog telephone lines. There one knows that the clean signal consists of exactly two pure frequencies. Second, we may have available a collection of typical signals: for example in medical imaging one can use a set of full-data CT slice images for measuring typical sparsity values and then use this knowledge as prior knowledge in limited-data tomography.

Let us now see how the sparsity-based choice rule works in practice. Let us compute the same example as in Figure 6.2 with $n = 128$. First of all, we compute the number of nonzero coefficients in our actual signal. Let the matrix L be given by formula (5.16), and compute $L\mathbf{f}$ with \mathbf{f} being the

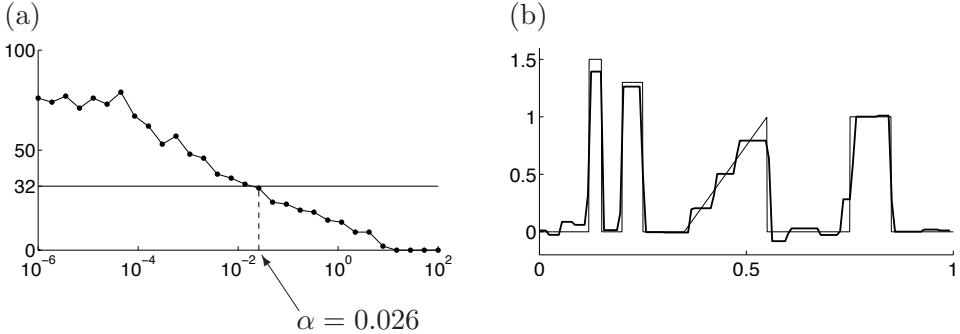


Figure 6.4: Sparsity-based choice of regularization parameter for total variation regularization. Here $n = 128$. (a) Number of jumps in the reconstruction as function of regularization parameter α . Note the logarithmic scale in the horizontal α -axis. (b) Reconstruction corresponding to the choice $\alpha = 0.026$ (thick line) and original signal (thin line).

true signal evaluated at the points x_j defined in (2.6). It turns out that the difference vector $L\mathbf{f}$ has exactly 32 nonzero entries. In other words, the signal \mathbf{f} has 32 jumps.

Next we choose 30 parameters $\alpha_1, \dots, \alpha_{30}$ in the interval $[10^{-6}, 100]$ with logarithmic spacing. We solve the minimization problem (6.4) for each α_j using quadratic programming as explained in Section 6.2. Figure 6.4(a) shows the numbers of jumps in each reconstruction. We want to choose a value of α that yields approximately 32 jumps in the reconstruction. However, the curve shown in Figure 6.4(a) is not monotonically decreasing, so there is in principle a possibility of non-uniqueness in the choice. We use the following simple idea: let α_{j_0} be the largest parameter used in the computation that yields *more than* 32 jumps in the reconstruction. Choose $\alpha := \alpha_{j_0+1}$. See Figure 6.4(b) for the resulting reconstruction.

Section 9.1.3 below presents a further example of the sparsity-based choice rule in the context of X-ray tomography and wavelet-based regularization.

6.4 Large-scale implementation

We apply the gradient-descent minimization method of Barzilai and Borwein [31] to (approximate) large-scale total variation regularization. Our aim is

to minimize

$$\begin{aligned} G(\mathbf{f}) &= \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha\|\mathbf{L}\mathbf{f}\|_1 \\ &= \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \sum_{j=1}^n |\mathbf{f}_j - \mathbf{f}_{j-1}|, \end{aligned} \quad (6.11)$$

where we use the convention $f_0 = f_n$ according to the periodic boundary condition. Now the *objective functional* $G : \mathbb{R}^n \rightarrow \mathbb{R}$ is not continuously differentiable because of the absolute values appearing in (6.11), so we cannot apply any derivative-based optimization method.

Let us replace the absolute value function $|t|$ by an approximation:

$$|t|_\beta := \sqrt{t^2 + \beta},$$

where $\beta > 0$ is small. (Another suitable choice is $|t|_\beta = \frac{1}{\beta} \log(\cosh(\beta t))$.) Then the modified objective functional

$$G_\beta(\mathbf{f}) = \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \sum_{j=1}^n |\mathbf{f}_j - \mathbf{f}_{j-1}|_\beta \quad (6.12)$$

is continuously differentiable and we can apply gradient-based optimization methods.

We need to compute the gradient of G_β . Write

$$\nabla G_\beta(\mathbf{f}) = \nabla \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \nabla \left(\sum_{j=1}^n |\mathbf{f}_j - \mathbf{f}_{j-1}|_\beta \right)$$

and calculate for the first term

$$\nabla \|A\mathbf{f} - \mathbf{m}\|_2^2 = 2A^T A\mathbf{f} - 2A^T \mathbf{m}. \quad (6.13)$$

Computing the gradient of the second term involves keeping track of the following kind of calculations:

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{f}_\nu} \left(\sum_{j=1}^n |\mathbf{f}_j - \mathbf{f}_{j-1}|_\beta \right) \\ &= \sum_{j=1}^n \frac{\partial}{\partial \mathbf{f}_\nu} ((\mathbf{f}_j - \mathbf{f}_{j-1})^2 + \beta)^{1/2} \\ &= \frac{\mathbf{f}_\nu - \mathbf{f}_{\nu-1}}{((\mathbf{f}_\nu - \mathbf{f}_{\nu-1})^2 + \beta)^{1/2}} - \frac{\mathbf{f}_{\nu+1} - \mathbf{f}_\nu}{((\mathbf{f}_{\nu+1} - \mathbf{f}_\nu)^2 + \beta)^{1/2}}. \end{aligned} \quad (6.14)$$

The details are left as an exercise.

The *steepest descent method* was introduced by Cauchy in 1847. It is an iterative method where the initial guess $\mathbf{f}^{(1)}$ is chosen some way (e.g. $\mathbf{f}^{(1)} = 0$) and the next iterates are found inductively by

$$\mathbf{f}^{(\ell+1)} = \mathbf{f}^{(\ell)} - \kappa_\ell \nabla G_\beta(\mathbf{f}^{(\ell)}),$$

where the steplength parameter κ_ℓ is determined from

$$\kappa_\ell = \arg \min_{\kappa \geq 0} G_\beta(\mathbf{f}^{(\ell)} - \kappa \nabla G_\beta(\mathbf{f}^{(\ell)})).$$

Note that κ_ℓ is called a *steplength parameter* instead of just *steplength* because the length of the gradient is not necessarily one in general; thus the actual length of the step is $|\kappa_\ell \nabla G_\beta(\mathbf{f}^{(\ell)})|$. The steepest descent method is known to converge very slowly.

Barzilai and Borwein introduced in 1988 [31] the following optimization strategy which differs from the steepest descent method only by the choice of the steplength parameter:

$$\mathbf{f}^{(\ell+1)} = \mathbf{f}^{(\ell)} - \delta_\ell \nabla G_\beta(\mathbf{f}^{(\ell)}),$$

where δ_ℓ is given by setting $y_\ell := \mathbf{f}^{(\ell)} - \mathbf{f}^{(\ell-1)}$ and $g_\ell := \nabla G_\beta(\mathbf{f}^{(\ell)}) - \nabla G_\beta(\mathbf{f}^{(\ell-1)})$ and

$$\delta_\ell = \frac{y_\ell^T y_\ell}{y_\ell^T g_\ell}.$$

This method converges faster and is less affected by ill-conditioning than the steepest descent method. (Especially for quadratic objective functionals) There are some practical problems with the method of Barzilai and Borwein:

- (i) How to choose $\beta > 0$? Too large value leads to smooth reconstructions, and too small value may cause convergence problems.
- (ii) How to choose the first steplength δ_1 ?
- (iii) The objective function is not guaranteed to get smaller with each step. What to do in the case it becomes bigger?

Usually (i) can be solved by some experimentation. The quick-and-dirty solution to (ii) is just choosing δ_1 to be small, for example $\delta_1 = \frac{1}{10\,000}$. Another practical way to choose δ_1 by line minimization.

One simple way to deal with (iii) is to check if f increases, and if so, half the steplength. However, this is not the best possible way to ensure the

convergence of the method, since just the increasing steps have turned out to be essential for the local convergence properties of the Barzilai-Borwein method. It is often advisable to just let the method run in spite of occasionally increasing objective function values.

Strategies to guarantee the global convergence of the Barzilai-Borwein method can be found, for instance, in [392, 104]. Constrained optimization, such as enforcing nonnegativity, using Barzilai-Borwein method is discussed in [104, 467].

Note that the storage need of the Barzilai-Borwein method is of the order n instead of n^2 typical for many other methods. If \mathbf{f} is a large $M \times N$ size image, then $n^2 = M^2N^2$ is too large for most computer memories!

Exercise 6.4.1 Use formulas (6.13) and (6.14) to determine the gradient of the objective functional G_β defined in (6.12).

Exercise 6.4.2 Plot $|t|_\beta = \frac{1}{\beta} \log(\cosh(\beta t))$ for various small values of $\beta > 0$ and explain why this is a suitable choice to approximate the absolute value function $|t|$.

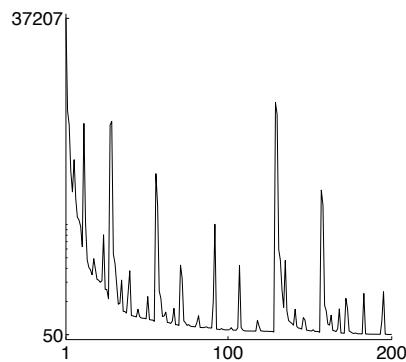
6.4.1 TV regularization for tomography

Consider the tomographic measurement matrix A constructed in Section 2.3.5 for the resolution 50×50 and with 50 projection directions. We simulate data avoiding inverse crime as explained in Section 2.3.6.

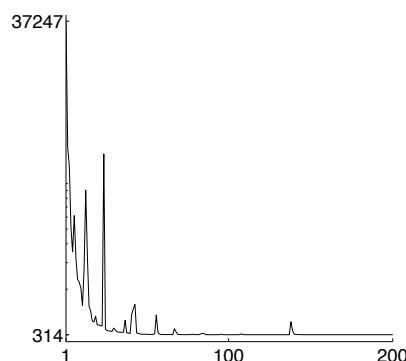
Next we apply the approximate large-scale total variation algorithm described in Section 6.4. We take $\beta = 0.0001$ and $\delta_1 = 0.0001$ and let the Barzilai-Borwein method run for 200 iteration steps. The initial guess is the all-zero image. See Figure 6.5 for the results using various regularization parameters.

Chapter 9 contains further tomographic examples based on total variation regularization, including reconstructions from measured X-ray data.

Too small $\alpha = 10^{-3}$, error 43%



Medium $\alpha = 0.8$, error 39%



Too large $\alpha = 20$, error 65%

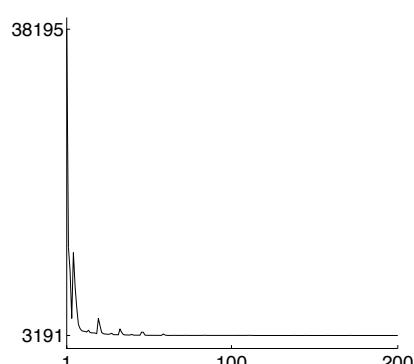
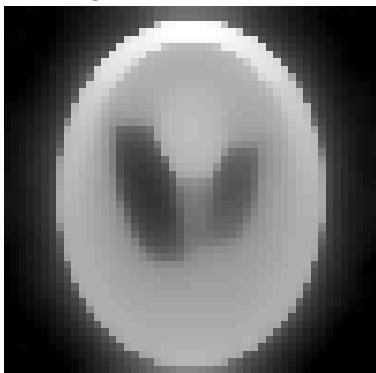


Figure 6.5: Left column: reconstructions of the 50×50 Shepp-Logan phantom using approximate total variation regularization and Barzilai-Borwein minimization with non-negativity constraint. Note the edge-preserving nature of the total variation penalty. Right column: Evolution of the objective functional during the minimization process. Note that the Barzilai-Borwein minimization does not decrease the value of the objective function at every iteration.

Chapter 7

Besov space regularization using wavelets

The Besov space norm provides another means of promoting solutions with sharp edges and introducing sparsity in the penalty function. It has great potential for capturing fine details in the reconstruction due to the *multiresolution* properties of wavelets. The reader is referred to texts such as [2, 85, 110, 247, 261, 333, 441] for a more thorough introduction to wavelets than what we provide here. Our introduction is mainly structured as in [110]. Examples using Besov space regularization for tomographic data are found in Section 9.1.3. For experimentation with wavelets, the reader may wish to try MATLAB's Wavelet Toolbox.

7.1 An introduction to wavelets

The first example of wavelets we will consider is the Haar wavelet basis. This basis for $L^2(\mathbb{R})$ has been known since 1910 and provides an example that is both intuitive and easy to compute. Define the Haar function

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \leq x < 1/2, \\ -1 & \text{for } 1/2 \leq x < 1, \\ 0 & \text{otherwise,} \end{cases} \quad (7.1)$$

and $\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n)$. The set $\{\psi_{m,n}\}$ is an orthonormal basis for $L^2(\mathbb{R})$. One proof of this result is found in [110], where it is shown that the $\psi_{m,n}$ are orthonormal and any L^2 function f can be approximated to arbitrarily small precision by a finite linear combination of the $\psi_{m,n}$.

A key concept to wavelet bases is the idea of multiresolution analysis (MRA).

Definition 7.1.1 A multiresolution analysis is a sequence of closed subspaces V_j that satisfy the following five properties.

$$1. \cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots$$

$$2. \overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}) \quad \text{and} \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\}$$

$$3. f \in V_j \iff f(2^j \cdot) \in V_0$$

$$4. f \in V_0 \implies f(\cdot - n) \in V_0 \text{ for all } n \in \mathbb{Z}$$

$$5. \text{there exists a function } \phi \in V_0 \text{ such that } \{\phi_{0,n} : n \in \mathbb{Z}\} \text{ where } \phi_{j,n} = 2^{-j/2} \phi(2^{-j}x - n) \text{ is an orthonormal basis in } V_0.$$

The function ϕ is called the scaling function of the MRA.

The third property is the one that epitomizes the idea of multiresolution: all of the spaces V_j are scaled versions of V_0 .

Here one choice for the scaling function ϕ is

$$\phi(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

related to the Haar multiresolution analysis with spaces

$$V_j = \{f \in L^2(\mathbb{R}) : f|_{[2^j k, 2^j(k+1)]} = \text{constant for all } k \in \mathbb{Z}\}. \quad (7.2)$$

The usefulness of a MRA is that whenever a collection of closed subspaces satisfies 1-5 of Definition 7.1.1, then there exists an orthonormal wavelet basis $\{\psi_{j,k}\}$ of $L^2(\mathbb{R})$ such that for all $f \in L^2(\mathbb{R})$,

$$P_{j-1}f = P_j f + \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}$$

where $P_j f$ is the orthogonal projection of f onto V_j . Moreover, the wavelet can be constructed explicitly. One possibility for the construction of ψ is to define its Fourier transform by

$$\hat{\psi}(\xi) = e^{i\xi/2} \overline{m_0(\xi/2 + \pi)} \hat{\phi}(\xi/2) \quad (7.3)$$

where

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_n h_n e^{-in\xi}, \quad h_n = \langle \phi, \phi_{-1,n} \rangle, \quad \text{and} \quad \sum_{n \in \mathbb{Z}} |h_n|^2 = 1. \quad (7.4)$$

Equivalently,

$$\psi(x) = \sum_n (-1)^{n-1} \overline{h_{-n-1}} \phi_{-1,n}(x) = \sqrt{2} \sum_n (-1)^{n-1} \overline{h_{-n-1}} \phi(2x - n). \quad (7.5)$$

This ψ is known as the *mother wavelet*. Note that this construction is not unique.

Exercises 7.1.3 and 7.1.4 illustrate that while the Haar wavelet basis is easy to use and has the attribute that the wavelets are compactly supported, it does not have very good time-frequency localization. Other examples of wavelet bases can be found in the references at the beginning of this Chapter. Those constructed by the method described in this section lead to wavelets with infinite support. An important breakthrough in wavelet analysis was the construction of a more general family of compactly supported wavelets [109]. Note that in general they cannot be written in closed form, and so we refer the reader to [110] for their construction, and we will use them in the remainder of this section.

With the deconvolution problem in mind, we construct a wavelet representation for 1-periodic functions on \mathbb{R} following Daubechies [110, section 9.3]. In other words, we will give a construction of a wavelet basis for functions on the one-dimensional torus \mathbb{T} constructed by identifying the end points of the interval $[0, 1]$.

Let ψ^C and ϕ^C be a compactly supported mother wavelet and scaling function, respectively, suitable for multi-resolution analysis in \mathbb{R} . Set

$$\phi_{j,k}(x) = \sum_{\ell \in \mathbb{Z}} \phi^C(2^j(x + \ell) - k), \quad (7.6)$$

$$\psi_{j,k}(x) = \sum_{\ell \in \mathbb{Z}} \psi^C(2^j(x + \ell) - k), \quad (7.7)$$

and define spaces $V_j := \overline{\text{span}\{\phi_{j,k} \mid k \in \mathbb{Z}\}}$ and $W_j := \overline{\text{span}\{\psi_{j,k} \mid k \in \mathbb{Z}\}}$. From [110] the V_j are spaces of constant functions for $j \leq 0$. Thus we have a ladder $V_0 \subset V_1 \subset V_2 \subset \dots$ of multiresolution spaces satisfying $\overline{\cup_{j \geq 0} V_j} = L^2(\mathbb{T})$. Notice that we have flipped notation from [110] in which the ladder goes in the opposite direction and the V_j are constant for negative j . We have replaced j by $-j$ in the notation of [110].

The successive orthogonal complements of V_j in V_{j+1} turn out to be W_j for $j \geq 0$. Each space W_j has the orthonormal basis $\{\psi_{j,k} \mid k = 0, \dots, 2^j - 1\}$, and we can represent functions as follows:

$$f(x) = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} w_{j,k} \psi_{j,k}(x), \quad (7.8)$$

where the coefficients are defined by

$$c_0 := \langle f, 1 \rangle = \int_{\mathbb{T}} f(x) dx, \quad (7.9)$$

$$w_{j,k} := \langle f, \psi_{j,k} \rangle = \int_{\mathbb{T}} f(x) \psi_{j,k}(x) dx. \quad (7.10)$$

Given a vector

$$\mathbf{f} = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_{n-1}]^T \in \mathbb{R}^n,$$

it is quite easy and computationally effective to evaluate approximately the wavelet coefficients c_0 and $w_{j,k}$. Namely, two discrete functions $g = [g_0 \cdots g_L]$ and $h = [h_0 \cdots h_L]$, are related to any fixed pair of mother wavelet $\psi_{0,0}$ and scaling function $\phi_{0,0}$. Denote the discrete periodic convolutions between the two discrete functions and the vector \mathbf{f} by

$$G\mathbf{f} := g * \mathbf{f} \in \mathbb{R}^n \quad \text{and} \quad H\mathbf{f} := h * \mathbf{f} \in \mathbb{R}^n.$$

Furthermore, define a *downsampling operator* $S : \mathbb{R}^n \rightarrow \mathbb{R}^{n/2}$ by

$$S\mathbf{v} = S[v_1 \ v_2 \ v_3 \cdots v_{n-1} \ v_n]^T := [v_1 \ v_3 \cdots v_{n-1}]^T.$$

The wavelet coefficients corresponding to the finest scale basis functions are then given by

$$SH\mathbf{f} = [w_{(N-1),0} \ w_{(N-1),1} \cdots w_{(N-1),(2^{N-1}-1)}]^T.$$

The next coarser level is given by

$$SHSG\mathbf{f} = [w_{(N-2),0} \ w_{(N-2),1} \cdots w_{(N-2),(2^{N-2}-1)}]^T,$$

and finally the process stops by the computation of the scalars $c_0 = (SG)^N \mathbf{f}$ and $w_{0,0} = SH(SG)^{N-1} \mathbf{f}$.

We organize the wavelet coefficients of $\mathbf{f} \in \mathbb{R}^n$ into a column vector $w \in \mathbb{R}^n$:

$$w = [c_0 \quad w_{0,0} \quad w_{1,0} \quad w_{1,1} \quad \cdots \quad w_{(N-1),(2^{N-1}-1)}]^T.$$

Furthermore, we denote the periodic wavelet reconstruction and decomposition by

$$\mathbf{f} = Bw \quad \text{and} \quad w = B^{-1}\mathbf{f}. \quad (7.11)$$

Exercise 7.1.1 Verify that the sets V_j defined by (7.2) satisfy conditions 1-4 of Definition 7.1.1.

Exercise 7.1.2 Show that the Haar mother wavelet computed from (7.5) agrees with (7.1).

Exercise 7.1.3 Consider the truncated sine function defined by

$$f(x) = \begin{cases} 0 & \text{for } x < 0, \\ \sin 2\pi x & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x > 1. \end{cases}$$

Express f in the Haar wavelet basis. Plot your results on the same axis as f for several choices of N in the expansion. How many wavelets are need to approximate f to an accuracy of 10^{-6} in ℓ^2 norm?

Exercise 7.1.4 Consider the more oscillatory function defined by

$$f(x) = \begin{cases} 0 & \text{for } x < 0, \\ \sin 8\pi x & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x > 1. \end{cases}$$

Express f in the Haar wavelet basis. Plot your results on the same axis as f for several choices of N in the expansion. How many wavelets are need to approximate f to an accuracy of 10^{-6} in ℓ^2 norm?

7.2 Besov spaces and wavelets

The Besov spaces are the natural function spaces to work with for wavelet representations of functions. Besov spaces are function spaces denoted by $B_{pq}^s(\mathbb{T})$, where $s \in \mathbb{R}$ is a smoothness index and $1 \leq p < \infty$ and $1 \leq q < \infty$ are integrability exponents. Roughly speaking, the greater s , the more derivatives a function $f \in B_{pq}^s(\mathbb{T})$ has with finite $L^p(\mathbb{T})$ norm. The following definition is from [333].

Definition 7.2.1 Suppose a wavelet family is r -regular or C^r regular (see [333]). A function f in $L^p(\mathbb{R})$, $1 \leq p \leq \infty$, represented as in (7.8) is in the Besov space $B_{pq}^s(\mathbb{T})$, $s < r$, $1 \leq p, q < \infty$ if the following norm is finite:

$$\|f\|_{B_{pq}^s(\mathbb{T})} := \left(|c_0|^q + \sum_{j=0}^{\infty} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_{k=0}^{2^j-1} |w_{j,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}$$

The choice $p = 1$ and $q = 1$ and $s = 1$ is interesting because it is related to the total variation norm; then the Besov norm takes the form

$$\|f\|_{B_{11}^1(\mathbb{T})} = |c_0| + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} 2^{j/2} |w_{j,k}|, \quad (7.12)$$

provided that $\psi_{0,0}$ and $\phi_{0,0}$ are once continuously differentiable [441, Thm 1.20].

7.3 Using B_{11}^1 regularization to promote sparsity

The regularized inverse problem by Besov space regularization is

$$\text{minimize} \left\{ \|Af - \mathbf{m}\|_2^2 + \alpha \|f\|_{B_{11}^1(\mathbb{T})} \right\}. \quad (7.13)$$

Such regularization was first suggested in [111], where the solution is based on a thresholding procedure. We will derive an alternative implementation following [282] based on constrained quadratic programming; the advantage of this approach is easy implementation of non-negativity constraints.

We truncate the wavelet expansion to a finite number of the coarseness scales in the following way. We only consider $n = 2^N$ with some $N \geq 0$, and define projection operator T_n acting on functions f expanded as (7.8) by

$$(T_n f)(x) = c_0 + \sum_{j=0}^{N-1} \sum_{k=0}^{2^j-1} w_{j,k} \psi_{j,k}(x). \quad (7.14)$$

Note that the right hand side of (7.14) spans an n -dimensional subspace of $B_{11}^1(\mathbb{T})$.

Now we can use (7.14) to construct a truncated version of the norm in (7.12). Defining $f_n \equiv T_n f$, the minimization problem (7.13) becomes

$$\text{minimize} \left\{ \|Af_n - \mathbf{m}\|_2^2 + \alpha \|f_n\|_{B_{11}^1(\mathbb{T})} \right\}, \quad (7.15)$$

where the Besov norm takes the finite form

$$\|f_n\|_{B_{11}^1(\mathbb{T})} = |c_0| + \sum_{j=0}^{N-1} \sum_{k=0}^{2^j-1} 2^{j/2} |w_{j,k}|. \quad (7.16)$$

The Besov norm defined by (7.16) takes the computationally effective form

$$\|f_n\|_{B_{11}^1(\mathbb{T})} = |(WB^{-1}\mathbf{f})_\nu|, \quad (7.17)$$

where W is a diagonal matrix containing the power-of-two weights that appear in formula (7.16).

The regularization task is an optimization problem with mixed ℓ^2 and ℓ^1 norms:

$$\arg \min_{\mathbf{f} \in \mathbb{R}^n} \left\{ \frac{1}{2\sigma^2} \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha_n \sum_{\nu=1}^n |(WB^{-1}\mathbf{f})_\nu| \right\}, \quad (7.18)$$

The minimization of (7.18) can be reformulated into a quadratic programming (QP) form as follows. Denote $WB^{-1}\mathbf{f} = \mathbf{u}^+ - \mathbf{u}^-$, where $\mathbf{u}^+, \mathbf{u}^- \geq 0$. Now the problem (7.18) can be written as

$$\arg \min_{\mathbf{z}} \left\{ \frac{1}{2\sigma^2} \mathbf{f}^T A^T A \mathbf{f} - \frac{1}{\sigma^2} \mathbf{f}^T A^T \mathbf{m} + \alpha_n \mathbf{e}^T \mathbf{u}^+ + \alpha_n \mathbf{e}^T \mathbf{u}^- + \frac{1}{2\sigma^2} \mathbf{m}^T \mathbf{m} \right\},$$

where $\mathbf{e} \in \mathbb{R}^n$ is vector of all ones, and we denote

$$\mathbf{z} = \begin{bmatrix} \mathbf{f} \\ \mathbf{u}^+ \\ \mathbf{u}^- \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{1}{\sigma^2} A^T A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} -\frac{1}{\sigma^2} A^T \mathbf{m} \\ \alpha_n \mathbf{e} \\ \alpha_n \mathbf{e} \end{bmatrix},$$

and $\alpha = \frac{1}{2\sigma^2} \mathbf{m}^T \mathbf{m}$. Now the minimization of (7.18) becomes

$$\begin{aligned} \min_{\mathbf{z}} \quad & \frac{1}{2} \mathbf{z}^T Q \mathbf{z} + \mathbf{c}^T \mathbf{z} + \alpha \\ \text{such that} \quad & \mathcal{A} \mathbf{z} = b \quad \text{and} \quad \mathbf{z} \geq \Lambda = \begin{bmatrix} -M \\ 0 \\ 0 \end{bmatrix}, \end{aligned} \quad (7.19)$$

where Λ is a lower bound for the primal variable z related to the following three inequalities: $u^+ \geq 0$ and $u^- \geq 0$ and $\mathbf{f} > -M$ (with some $M \gg 0$ so that the constraint on \mathbf{f} is practically ineffective). The matrix \mathcal{A} is an equality constraint matrix related to $WB^{-1}\mathbf{f} = u^+ - u^-$:

$$\mathcal{A} = [WB^{-1} \quad -I \quad I].$$

The regularized solution can then be computed using QP optimization methods, such as primal-dual path-following interior-point methods [478, 328, 451, 142, 363]. See Figure 7.1 for regularized deconvolutions computed as explained in [282].

See Section 9.1.3 for real-data tomographic reconstructions using Besov space regularization and sparsity-based parameter choice.

Exercise 7.3.1 Using formula (7.17), compute the Besov norm of the approximation to the function $f(x)$ in Exercise 7.1.3 for several values of N . What happens as N increases?

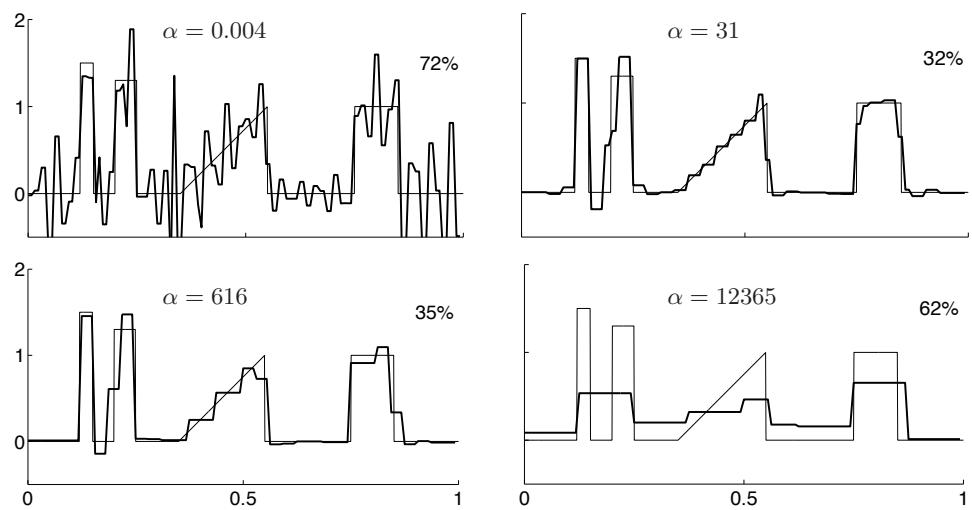


Figure 7.1: Besov space regularization using the Haar wavelet basis and $n = 128$. Note the edge-preserving nature of the reconstruction similar to total variation regularization.

Chapter 8

Discretization-invariance

In numerical analysis it is standard practice to analyze the error in the computational result as a function of the size of the grid. For example, the approximation in numerical integration quadrature is shown to be better when the number of quadrature points increases. Also, in finite element methods it is proven that if the diameter of the triangles in the mesh is proportional to h , then the solution of a partial differential equation is approximated better by FEM when the mesh is refined, or $h \rightarrow 0$.

Inverse problems often arise from continuous physical models. Such models must be discretized, or approximated by a finite-dimensional computational model, for calculating practical reconstructions using a computer. Usually there is no canonical computational grid but instead it can be rather freely chosen. As we saw in Section 3.3, the continuous model for the inverse problem $\mathcal{A}f = \mathbf{m}$ is ill-posed when \mathcal{A} is a compact operator, and so finite dimensional discretizations of this problem will result in increasingly ill-conditioned matrices as the mesh is refined and the discrete problem more closely approximates the continuous problem. However, this is not necessarily the case when we study the regularized problem.

Assume that there are k indirect measurements available in an inverse problem, and we choose to use n degrees of freedom in the computational representation of the unknown. It is natural to ask what happens to the regularized reconstruction when $n \rightarrow \infty$? This question has been studied surprisingly little so far in the inverse problems research community. In this section we use numerical examples to illustrate the concept of *discretization-invariance* related to the above question.

Consider an ill-posed inverse problem where a periodic function $f : \mathbb{T} \rightarrow$

\mathbb{R} needs to be recovered from an indirect measurement $\mathbf{m} \in \mathbb{R}^k$ modeled by

$$\mathbf{m} = \mathcal{A}f + \varepsilon, \quad (8.1)$$

where \mathcal{A} is a linear operator modeling the measurement and $\varepsilon \in \mathbb{R}^k$ is measurement error, here assumed to be white noise.

Define a grid of n points in the interval $[0, 1]$:

$$x_n(j) = \frac{j-1}{n} \quad \text{for } j = 1, 2, \dots, n, \quad (8.2)$$

and piecewise linear roof-top functions $\theta_j^n : \mathbb{T} \rightarrow \mathbb{R}$ by the requirement $\theta_\nu^n(x_n(j)) = \delta_{j\nu}$ for $j = 1, \dots, n$ and $\nu = 1, \dots, n$, where $\delta_{j\nu}$ is the Kronecker delta:

$$\delta_{j\nu} = \begin{cases} 1 & \text{when } j = \nu, \\ 0 & \text{when } j \neq \nu. \end{cases}$$

To discretize the inverse problem we use vectors

$$\mathbf{f} = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_{n-1}]^T \in \mathbb{R}^n, \quad (8.3)$$

and consider piecewise linear functions $f_n : \mathbb{T} \rightarrow \mathbb{R}$ of the form

$$f_n(x) = \sum_{\nu=1}^n \mathbf{f}_\nu \theta_\nu^n(x). \quad (8.4)$$

Now functions of the form (8.4) form a linear n -dimensional vector space. Also, it follows from (8.3) and (8.4) that

$$\mathbf{f} = [f_n(x_n(1)) \quad f_n(x_n(2)) \quad \cdots \quad f_n(x_n(n))]^T. \quad (8.5)$$

Let us return to the inverse problem of recovering the function f approximately from the measurement \mathbf{m} defined by (8.1). Fix n and assume that we have computational algorithms available for computing $\mathcal{A}f_n \in \mathbb{R}^k$ approximately for any given $\mathbf{f} \in \mathbb{R}^n$ and f_n defined by (8.4). Then we can compute regularized reconstructions by solving the following optimization problem:

$$\arg \min_{\mathbf{f} \in \mathbb{R}^n} \{\|\mathcal{A}f_n - \mathbf{m}\|_2^2 + \alpha_n \Psi(f_n)\}, \quad (8.6)$$

where $0 < \alpha_n < \infty$ is a regularization parameter and $\Psi(f_n)$ is a suitable penalty term. Note that Ψ takes as an argument a function defined on \mathbb{T} , not a vector. The idea behind this is to have a well-defined penalty term for functions that is naturally discretized by restricting it to finite-dimensional function spaces.

The concept of discretization-invariance arises from the observation that the dimension n in the minimization problem (8.6) can be chosen arbitrarily, in particular independently of the number k of measurements. It is natural to require that computational reconstructions (8.6) converge as the grid is refined arbitrarily, in other words as $n \rightarrow \infty$. Choices of Ψ and α_n in (8.6) satisfying this condition are called *discretization-invariant*.

Exercise 8.0.2 *Prove formula (8.5).*

8.1 Tikhonov regularization and discretizations

In the case of Tikhonov regularization, recall that the penalty term in (8.6) is defined as follows:

$$\Psi(f_n) = \|f'_n\|_{L^2(\mathbb{T})}^2 = \int_{\mathbb{T}} \left| \frac{df_n(x)}{dx} \right|^2 dx. \quad (8.7)$$

We set $\Delta_n := x_n(2) - x_n(1)$ and proceed to discretize the integral in (8.7) as a Riemann sum, and we use a finite difference approximation for the derivative. These approximations are actually exact for the piecewise linear functions of the form (8.4), resulting in

$$\begin{aligned} \int_{\mathbb{T}} \left| \frac{df_n}{dx}(x) \right|^2 dx &= \Delta_n \sum_{\nu=1}^n \left| \frac{f_n(x_n(\nu+1)) - f_n(x_n(\nu))}{\Delta_n} \right|^2 \\ &= \frac{1}{\Delta_n} \sum_{\nu=1}^n |\mathbf{f}_{\nu+1} - \mathbf{f}_\nu|^2 \\ &= \Delta_n \sum_{\nu=1}^n |(L\mathbf{f})_\nu|^2, \end{aligned} \quad (8.8)$$

where we use the notations $x_n(n+1) = x_n(1)$ and $\mathbf{f}_{n+1} = \mathbf{f}_1$ in accordance with the periodic boundary condition, and the $n \times n$ difference matrix L is given by

$$L = \frac{1}{\Delta_n} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \ddots & \\ \vdots & & & & & & \ddots \\ 0 & \cdots & & 0 & -1 & 1 & 0 \\ 0 & \cdots & & 0 & 0 & -1 & 1 \\ 1 & \cdots & & 0 & 0 & 0 & -1 \end{bmatrix} \quad (8.9)$$

Let us consider Tikhonov regularized deconvolutions at different resolutions n . Problem (8.6) takes the discrete form

$$T_{\alpha}^{(n)}(\mathbf{m}) = \arg \min_{\mathbf{f} \in \mathbb{R}^n} \left\{ \|A\mathbf{f} - \mathbf{m}\|_2^2 + \alpha_n \Delta_n \sum_{\nu=1}^n |(L\mathbf{f})_{\nu}|^2 \right\}, \quad (8.10)$$

which we solve using the stacked form formulation

$$T_{\alpha}^{(n)}(\mathbf{m}) = \begin{bmatrix} A \\ (\alpha_n \Delta_n)^{1/2} L \end{bmatrix} \backslash \begin{bmatrix} \mathbf{m} \\ 0 \end{bmatrix}, \quad (8.11)$$

where the backslash denotes minimum square norm solution.

As an example, we consider the 1-D deconvolution problem, and we simulate the measurement vector $\mathbf{m} \in \mathbb{R}^k = \mathbb{R}^{64}$ by computing the convolution with a highly accurate numerical integration at the 64 sample points, thus avoiding an inverse crime. White noise with standard deviation 0.01 is added to each component of \mathbf{m} .

We build a 64×64 measurement matrix A and use the L-curve method to find an appropriate regularization parameter α_{64} at the discretization $n = 64$. A suitable value is found to be $\alpha_{64} = 3.1 \cdot 10^{-5}$. Note that in this case $k = n = 64$.

Next we repeat the same reconstruction task with greater values of n while keeping \mathbf{m} and k fixed. Note that the inverse problem is now underdetermined. We keep things simple by choosing $n = 2^N$ with $N = 6, 7, 8, \dots, 13$; then all of the computational grids we use contain the 64-point grid as a subset. Consequently we can construct the $k \times n$ measurement matrix A as follows. Given $N > 6$, take $n = 2^N$ and let \tilde{A} be the $n \times n$ convolution matrix defined analogously to (2.14). Then we get the $k \times n$ matrix A by taking every $\frac{n}{64}$ th row of the larger matrix \tilde{A} .

Define $\alpha_n = \alpha_{64}$ for all $n > 64$. Disregarding a slightly different boundary condition, [303, Theorem 4.1.(i)] implies that the regularized solutions $T_{\alpha}^{(n)}(\mathbf{m})$ converge to a well-defined limit function as $n \rightarrow \infty$. Indeed, evaluating (8.11) numerically for $n = 64, 128, 256, \dots, 8192$ leads to almost identical functions, as shown in Figures 8.1 and 8.2. This is discretization-invariance.

Exercise 8.1.1 Verify that the choice $\alpha_{64} = 3.1 \cdot 10^{-5}$ is a suitable choice for the example above based on the L-curve method.

Exercise 8.1.2 Investigate the property of discretization invariance for the backward heat equation with Tikhonov regularization.

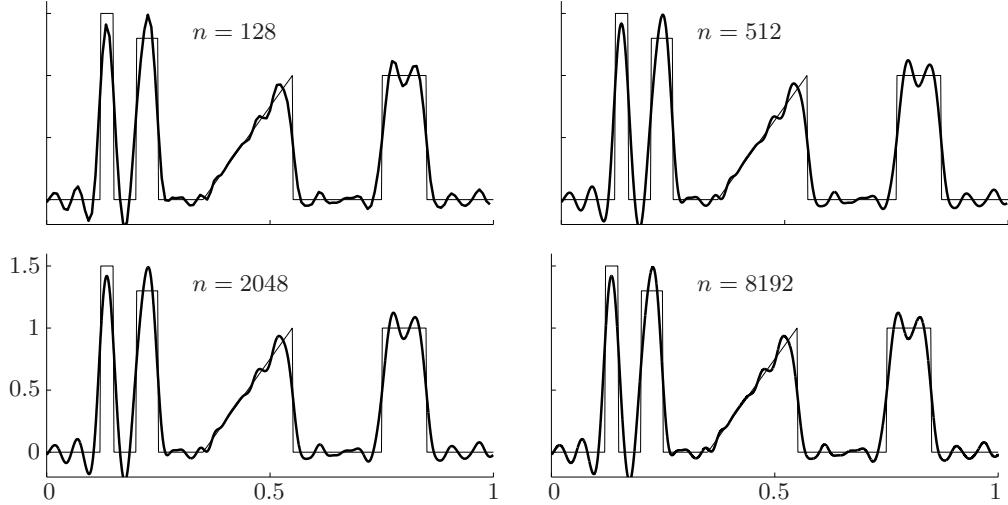


Figure 8.1: Discretization-invariant deconvolutions. Original signal (thin line) and Tikhonov regularized reconstructions (thick line). The number of measurements was $k = 64$ in each case. See also Figure 8.2.

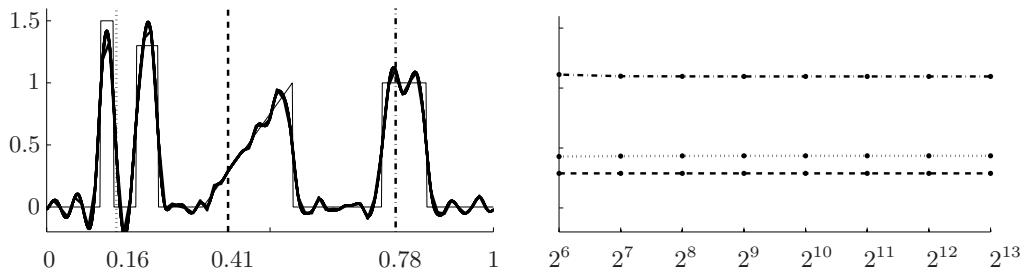


Figure 8.2: Discretization-invariant deconvolutions. Left: original signal (thin line) and Tikhonov regularized reconstructions (thick line) at resolutions $n = 64, 128, 256, \dots, 8192$. The reconstructions shown in Figure 8.1 are plotted here as well. Right: convergence of estimates at three points when the resolution n ranges from $2^6 = 64$ to $2^{13} = 8192$.

8.2 Total variation regularization and discretizations

For total variation regularization, the penalty term in (8.6) is defined as follows:

$$\Psi(f_n) = \|f'_n\|_{L^1(\mathbb{T})} = \int_{\mathbb{T}} \left| \frac{df_n(x)}{dx} \right| dx. \quad (8.12)$$

As above, we set $\Delta_n := x_n(2) - x_n(1)$, discretize the integral in (8.7) as a Riemann sum, and employ a finite difference approximation for the derivative. Similarly to (8.8) for functions of the form (8.4) the penalty term becomes

$$\begin{aligned} \int_{\mathbb{T}} \left| \frac{df_n}{dx}(x) \right| dx &= \Delta_n \sum_{\nu=1}^n \left| \frac{f_n(x_n(\nu+1)) - f_n(x_n(\nu))}{\Delta_n} \right| \\ &= \sum_{\nu=1}^n |\mathbf{f}_{\nu+1} - \mathbf{f}_\nu|. \end{aligned} \quad (8.13)$$

Compare (8.13) with (8.8); note especially how the factor Δ_n cancels in (8.13).

We compute total variation regularized deconvolutions with various resolutions n . See Figures 8.3 and 8.4 for the results. The close agreement of the reconstructions computed at different resolutions reflects the fact that total variation regularization is discretization-invariant [303, Theorem 4.1(ii)].

However, the statistical (Bayesian) interpretation of total variation is shown in [303] not to be discretization-invariant. This observation started a quest for finding edge-preserving or sparsity-promoting regularization methods that would allow a well-defined and discretization-invariant Bayesian interpretation. One possibility to achieve that is to use wavelets and Besov spaces, as shown theoretically in [302] and demonstrated numerically in [282].

8.3 Besov norm regularization and discretizations

The idea of using Besov space regularization for linear inverse problems was first introduced in [110]. Their Bayesian extensions are discretization-invariant, see [302, 282]. Discretization-invariance of deterministic Besov space regularization is proved in [169, 170, 282].

We define the penalty term in (8.6) as follows:

$$\Psi(f_n) = \|f_n\|_{B_{11}^1(\mathbb{T})}. \quad (8.14)$$

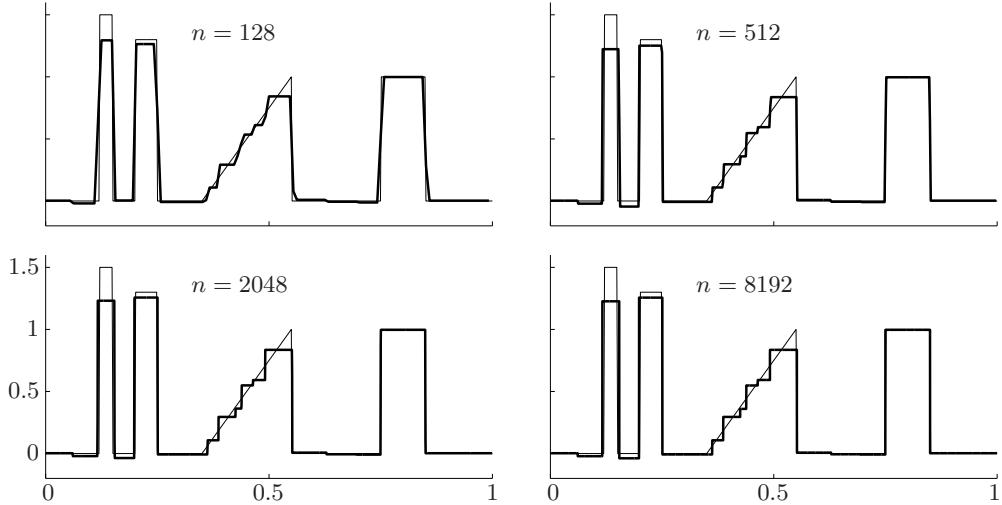


Figure 8.3: Discretization-invariant deconvolutions based on total variation regularization. Original signal is plotted using thin line and regularized reconstructions are plotted using thick line. Computations courtesy of Kati Niinimäki; see [282] for more details. See also Figure 8.4.

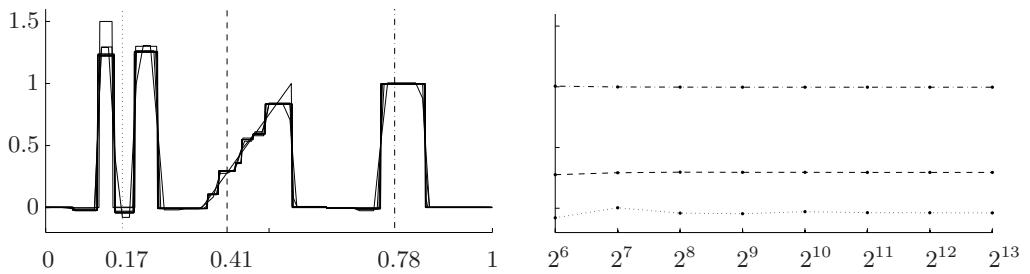


Figure 8.4: Discretization-invariant deconvolutions. Left: original signal (thin line) and total variation regularized reconstructions (thick line) at resolutions $n = 64, 128, 256, \dots, 8192$. The reconstructions shown in Figure 8.3 are plotted here as well. Right: convergence of estimates at three points when the resolution n ranges from $2^6 = 64$ to $2^{13} = 8192$.

Here $n = 2^N$ with some $N \geq 0$ and

$$f_n(x) = c_0 + \sum_{j=0}^{N-1} \sum_{k=0}^{2^j-1} w_{j,k} \psi_{j,k}(x). \quad (8.15)$$

According to [441, Theorem 1.20], if the mother wavelet $\psi_{0,0}$ is continuously differentiable, then $B_{11}^1(\mathbf{t})$ functions can be written in the form (8.15). We choose the orthonormal Daubechies 7 wavelet basis since its mother wavelet satisfies that condition. See the right column in Figure 8.5 and Figure 8.7 for the resulting reconstructions. The reconstructions seem to converge to a well-defined continuous limit, as predicted theoretically.

Since the wavelet expansion needs to be finite for computational reasons, the smoothness of the Daubechies 7 basis functions results in smooth reconstructions. Jumps in the signal are not recovered well, but the linear ramp in the middle of the interval is nicely recovered.

To promote edge-preservation we try using the discontinuous Haar wavelet basis in (8.15). Then it is unclear whether we are working with a norm of the space $B_{11}^1(\mathbf{t})$, but computationally there is no problem in proceeding like this. See the left column in Figure 8.5 and Figure 8.6 for reconstructions using the Haar wavelet basis. Numerical evidence suggests that these reconstructions are discretization-invariant as well.

The Haar-based reconstructions are edge-preserving, which is similar to the total variation regularized reconstructions shown in Figures 8.3 and 8.4. Note also the staircasing effect along the ramp in the middle of the interval; this feature is shared by both total variation regularization and Haar- $B_{11}^1(\mathbf{t})$ regularization.

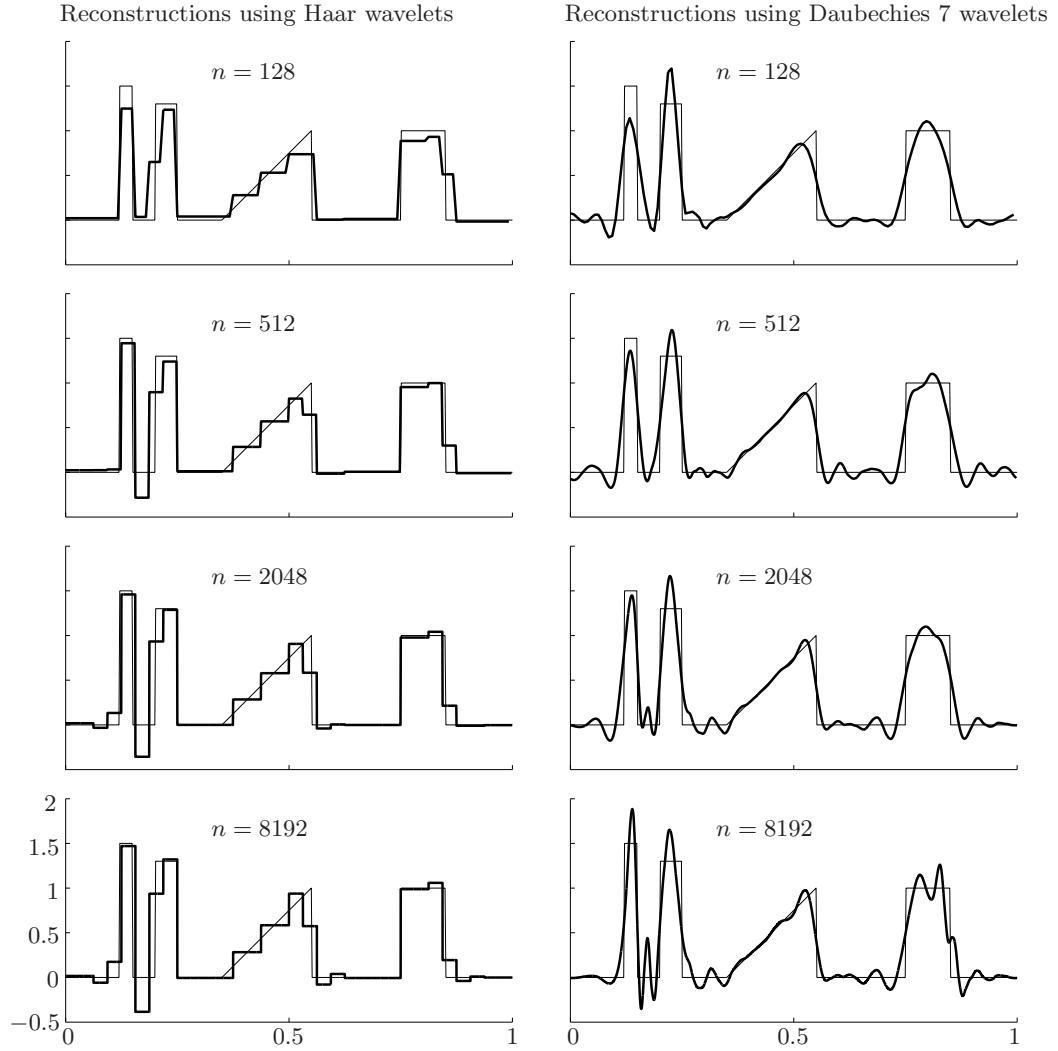


Figure 8.5: Discretization-invariant deconvolutions. Left column: original signal (thin line) and Besov norm regularized reconstructions (thick line) using Haar wavelet basis at various resolutions. See also Figure 8.6. Right column: original signal (thin line) and Besov norm regularized reconstructions (thick line) using Daubechies 7 wavelet basis at various resolutions. See also Figure 8.7. Computations illustrated in this Figure were implemented by Kati Niinimäki; see [282] for more details.

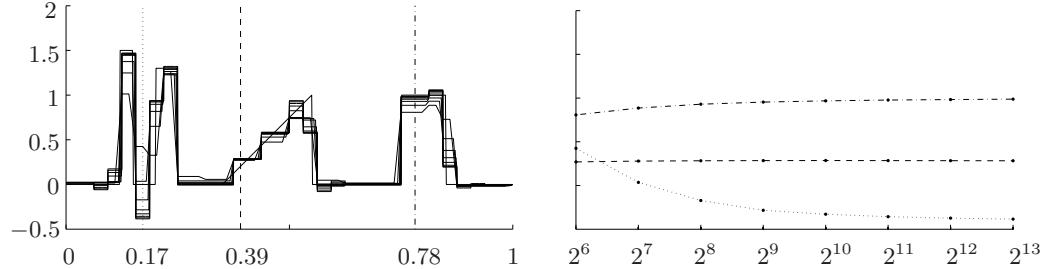


Figure 8.6: Discretization-invariant deconvolutions. Left: original signal (thin line) and Besov norm regularized reconstructions (thick line) using Haar wavelet basis at resolutions $n = 64, 128, 256, \dots, 8192$. The reconstructions shown in the left column of Figure 8.5 are plotted here as well. Right: convergence of estimates at three points when the resolution n ranges from $2^6 = 64$ to $2^{13} = 8192$.

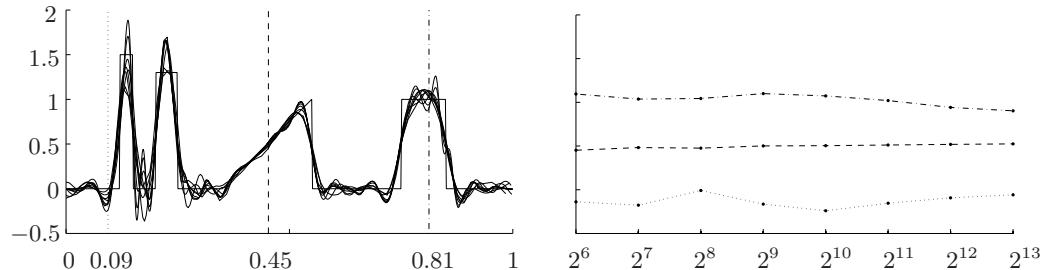


Figure 8.7: Discretization-invariant deconvolutions. Left: original signal (thin line) and Besov norm regularized reconstructions (thick line) using Daubechies 7 wavelet basis at resolutions $n = 64, 128, 256, \dots, 8192$. The reconstructions shown in the right column of Figure 8.5 are plotted here as well. Right: convergence of estimates at three points when the resolution n ranges from $2^6 = 64$ to $2^{13} = 8192$.

Chapter 9

Practical X-ray tomography with limited data

In many practical tomographic imaging situations there are restrictions on the number of available projection directions or on the angle of view. The reconstruction problem is more ill-posed when the data set is limited, as will be explained below in Sections 9.1 and 9.2. Recently there has been a lot of interest in limited-data imaging, and even some commercial products are available. We will describe one such product, intended for dental implant planning, in Section 9.3.

Before discussing limited-data tomography, let us comment on full-data tomography. Currently almost all commercial tomographic imaging is based on very comprehensive data sets and on the use of filtered back-projection (FBP) type algorithms. Those algorithms are proven to recover the attenuation perfectly in the idealized continuum case, which is approximated closely by collecting data with very small angular increments. The effect of measurement noise can be typically kept very low using mild regularization, such as an appropriate filter in the frequency domain.

When dealing with comprehensive tomographic data, there is usually not much difference in the outcomes of various inversion algorithms. To demonstrate this, let us simulate tomographic data from a two-dimensional 512×512 Shepp-Logan phantom using 512 angles uniformly distributed in the interval $0^\circ \leq \theta < 180^\circ$, avoiding inverse crime by simulating on a twice finer grid first and then interpolating as explained in Section 2.3.6. We call the resulting sinogram $\mathbf{m} \in \mathbb{R}^k$; with the above choices we have $k = 373248$. We simulate measurement noise by adding an independent normally distributed noise component with standard deviation $0.01 \cdot \max_{1 \leq j \leq k} |\mathbf{m}_j|$.

to each pixel in the sinogram. We recover the phantom using

1. filtered back-projection (FBP) as explained in Section 2.3.3,
2. classical Tikhonov regularization with the matrix-free large-scale variant of Section 5.5, and
3. matrix-free large-scale approximate total variation regularization with non-negativity constraint as explained in Section 6.4.1.

By *matrix-free* we mean the following. Note that $n = 512^2 = 262144$ and $k = 373248$, so the number of elements in A is $kn \approx 10^{11}$. Therefore, even storing A in computer memory is not advisable, not to mention the computation of the singular value decomposition. We do not construct the matrix A at all but instead use the Matlab routines `radon.m` and `iradon.m` for computing $A\mathbf{f}$ and $A^T\mathbf{g}$ for given vectors $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^k$. This is enough for implementing the large-scale methods described in Sections 5.5 and 6.4.1.

The reconstructions computed with the three different methods are shown in Figure 9.1. While the relative square norms errors are roughly 20% for all three methods, there is a big difference in computation times: FBP is done in less than half a second, but both matrix-free regularization methods take more than 200 seconds to run.

Our numerical example demonstrates the fact that whenever dealing with comprehensive, densely sampled tomographic data, it is most efficient to use FBP. In dimension two, good references include [351, 248, 352, 132]. In dimension three one typically uses helical trajectories for the X-ray source. Algorithms of the FBP flavor for various comprehensive 3D data sets have been developed in [258, 260, 259]. For reviews of the mathematics of tomography, see [418, 139].

Let us now turn to limited-data imaging. There are many kinds of naturally appearing restricted data sets, for example limited-angle data, sparse-angle data, and exterior or local tomographic data. Possible reasons for data limitations include

- Geometrically limited angle of view. In medical imaging this happens for example in mammographic imaging with the breast in a stationary position between the detector surface and a compression plate and the X-ray source moving in an arc [358, 480, 391, 494]. Another example is electron microscope imaging [138].
- Impossibility of probing through a ball in the center of the target. These so-called limited-angle exterior data sets arise for instance in

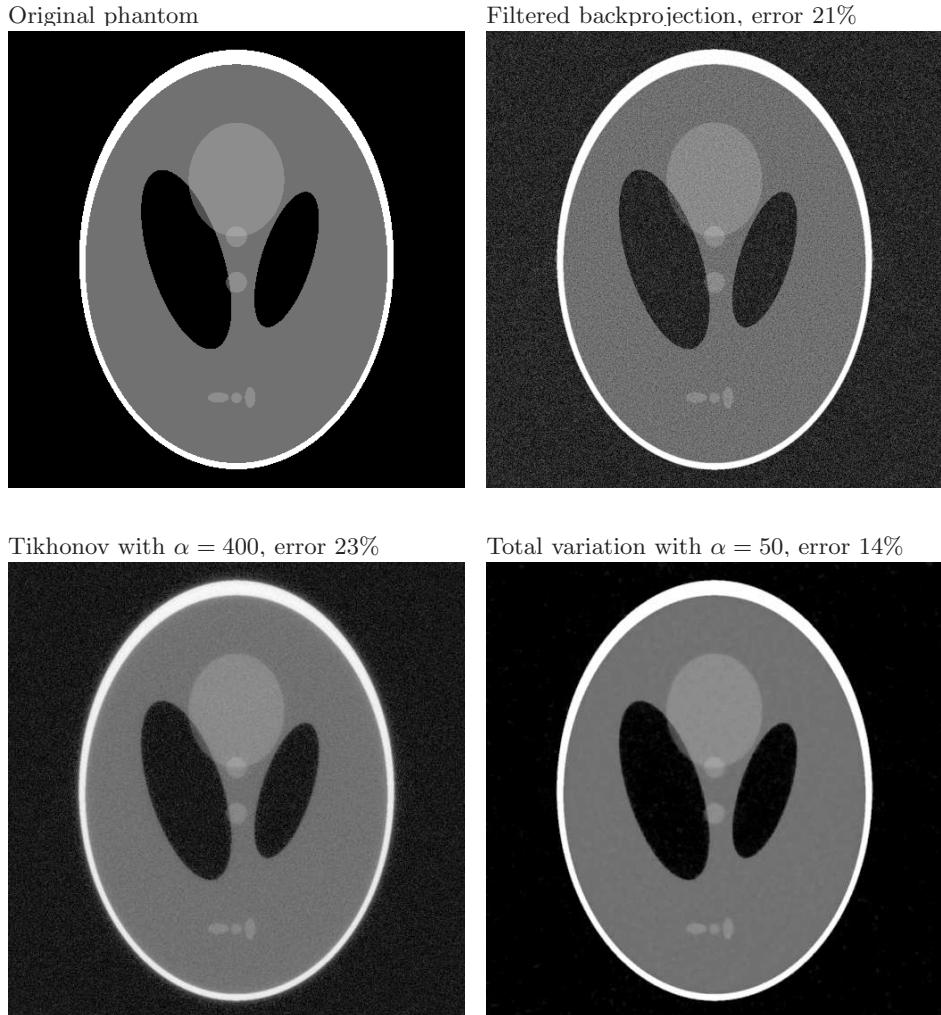


Figure 9.1: Comparison of reconstruction methods for densely sampled full-angle tomography. Here the phantom has size 512×512 and we use parallel beam geometry and 512 projection directions. The computation times of both regularization methods are more than 400 times longer than the time for the FBP algorithm, but there is no visible benefit for the extra investment in computational resources. In this kind of full-data situations FBP is the method of choice.

non-destructive testing [383, 382], reconstructing the structure of the solar corona [7], and in the recovery of ozone profiles around the Earth using space-borne star occultation data [295].

- The desire to lower the radiation dose to a patient by taking fewer images. (There is considerable recent interest in CT dose reduction, see [444, 292, 326, 492].) We discuss such a case in three-dimensional dental X-ray imaging in Section 9.3.
- Cost of equipment in cases when each projection direction needs its own X-ray source. This is often the case with luggage screening at airports using one or two sources [394]. Another tomographic example (not based on X-rays) is magnetospheric imaging where each projection direction requires an own satellite orbiting the Earth [153].

The filtered back-projection algorithm is not well-suited for limited tomographic data [351]. This well-known fact is demonstrated numerically in Sections 9.1 and 9.2, where we show how Tikhonov regularization and non-negativity constrained total variation regularization outperform FBP in the cases of sparse-angle and limited-angle tomography, respectively. See also the review article [373].

There are many successful computational approaches to sparse-data tomography discussed in the literature. Here is an incomplete list of results, based on Tikhonov regularization [217], total variation regularization [115, 375, 288, 409, 410, 289, 313, 411, 205, 432, 128, 46, 45, 238, 236, 435], strict piecewise-constant attenuation assumption [336, 424, 6, 396, 416, 439], level set methods [489, 140, 388, 287, 485, 271], modified algebraic reconstruction [377, 312], deformable models [196, 32, 197, 33, 335], variational methods [294, 121], sparsity-promoting methods [68, 67, 490, 491, 84, 237, 296], Bayesian methods [413, 246, 65], and multiresolution-sparsity methods [391, 357, 423, 452, 270]. See also the books [206, 207] for discrete tomography approaches.

We stress that any given solution method is probably not the best choice for *every* limited-data tomographic application. With severely limited data, each method produces errors and artefacts of its own kind, and it is up to the application area to choose the most suitable one. There are big differences in the computational burden of various methods, too.

9.1 Sparse full-angle tomography

In some tomographic applications it is possible to take images from all around the object, but the total number of images is restricted. We consider here taking images from full angle but with large uniform angular steps and call such datasets *sparse-angle data*.

One might think that it is impossible to compute reasonable reconstructions from sparse-angle data. Namely, it was noted already in [99], and later analyzed in [419, Theorem 4.2], that a finite number of line integrals does not determine the target uniquely since the measurement operator has a non-trivial nullspace. However, the “ghosts”, or the objects invisible in the tomographic data, are known to be high-frequency functions [317, 319, 318, 324] and can be effectively suppressed by regularization.

Sparse-angle data violates the assumptions of filtered back-projection algorithms, and consequently FBP reconstructions typically contain a lot of artefacts as we will see below. In Sections 9.1.1–9.1.3 we demonstrate numerically Tikhonov, total variation, and Besov norm regularization for sparse-angle tomography and compare the results to those from FBP.

Exercise 9.1.1 *Compute and plot an invisible structure in tomographic imaging following the analysis in [401].*

9.1.1 Simulated data

We consider tomographic data taken from full angle of view but consisting only of 12 projection images with 30 degrees angular steps.

Note that here $n = 512^2 = 262144$, and the sinogram has $k = 128 \times 185 = 23680$ data points. The measurement matrix A has size 8748×262144 , so the number of elements in A is approximately $2.3 \cdot 10^9$. We use the matrix-free algorithms described in Sections 5.5 and 6.4.1 for the Tikhonov and total variation regularization, respectively. See Figure 9.2 for the reconstructions.

Table 9.1 shows the relative square-norm errors in the reconstructions and the computation times of the three algorithms. The computation time of the filtered back-projection algorithm is very short, but the result has low quality. In this case it is beneficial to invest in computational resources and use total variation regularization, as the reconstruction error can be reduced from 126% to 19%. This is in contrast with the dense-data case shown in Figure 9.1.

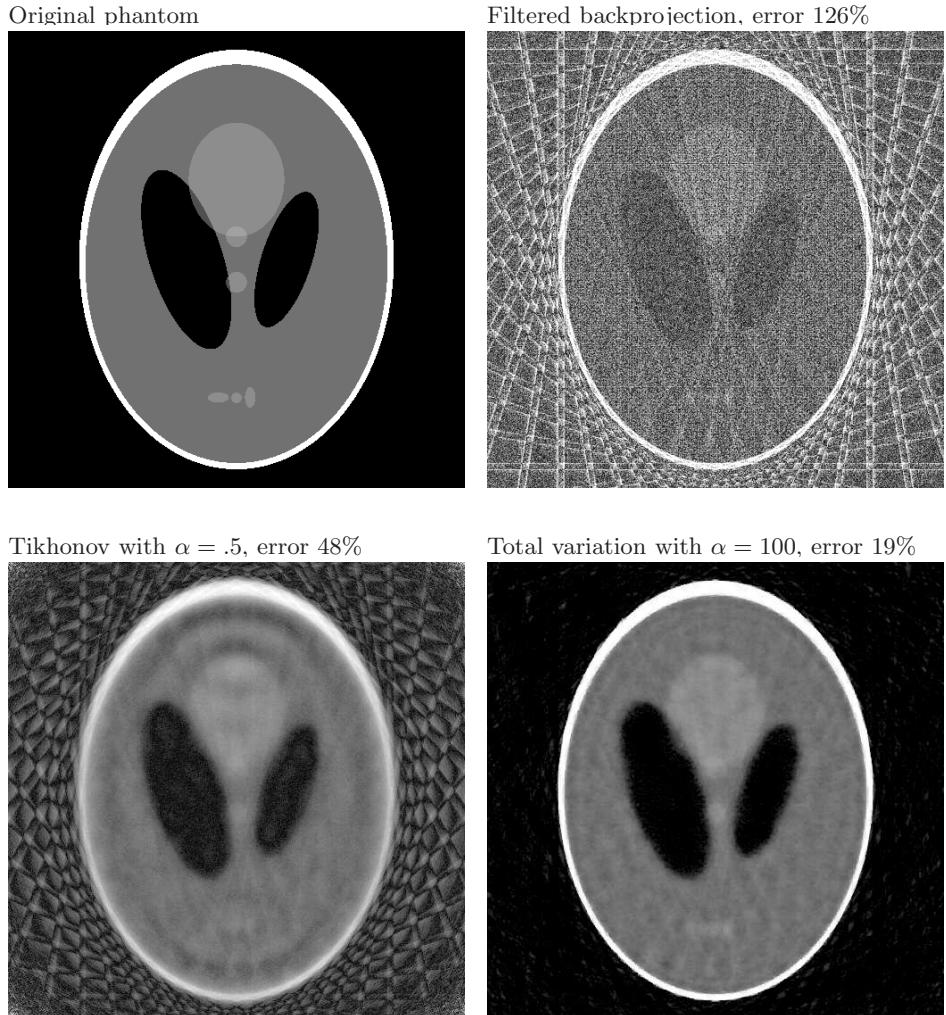


Figure 9.2: Comparison of reconstruction methods for sparsely sampled full-angle tomography. Here we simulated parallel-beam data from 20 uniformly spaced directions. The size of the reconstruction is 512×512 . Approximate total variation regularization with non-negativity constraint outperforms the competing methods clearly. See Table 9.2 for the computation times of the various methods. Also, compare these reconstructions with the dense-angle case shown in Figure 9.1; in the sparse-angle case there is much clearer difference between the reconstructions.

Table 9.1: Computation times and relative square-norm errors of reconstructions from sparsely sampled full-angle tomographic data. See Figure 9.2 for the reconstructions.

Reconstruction method	Error	Computation time (s)
Filtered back-projection	126%	0.4
Tikhonov regularization	48%	69.5
Total variation regularization	19%	33.4

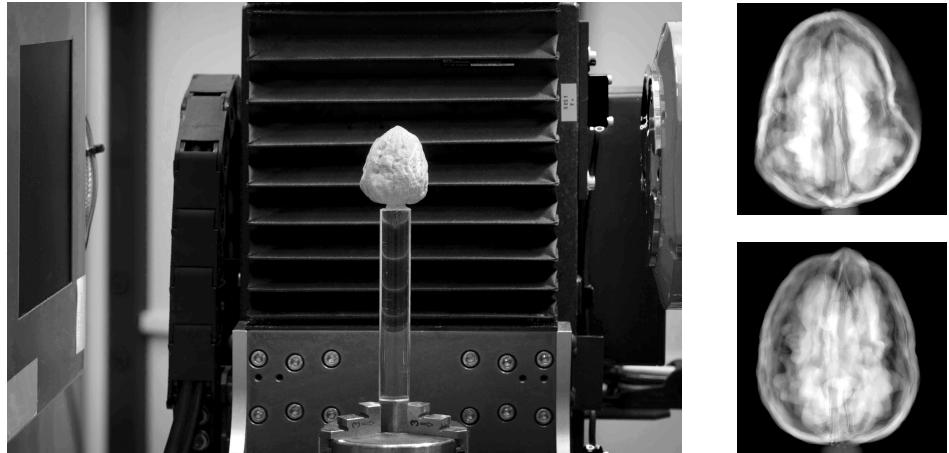


Figure 9.3: Left: Experimental setup for collecting tomographic X-ray data of a walnut. The detector plane is on the left and the X-ray source on the right in the picture. The walnut is attached to a computer-controlled rotator platform. Right: Two examples of the resulting projection images.

9.1.2 Real data: total variation regularization

We demonstrate the large-scale inversion methods introduced earlier in this book in the case of measured X-ray projection data. Our target is a walnut, and Keijo Hämäläinen and Aki Kallonen from University of Helsinki were kind enough to measure projection data for us. The nut was measured using cone-beam geometry and collecting data from 90 directions.

The problem was reduced to a two-dimensional reconstruction task by choosing only the middle row of each projection image; this resulted in a fan-beam geometry. Each (one-dimensional) projection image consists of 512 measured attenuation values. We measured 1200 projection images from uniformly distributed angles around 360° and computed a filtered backprojection reconstruction using all this data. The resulting image is shown in Figure 9.4 and is used as ground truth to which reconstructions from fewer projections can be compared.

We wish to study the performance of filtered backprojection and total variation regularization as the number of projection angles is reduced progressively.

The data was downsampled to 512 pixels per projection image, and rebinned to a parallel-beam geometry. This operation contains some interpolation error which is not involved in the computation of the ground truth

image shown in Figure 9.4 as it used a built-in cone-beam reconstruction routine of the X-ray laboratory system. We pick three subsets from the original dense-angle dataset. The subsets consist of 10, 15 and 23 projections, respectively, each set spanning 180° . The number of data points in each case is $k = 10 \cdot 512 = 5120$ and $k = 15 \cdot 512 = 7680$ and $k = 90 \cdot 512 = 11776$, respectively.

After rebinning, parallel-beam projections and back-projections were readily implemented for the three datasets using Matlab's `radon.m` and `iradon.m` functions. The total variation regularized reconstruction is computed in all three cases on a 256×256 grid, so $n = 65536$. The computations were performed by Kati Niinimäki using a quadratic programming approach as explained in Section 6.2. See Figure 9.5 for the results.

It is clear from Figure 9.5 that filtered back-projection is not suitable for sparse-angle imaging, and that total variation regularization delivers useful piecewise-constant low-resolution reconstructions.

The results and images presented in this section are due to a collaboration with Keijo Hämäläinen, Aki Kallonen, Ville Kolehmainen, Matti Lassas and Kati Niinimäki. For more details, see [186].

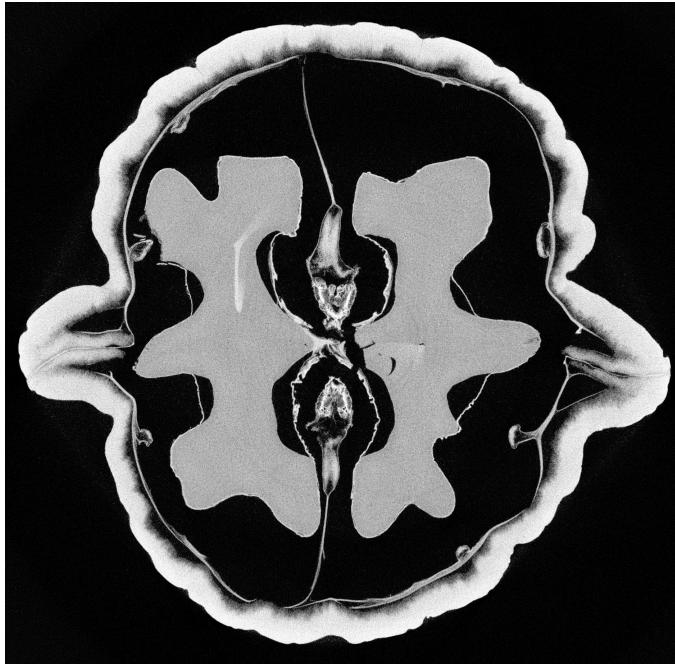


Figure 9.4: The tomographic measurement shown in Figure 9.3 involved recording projections from 1200 uniformly distributed angles. This image shows the filtered back-projection reconstruction from the full dense-angle data set. This is the ground truth to which the reconstructions computed using fewer projections can be compared.

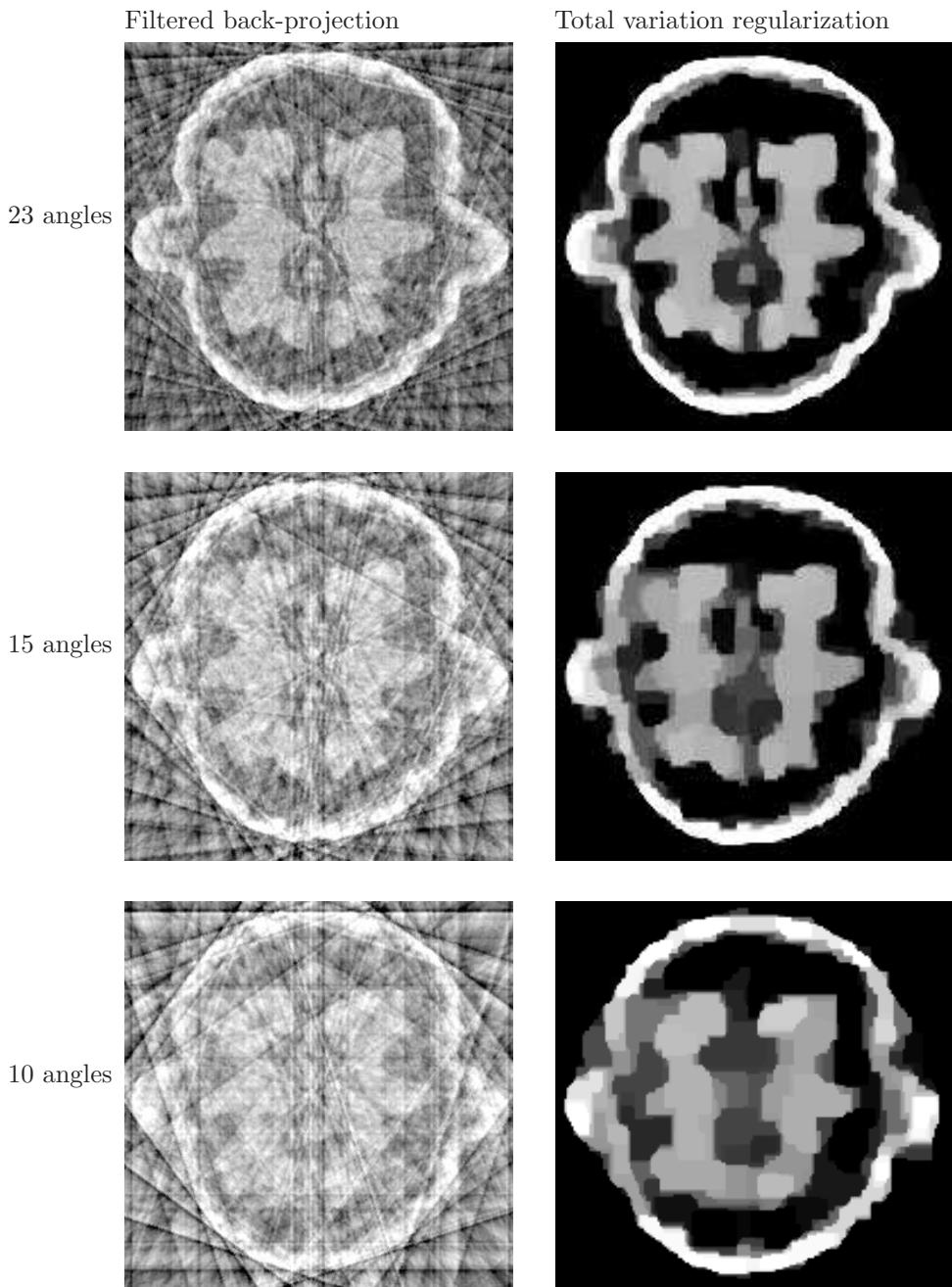


Figure 9.5: Sparse-angle tomographic reconstruction of a slice of a walnut using filtered back-projection (left column) and total variation regularization (right column). The number of equally distributed projections is indicated on the left. These images are computed by Kati Niinimäki. Compare the reconstructions to the ground truth shown in Figure 9.4.

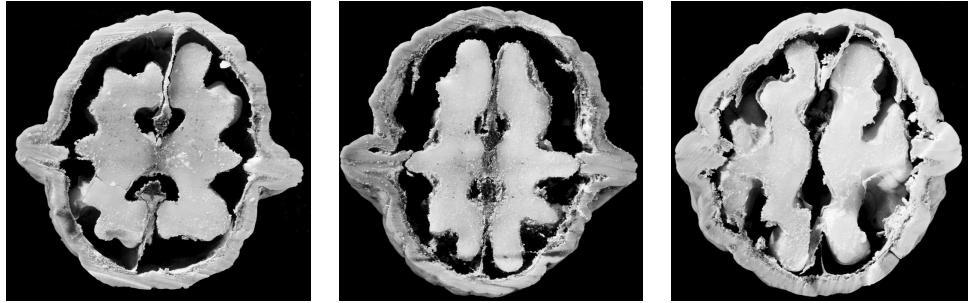


Figure 9.6: Photographs of walnuts sawed in half. These photos are used for estimating the expected number of nonzero wavelet coefficients in a two-dimensional tomographic reconstruction. Special thanks go to Esa Niemi for his careful job in sawing the walnuts.

9.1.3 Real data: sparsity-promoting tomography

We demonstrate the sparsity-based parameter choice rule introduced in Section 6.3 here using X-ray tomography and Besov space regularization. The method requires *a priori* knowledge about the number of nonzero wavelet coefficients in the unknown attenuation coefficient. How would we know such a thing? In medical imaging applications, for example, we might have available an atlas of high-resolution CT slices of patients. We could then analyze the sparsity levels of those slices and use that knowledge to set the desired number of nonzero coefficients in the new (unknown) target.

We model the medical atlas scenario by sawing three walnuts in half and photographing the exposed slice. See Figure 9.6 for the photos. Further, we compute the wavelet transforms of the digital photographs and see how many of them are above a small threshold. It turns out that in size 128×128 , in average 5936 wavelet coefficients out of the total of $128^2 = 16384$ are essentially nonzero.

We use the same 90-angle dataset than in Section 9.1.2. The reconstruction size is chosen to be $n = 128 \times 128$. We solve the minimization problem

$$\text{minimize} \left\{ \|Af - \mathbf{m}\|_2^2 + \alpha \|f\|_{B_{11}^1(\mathbb{T}^2)} \right\}$$

using Haar wavelet basis and a quadratic programming code developed by Kati Niinimäki. The regularization parameter ranges between 10^{-6} and 10^4 , and we compute reconstructions using 19 values of α in that interval. The result is readily expressed in wavelet transform domain, and we compute

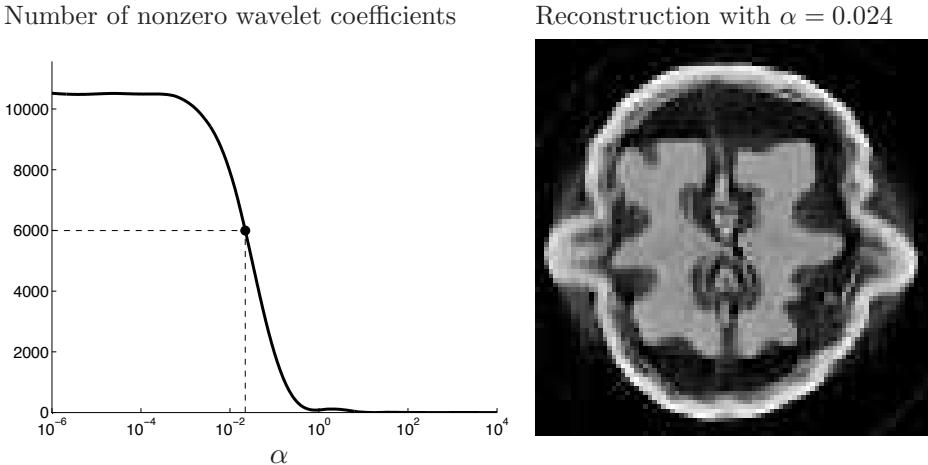


Figure 9.7: Reconstruction using Besov space regularization and sparsity-based choice of regularization parameter. Assume that we know *a priori* that the reconstruction should have 5936 nonzero wavelet coefficients (out of the maximal number of $128^2 = 16384$). Left: plot of a curve interpolating the numbers of nonzero wavelet coefficients in reconstructions computed using different regularization parameters α . Note that the horizontal axis is logarithmic. Right: reconstruction with the parameter $\alpha = 0.024$ leading to correct sparsity. Compare to the ground truth shown in Figure 9.4.

the number of (essentially) nonzero wavelet coefficients of each of the 19 reconstructions. Further, we fit a smooth interpolation curve to the numbers of nonzeros. See the left plot in Figure 9.7.

According to the sparsity-based choice rule for regularization parameter, we use the interpolation curve (the S-curve) in the left plot in Figure 9.7 to find the value of α that gives most closely the *a priori* known number 5936 of nonzero coefficients. The result is $\alpha = 0.024$. The resulting reconstruction is shown on the right in Figure 9.7.

The results and images presented in this section are due to a collaboration with Keijo Hämäläinen, Aki Kallonen, Ville Kolehmainen, Matti Lassas, Esa Niemi and Kati Niinimäki. For more details, see [185].

9.2 Limited-angle tomography

In limited angle tomography the projection images are available only from a restricted angle of view. From a purely theoretical point of view this may appear not to be a big problem. Namely, as discussed in [419, 351], having projection data available for example in dimension two from an angular range $\theta \in [-\varepsilon, \varepsilon]$ for any $\varepsilon > 0$ amounts to knowing the Fourier transform of the attenuation coefficient in a bowtie-shaped set in the frequency domain. In case the target has finite extent, the attenuation coefficient is compactly supported, and therefore its Fourier transform is real-analytic. Since we know the Fourier transform in a set containing an open disc, by analyticity we know it in the whole plane. Consequently we can recover the attenuation coefficient perfectly.

The above reasoning is based on the notoriously unstable theoretical device of analytic continuation. Practically, when the angle of view decreases from 180 degrees towards zero, the reconstruction problem becomes increasingly ill-posed. The sensitivity of the limited-angle reconstruction to measurement noise is analysed quantitatively using singular value expansions in [431, 112, 320]. These studies show that limited angle tomography is a very challenging inverse problem, especially with narrow angles of view.

We demonstrate Tikhonov and total variation regularization for limited-angle tomography with simulated data. We compute limited-angle tomographic data for the 512×512 Shepp-Logan phantom using 12 angles from 60 degrees angle of view. See Figure 9.8 for the reconstructions. Note the typical “stretching” of the reconstructions in the general direction of the X-rays. Table 9.2 lists relative errors in the reconstructions.

Table 9.2: Computation times and relative errors of reconstructions from limited-angle tomographic data. See Figures 9.8 and 9.9 for the reconstructions.

Reconstruction method	Error	Error (profile)	Comp. time (s)
Filtered back-projection	212%	230%	0.1
Tikhonov regularization	64%	38%	10.2
Total variation regularization	58%	21%	14.7

Despite the distorted nature of the reconstructions in Figure 9.8, some features are actually recovered quite reliably. According to the microlocal

analysis presented in [177, 381], there are some boundaries of objects inside the target that are stably represented in the limited-angle data. Roughly speaking, those boundaries that are tangent to X-rays can be reliably reconstructed. See Figure 9.9 for horizontal profiles of the reconstructions in Figure 9.8 and note that the locations of the jumps are rather well reconstructed, as predicted by microlocal analysis. See Table 9.2 for relative errors in the reconstructed profiles.

Three-dimensional reconstructions computed from measured limited-angle X-ray data are shown in Section 9.3.4 below.

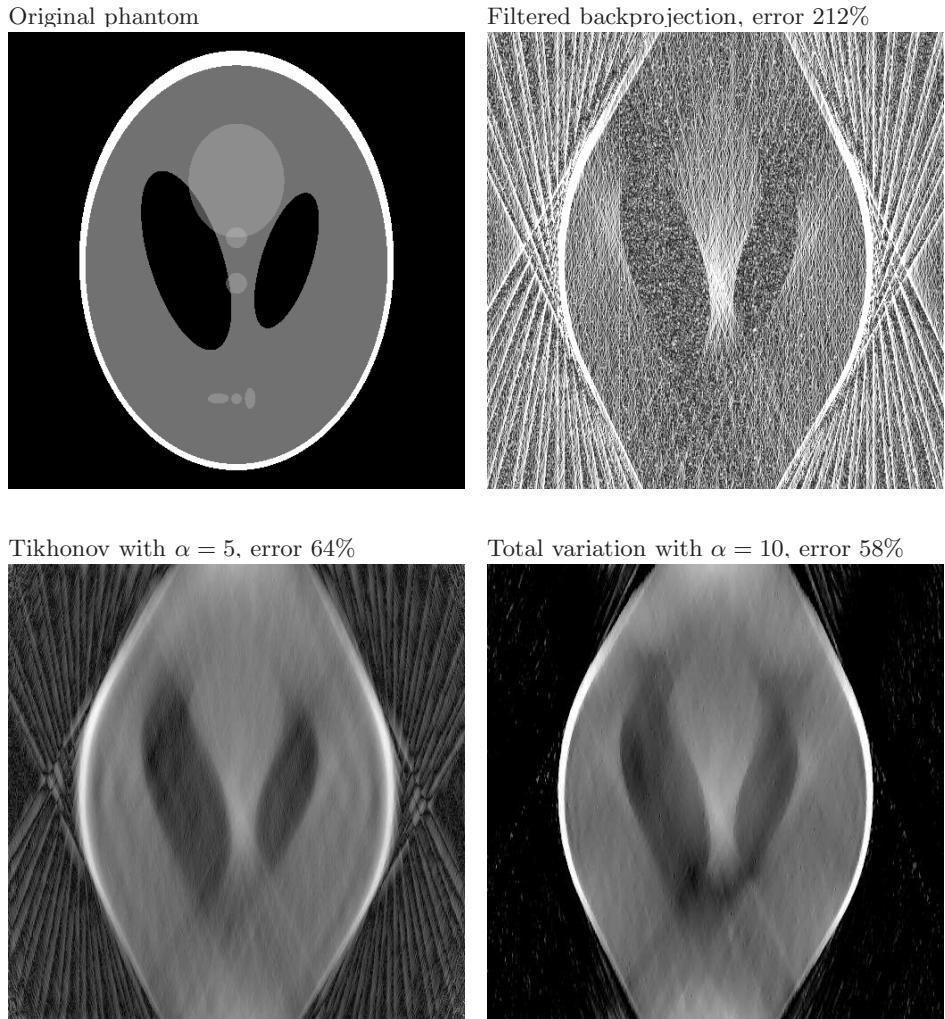


Figure 9.8: Comparison of reconstruction methods for limited-angle tomography. Note the typical “stretching” of the reconstructions along the directions of view. According to the microlocal analysis presented in [177, 381], some parts of the boundaries can be reconstructed better than others; in this case close to vertical boundaries. See Figure 9.9 for plots of horizontal profiles through the above reconstructions.

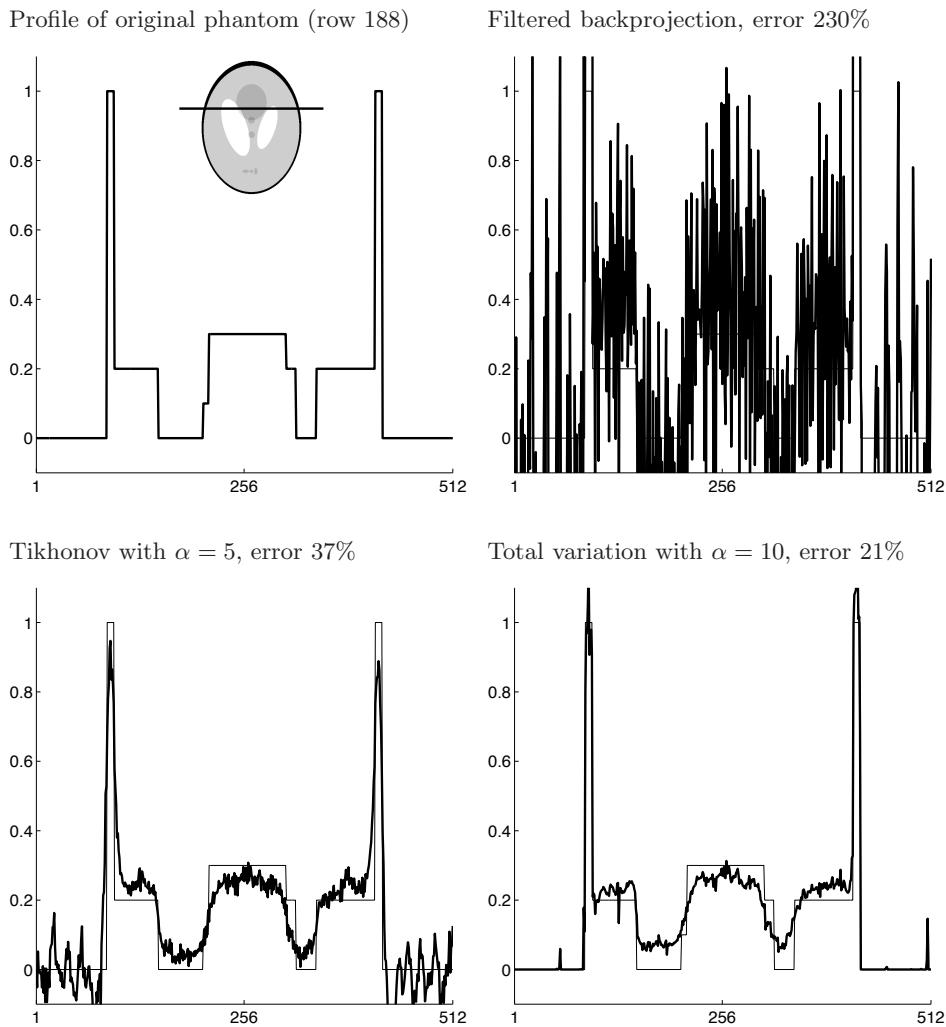


Figure 9.9: Horizontal profiles through the limited-angle reconstructions of Figure 9.8.

9.3 Low-dose 3D dental X-ray imaging

In filtered back-projection the mathematics requires dense sampling of data, leading to high radiation doses. In low-dose three-dimensional X-ray imaging the thinking is reversed. Only the least possible amount of X-ray data is collected, and limitations of data are compensated for with tailored inversion algorithms. The reconstructions are not perfect, but they only need to be good enough for the clinical application at hand.

Now the mathematics does not dictate the dose, but instead the dose dictates the mathematics!

This section is dedicated to the description of the first commercial medical imaging apparatus based on limited-angle X-ray tomography. It is a novel application of a so-called panoramic imaging device that every dental clinic has nowadays. Programming new movements to the machine allows the collection of tomographic projection data. In essence, a two-dimensional imaging device that the clinic already has is reprogrammed, and suitable mathematics is developed for compensating for the incomplete dataset. Here is an example where computational inversion is the core feature enabling a breakthrough in imaging technology.

Before discussing the new imaging device in Section 9.3.4 we need a couple of historical digressions.

9.3.1 Tomosynthesis

Tomosynthesis was invented (or at least first published) by a Dutch scientist called Ziedses des Plantes [119]. Later, tomosynthesis was named and developed further by David G. Grant [168]. The idea is to take X-ray images of a three-dimensional object from different directions and then use a simple shift-and-add algorithm to combine the images in such a way that details in a so-called *sharp layer* are more clearly visible than details away from that layer. We illustrate the shift-and-add process in Figures 9.10 and 9.11.

Mathematically, tomosynthesis is easily seen to be nothing else but *unfiltered* back-projection.

See the articles [179, 496, 165, 120, 123, 122, 440, 25] for more information on tomosynthesis and its applications.

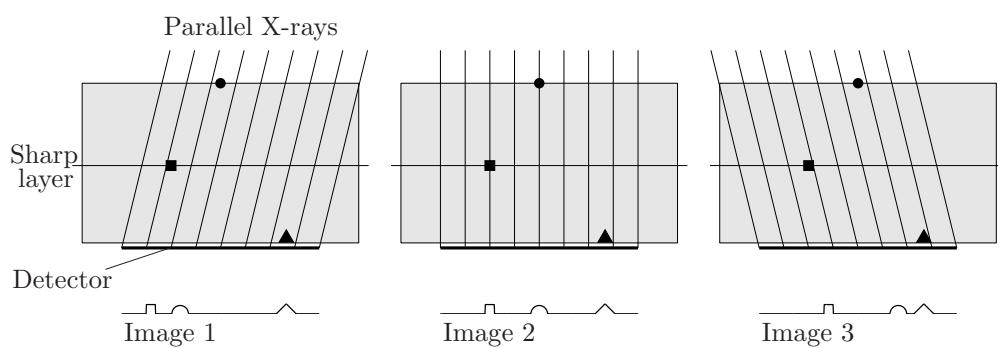
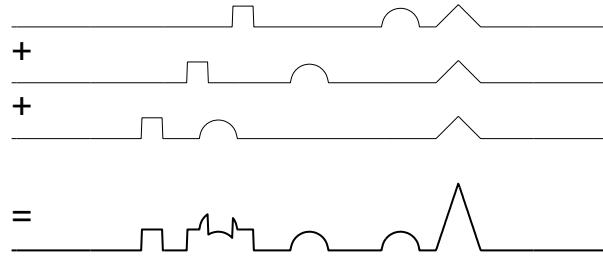
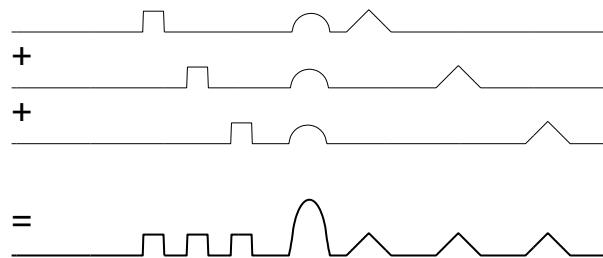


Figure 9.10: Image acquisition in tomosynthesis. X-ray images of the same object are taken from different directions. In this schematic two-dimensional illustration we use parallel-beam geometry for simplicity.

(a) Adding the images without shifting aligns the features near the detector:



(b) Shifting the images so that the images of the ball overlap:



(c) Shifting half as much as with the ball aligns features at the sharp layer:

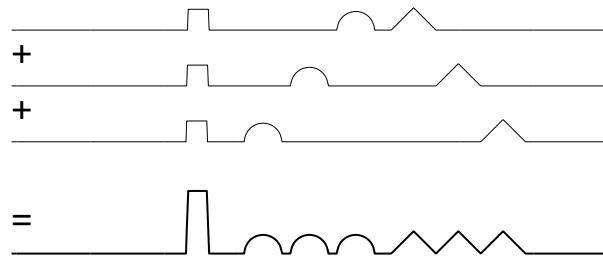


Figure 9.11: Computation of sharp layers in tomosynthesis. This is the so-called shift-and-add algorithm. See Figure 9.10 for the structure of the object under imaging.

9.3.2 Tuned-aperture computed tomography

Richard L. Webber patented in 1994 an imaging method called *tuned aperture computed tomography* (TACT), where the idea is to record X-ray images of a target from different directions, calibrate the directions using a fiducial reference ball visible in each image, and use tomosynthesis for three-dimensional reconstruction. The reference ball is typically placed on top of the object to be imaged (as in Figure 9.10) and keep the object and detector in fixed positions relative to each other.

The goal of Webber's work is to perform three-dimensional medical imaging with simple and cost-effective equipment. Any combination of an X-ray source and digital detector will do. This approach is in stark contrast with the trend of using big and expensive MRI and CT machines for 3D imaging.

"Tuning" the aperture is related to the possibility of modifying the thickness of the sharp layer by choosing the imaging angles suitably. Roughly speaking, a wider angle of view results in narrower depth-of-field in the tomosynthetic slice.

The TACT imaging approach is useful for several applications in dental radiography, see [471, 472, 473, 474].

9.3.3 Panoramic dental imaging

Starting in the beginning of twentieth century there was a race to develop a device that could take an X-ray image of all the teeth simultaneously. The first approaches were based in having a film inside the mouth and using a narrow X-ray beam as shown in Figure 9.12. The pioneering work of Hisatugu Numata [366] deserves special mention.

Yrjö Veli Paatero (1901-1963) defended his thesis in dentistry in 1939 and was nominated in 1945 to a teaching position in the Department of Dentistry of University of Helsinki, Finland. He was responsible for X-ray diagnostics, and spent a lot of time in the tedious routine work of imaging the teeth of individual patients using several small intraoral films. He started to develop better methods for imaging all the teeth, starting with intraoral approaches [371]. Soon thereafter he came up with the idea of imaging a sharp layer following the dental arc using a geometric movement; in this approach both the film and the X-ray source are outside the mouth [372]. Paatero's invention received international attention, and he received an invitation to University of Washington, Seattle, for the period 1950–51. During that time he developed panoramic imaging further, and continued his work after returning to Finland on 1951.

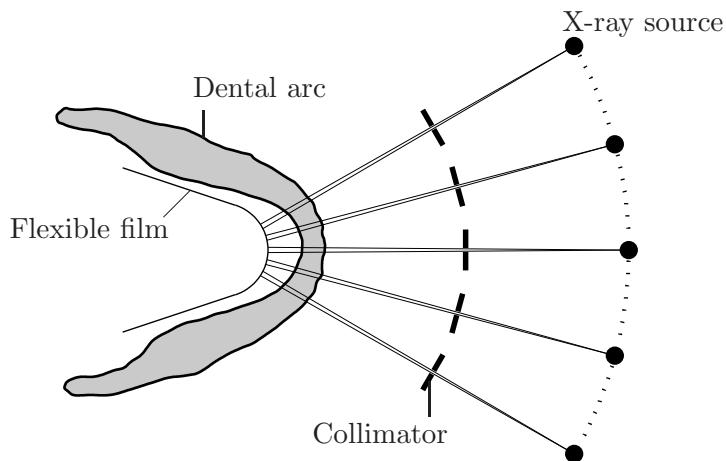


Figure 9.12: Principle of film-based intraoral (film inside the mouth) panoramic dental X-ray imaging. The X-ray beam is narrowed using a collimator, which is essentially a slit in a radiopaque lead plate.

The principle of the panoramic image formation is based on an ingenious combination of two simultaneous movements. A fixed arrangement of an X-ray source and film holder rotates with the patient's head located at the center of rotation. X-rays reach the film through a collimator (a narrow slit in a radiopaque lead plate). The film moves linearly with respect to the holder, effectively picking out features located in the sharp layer and blurring features away from the sharp layer. The faster the linear movement of the film, the larger is the radius of the circular sharp layer. See Figure 9.13 for a simplified illustration.

The formation of the sharp layer is basically a continuum version of the tomosynthesis principle explained in Section 9.3.1 and images 9.10 and 9.11.

In modern digital panoramic devices the physical movement of the film is replaced by electronic movement of pixels in a narrow charge coupled device detecting the radiation. See Figure 9.14 for a picture of a modern panoramic imaging device and Figure 9.15 for a digital panoramic image.

In recognition of his groundbreaking work, Paatero was made the chairman of the Oral Radiology at the University of Turku, Finland, in 1961. However, he died tragically soon thereafter in 1963. His long-time collaborator Timo Nieminen founded a company called Palomex in 1964 and started

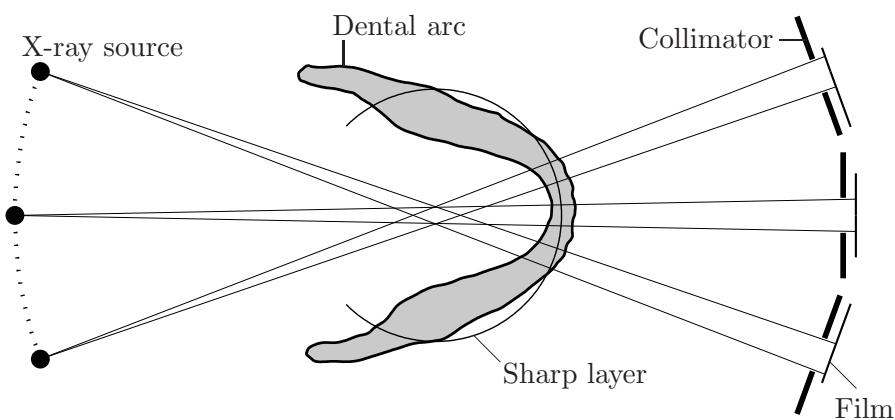


Figure 9.13: Principle of film-based extraoral (film outside the mouth) panoramic dental X-ray imaging, shown in a simplified cylindrical geometry and exaggerated angles for clarity. Note the crucial linear movement of the film with respect to the collimator slit. Due to the simplified geometry, the sharp layer does not follow the dental arc properly. In practice this problem is solved by varying the linear speed of the film and moving the center of rotation during the exposure.



Figure 9.14: Digital panoramic imaging device (Instrumentarium Dental OP200D). Photograph courtesy of PaloDEx Group. See Figure 9.15 for a digital panoramic image.



Figure 9.15: A typical digital panoramic image, taken with the Instrumentarium Dental OP200D device shown in Figure 9.14. Image courtesy of PaloDEx Group.

manufacturing panoramic X-ray devices. The idea of panoramic imaging caught on in the 1970's, and since then a panoramic device has been considered a necessary tool for every dental clinic.

For more details on the history of panoramic imaging the reader is referred to [184].

9.3.4 The VT device for dental implant planning

The Palomex company mentioned in Section 9.3.3 was acquired in 1977 by Instrumentarium Corporation, a Finnish company founded in the year 1900. The Instrumentarium Imaging division concentrated on various medical X-ray technologies. The story of the VT device started in 1998, when the company licensed the patent for TACT imaging (discussed in Section 9.3.2).

Instrumentarium Imaging planned to use TACT technology for several applications, including mammography, surgical C-arm imaging and dental imaging. Research projects were initiated in 2001 together with academic partners and funding from the Finnish Technology Agency. Bayesian inversion was studied as a flexible framework of reconstruction instead of the outdated and rigid tomosynthesis approach. The first results were published in 2003 as the two-part article [413, 288].

Big changes followed on the corporate level. General Electric acquired Instrumentarium in 2003 and later sold the dental imaging division in 2005, resulting in the founding of PaloDEx Group. As a result, the originally quite general 3D imaging project was refocused to dental imaging applications only. Some 3D results concerning mammography were published, though [391]. Interestingly, pretty much the same team of researchers and engineers worked together regardless of the changing names of the parent corporation.

At the time of writing, these research and development projects during 2001–2007 have resulted in one commercial product: the so-called Volumetric Tomography (VT) device designed for dental implant planning.

Dental implants replace missing teeth. Implants are attached by drilling a hole into the bone and screwing a titanium screw to the bone and attaching the implant to the screw. The hole must be deep enough for the implant to be sturdy, but not so deep that it damages nerves located inside the bone. See Figure 9.16 for illustration. Planning the direction, width and depth of an appropriate hole requires three-dimensional information on the tissue. Mere projection X-ray images or panoramic X-ray images are not enough as they involve overlapping of structures and geometric distortions.

The idea of the VT device is to use a digital panoramic imaging device, shown in Figure 9.14, in a new way for collecting limited-angle tomographic

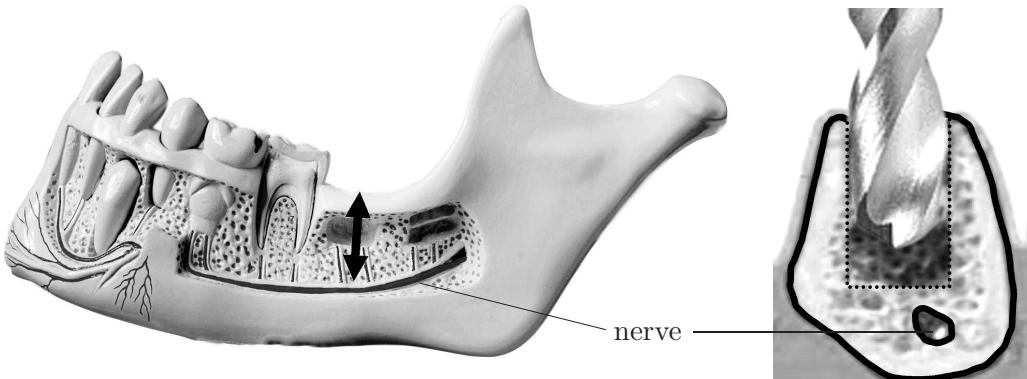


Figure 9.16: Left: typical situation for implant planning. A tooth is missing, and a hole should be drilled as indicated by the black arrow. However, it is important not to drill so deep that the nerve would be damaged. The photograph is taken from a plastic anatomical model. Right: cross-section of the mandible showing the location of the nerve. The appropriate depth of drilling can be measured from a reconstructed cross-sectional slice oriented like this.

data. See Figure 9.17 for an illustration of the imaging geometry. These tailored movements of the panoramic device need to be specially designed because the detector is very narrow (as explained in Section 9.3.3, the detector is designed to mimic the moving film behind a narrow slit collimator). Thus 2D projection images need to be formed by a linear movement, much as in a xerox machine or a flatbed scanner. See Figure 9.18 for projection images of a dry skull.

The limited-angle data set is so incomplete that filtered back-projection type algorithms do not give satisfactory results. Tikhonov regularization, iterative frequency-domain reconstruction and total variation regularization do provide good enough reconstructions, as shown in [289, 290, 69, 217]. See Figure 9.19 for an example of actual *in vivo* imaging using the VT device. The computation is done according to [251].

It is also possible to augment the limited-angle projection data set by including an panoramic image. Now the panoramic image is not a projection image but rather a curved tomosynthetic slice. However, proper modelling of the sharp layer formation allows one to combine information from perpendicular directions, thus adding the most badly missing information to the limited-angle data set. See [217] for reconstructions from such hybrid data sets.

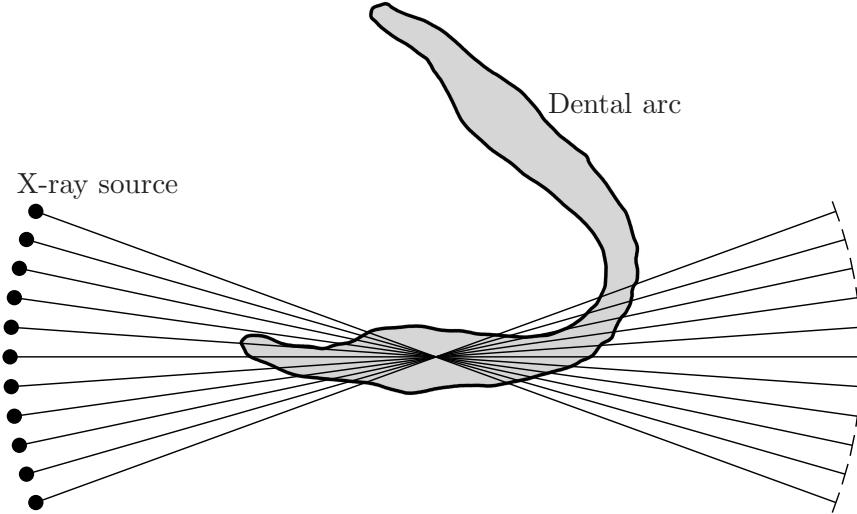


Figure 9.17: Imaging geometry of the VT device. A total of 11 projection images is taken from a ± 20 degree angle of view. See Figure 9.18 for examples of actual projection images.

The sparse-data VT product has the lowest radiation dose among its dense-data competitors, as shown in Table 9.3. The three-dimensional information provided by VT is, nevertheless, enough for the clinical task of dental implant planning. This way a device that the clinic already has can be upgraded by mathematical software to yield 3D images.

Table 9.3: X-radiation patient doses delivered by various devices used for three-dimensional dental imaging. The dose of the VT device is known to be 1–2 doses of a panoramic 2D image, which is roughly $6.7 \mu\text{Sv}$ [108, 477]. The data for the other devices is taken from [322].

Modality	Dose (μSv)
Head CT	2100
CB Mercuray	558
NewTom 3G	59
VT device	13

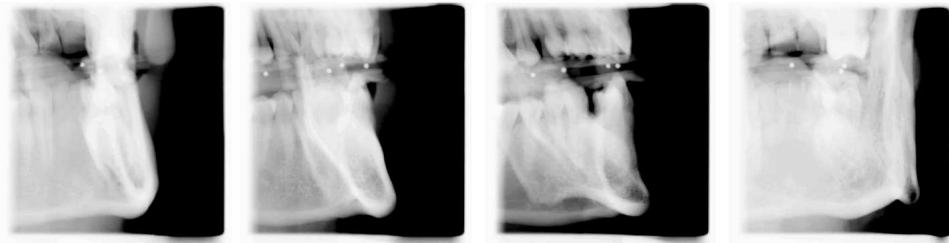


Figure 9.18: Some projection images of a dry scull taken with the VT device. The detector is very narrow in the horizontal direction, so the images are recorded using a scanning movement. Notice the fiducial reference balls used for calibrating the imaging geometry. Images courtesy of PaloDEx Group.

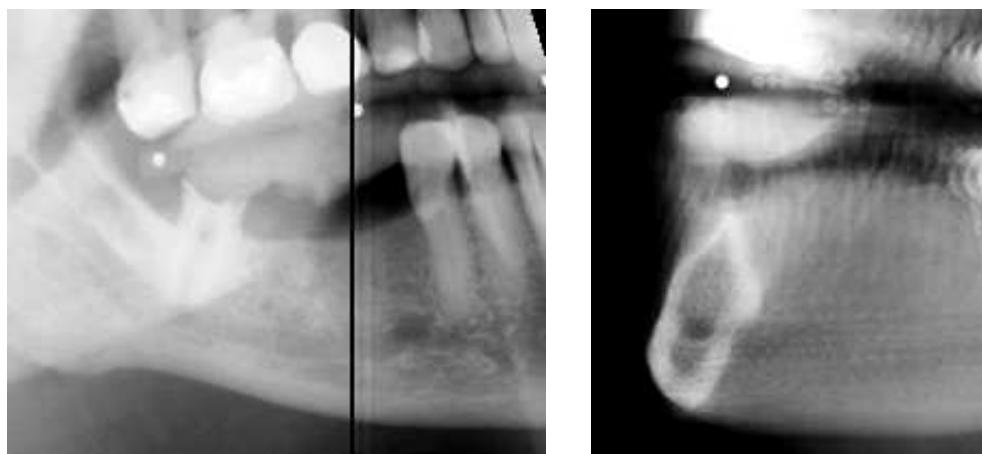


Figure 9.19: Slices from an *in vivo* three-dimensional reconstruction produced by the VT device from a real patient with missing teeth. Left: navigation slice showing the location of the cross-section on the right. Right: cross-section of the mandible with nerve canal clearly visible. Compare to Figure 9.16. Images courtesy of PaloDEx Group.

Chapter 10

Projects

In this section we outline several computational projects related to linear inverse problems. Each project involves building a computational model for a practically relevant direct problem and applying some of the inversion methods developed in Part I of this book. The projects may serve as individual or group assignments in a course, or as starting points of academic theses. Also, successful execution and formal reporting of a project below serves as training for writing scientific articles or engineering reports.

Let us point out some basic principles related to this kind of projects combining theoretical and computational aspects as well as dealing with ill-posedness.

Every piece of software should be tested thoroughly until you can completely trust the results. It's a good idea to start with the simplest possible case (perhaps allowing analytical solution by an explicit formula), write an algorithm, and compare the outcome with an analytical formula or another algorithm which is as different as possible. Then add the slightest possible complication to the code and test it as well. This way you can work your way to the full complexity of the computation by taking as small completely reliable steps as possible.

If you are working with simulated data (as opposed to measuring real data), be sure to avoid an inverse crime. Simulate the data with a different algorithm (or at least with a different grid and slightly perturbed parameters) than what was used in the inversion method, otherwise you may end up with computations that produce wonderfully accurate reconstructions from non-noisy data but break down completely when noise is added. Such algorithms are utterly useless in practice as real data always contains noise. Let us stress once more: noise-robustness is the most important feature of

practical inversion methods.

If your project involves data measured in the real world, start by building both the computational forward model (with accurate simulation of measurement noise) and the inversion algorithm. Then test the inversion thoroughly using simulated data and avoiding any inverse crime. Only then apply the method to real data. The reason for this procedure is efficient debugging. If you go right ahead and try to invert the measured data, chances are that something goes wrong (it always does), but you have no chance of knowing where the problem is: in the forward model, in the inversion algorithm, in the choice of parameters, in the representation of data, or something else. However, if you work your way from the simplest simulated case step by step to the most complicated simulated case, then compare the properties of the measured and simulated data, and only then attempt inversion of the real data, then you always know that the possible problem arises from the latest small addition to the so far reliable computational framework.

10.1 Image deblurring

The aim of this project is to build an algorithm that sharpens misfocused photographs. The data is measured using a digital camera, and consequently the reconstruction methods must be applicable to large-scale data.

The optical construction of regular consumer cameras produces a plane of sharp focus. This means that details located in a two-dimensional plane in the three-dimensional scene under imaging show up sharp and crisp in the photograph, and details away from that plane are blurred. Moreover, the blur becomes worse as the distance grows between a detail and the plane of sharp focus. By “worse blur” we mean wider point spread function.

To keep the project simple, we choose a planar scene so that the point spread function is the same in all parts of the image. Take a printed paper with some text and an image; a newspaper page will do. Make sure that the page contains a small isolated black dot.

Now you need to take a misfocused photograph of the printed paper. The use of a tripod to keep the camera steady is recommended; it’s also a good idea to use two light sources illuminating the paper with roughly 45 degree angles. If you have a camera with manual focus, you can simply focus perfectly on the plane of the paper and then turn the focus a little (or move the camera slightly away from the paper). A simple way to achieve the same result with an autofocus camera is to place the paper on the floor

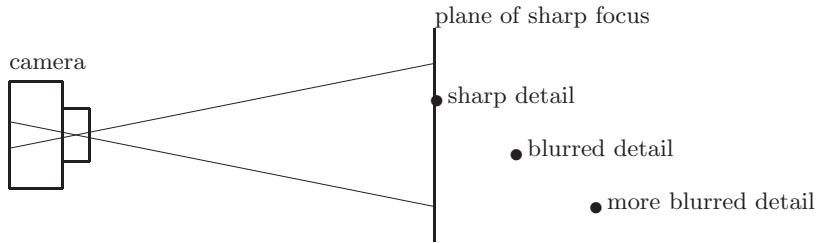


Figure 10.1: Schematic illustration of the focal plane of a camera. Details in the focal plane appear sharp, and details off the focal plane appear blurred. The blurring is stronger if the detail is more far away from the focal plane.

and put a book on top of the paper. (Of course, the book should not cover the important part of the paper.) Let the camera focus on the book cover; then the paper is out of focus. Varying the distance between the camera and the book will adjust the amount of blur.

It's a good idea to take a sharp picture for comparison and a few misfocused versions with different amount of blur. Also, take photographs with different sensitivities (ISO settings in the camera), resulting in various amounts of measurement noise.

Read the photograph into MATLAB using the `imread` command and pick out only one of the three color components (red, green or blue). Then you have a grayscale pixel image that can be viewed as a discrete approximation of a real-valued light intensity distribution on the paper. You can read off the shape of the point spread function from the image of the isolated black dot.

10.2 Inversion of the Laplace transform

Let $f : [0, \infty) \rightarrow \mathbb{R}$. The Laplace transform F of f is defined by

$$F(s) = \int_0^\infty e^{-st} f(t) dt, \quad s \in \mathbb{C}, \quad (10.1)$$

provided that the integral converges. The direct problem is to find the Laplace transform for a given function f according to (10.1). The inverse problem is given a Laplace transform F , find the corresponding function f .

Assume we know the values of F at real points $0 < s_1 < s_2 < \dots < s_n < \infty$. Then we may approximate the integral in (10.1) with, for example, the

trapezoidal rule as

$$\int_0^\infty e^{-st} f(t) dt \approx \frac{t_k}{k} \left(\frac{1}{2} e^{-st_1} f(t_1) + e^{-st_2} f(t_2) + e^{-st_3} f(t_3) + \dots + e^{-st_{k-1}} f(t_{k-1}) + \frac{1}{2} e^{-st_k} f(t_k) \right), \quad (10.2)$$

where the vector $t = [t_1 \ t_2 \ \dots \ t_k]^T \in \mathbb{R}^k$, $0 \leq t_1 < t_2 < \dots < t_k$, contains the points at which the unknown function f will be evaluated. By denoting $x_l = f(t_l)$, $l = 1, \dots, k$, and $m_j = F(s_j)$, $j = 1, \dots, n$, and using (10.2), we get a linear model of the form $m = Ax + \varepsilon$ with

$$A = \frac{t_k}{k} \begin{bmatrix} \frac{1}{2} e^{-s_1 t_1} & e^{-s_1 t_2} & e^{-s_1 t_3} & \dots & e^{-s_1 t_{k-1}} & \frac{1}{2} e^{-s_1 t_k} \\ \frac{1}{2} e^{-s_2 t_1} & e^{-s_2 t_2} & e^{-s_2 t_3} & \dots & e^{-s_2 t_{k-1}} & \frac{1}{2} e^{-s_2 t_k} \\ \vdots & & & & & \vdots \\ \frac{1}{2} e^{-s_n t_1} & e^{-s_n t_2} & e^{-s_n t_3} & \dots & e^{-s_n t_{k-1}} & \frac{1}{2} e^{-s_n t_k} \end{bmatrix}.$$

This problem is addressed in the article *The Bad Truth about the Laplace Transform* by Epstein and Schotland [133].

10.3 Backward parabolic problem

The backward heat equation introduced in Chapter 2 included the simplifying assumptions that the problem is posed in one spatial dimension and that the diffusion coefficient D is constant. Relaxing these assumptions leads to a more versatile model, although it will have more computational complexity. The solution can be approximated with a regularized least-squares approach or by methods of quasi-reversibility [306], in which the problem is transformed into a new problem that is well-posed in the backward direction, but ill-posed in the forward direction. The disadvantage of this approach is that the transformation is achieved by adding the biharmonic operator $\epsilon \Delta^2$ to the elliptic operator, resulting in a fourth-order PDE. A good project is to experiment with either of these approaches on a backward heat equation with one or both of these assumptions relaxed.

A more general model that includes a transport term, a source term, and a decay term is applicable not only to heat transfer [40], but also to contaminant transport in an underground aquifer [39]. In such a model, $u(x, y, t)$ represents the concentration of the contaminant at time t and point (x, y) in the aquifer. Consider the following backward parabolic problem for a two-dimensional convection-diffusion equation.

Find $\phi(x, y)$ such that $u(x, y, t_0) = \phi(x, y)$ is the solution to

$$\begin{aligned} u_t &= (D_1(x, y)u_x)_x + (D_2(x, y)u_y)_y - (v(x, y)u)_x - \lambda(x, y)u + F(x, y) \\ u(x, y, t) &\rightarrow 0 \quad \text{as } x^2 + y^2 \rightarrow \infty \\ u(x, y, T) &= \psi(x, y) > 0, \quad T > t_0 > 0, \quad (x, y) \in \mathbb{R}^2 \end{aligned} \tag{10.3}$$

The initial profile ϕ is assumed to have compact support. The coefficients $D_i(x, y)$, $i = 1, 2$ are dispersion-diffusion coefficients, and the function $v(x, y)$ is the average linear velocity. The functions D_1, D_2 , and v are assumed to be positive piecewise continuous bounded functions in a domain Ω . The function $\lambda > 0$ is the decay rate, and $F(x, y)$ is the source of contaminant concentration per unit mass. The inverse problem is to determine the release history of the groundwater contaminant from knowledge of $\psi(x, y)$. The problem is ill-posed, as was the backward heat equation.

Note that the differential operator on the right-hand side of (10.3) is non-selfadjoint, and due to the presence of the transport term, the solution to the forward problem has a “wave-like” character. As a result, contrary to the self-adjoint case, the computational mesh depends on T , requiring the use of a forward solver with sufficient accuracy and stability for physically realistic time steps.

The continuous dependence of the solution to the forward problem on the initial data ϕ is proved in [199] where a stability estimate on the final time T is also proved that shows the decrease in stability as T increases. Using the stability result the existence of a least-squares solution to the backward parabolic problem is proved. Expressing the unknown initial profile in terms of a finite set of linearly independent basis functions $\xi_m(x, y)$, $m = 1, \dots, M$

$$\phi(x, y) = \sum_{m=1}^M \beta_m \xi_m(x, y)$$

reduces the numerical solution of the backward parabolic problem to a finite number of well-posed forward problems with initial profile ξ_m . Expressing the solution at time T to these forward problems by $u[\xi_m]|_{t=T}$, the minimization functional J can be defined by

$$J(\beta) = \int_{\Omega} \left| \sum_{m=1}^M \beta_m u[\xi_m]|_{t=T} - \psi(x, y) \right|^2 dx dy.$$

The minimum occurs when

$$\frac{\partial}{\partial \beta_k} J(\beta_1, \dots, \beta_M) = 0, \quad k = 1, \dots, M,$$

resulting in an ill-posed system of linear algebraic equations $A\beta = b$ where $A = [a_{km}]$, $b = [b_1, \dots, b_M]^T$, and

$$\begin{aligned} a_{km} &= \int_{\Omega} u[\xi_k]|_{t=T} u[\xi_m]|_{t=T} dx dy \\ b_k &= \int_{\Omega} \psi(x, y) u[\xi_k]|_{t=T} dx dy \end{aligned}$$

A modified alternating-direction finite difference scheme for the solution of the intermediate forward problems is presented in [199], and another approach is the use of a finite element method. The minimization problem can be regularized by the methods of Part I of this book, resulting in many possibilities for a project solving the backward parabolic problem (10.3).

Part II

Nonlinear Inverse Problems

Bibliography

- [1] M. J. ABLOWITZ AND A. S. FOKAS, *Complex variables: introduction and applications*, Cambridge University Press, 2003.
- [2] E. ABOUFADEL AND S. SCHLICKER, *Discovering wavelets*, John Wiley & Sons, 1999.
- [3] A. ABUBAKAR, T. M. HABASHY, M. LI, AND J. LIU, *Inversion algorithms for large-scale geophysical electromagnetic measurements*, Inverse Problems, 25 (2009), p. 123012.
- [4] R. A. ADAMS, *Sobolev Spaces*, Academic Press, 1975.
- [5] G. ALESSANDRINI, *Stable determination of conductivity by boundary measurements*, Applicable Analysis, 27 (1988), pp. 153–172.
- [6] M. ALLAIN AND J. IDIER, *Efficient binary reconstruction for non-destructive evaluation using gammagraphy*, Inverse Problems, 23 (2007), p. 1371.
- [7] M. ALTSCHULER, *Reconstruction of the global-scale three-dimensional solar corona*, in Image Reconstruction from Projections, G. Herman, ed., vol. 32 of Topics in Applied Physics, Springer Berlin / Heidelberg, 1979, pp. 105–145.
- [8] A. ALÙ AND N. ENGHETA, *Achieving transparency with plasmonic and metamaterial coatings*, Phys. Rev. E, 72 (2005), p. 016623.
- [9] W. AMES, *Numerical methods for partial differential equations*, Academic Press, San Diego, 1992.
- [10] H. AMMARI, O. KWON, J. K. SEO, AND E. J. WOO, *Anomaly detection in T-SCAN trans-admittance imaging system*, SIAM Journal on Applied Mathematics, 65 (2004), pp. 252–266.

- [11] S. R. ARRIDGE, *Optical tomography in medical imaging*, Inverse Problems, 15 (1999), pp. R41–R93.
- [12] S. R. ARRIDGE, O. DORN, J. P. KAIPIO, V. KOLEHMAINEN, M. SCHWEIGER, T. TARVAINEN, M. VAUHKONEN, AND A. ZACHAROPOULOS, *Reconstruction of subdomain boundaries of piecewise constant coefficients of the radiative transfer equation from optical tomography data*, Inverse Problems, 22 (2006), pp. 2175–2196.
- [13] M. ASSENHEIMER, O. LAVER-MOSKOVITZ, D. MALONEK, D. MANOR, U. NAHALIEL, R. NITZAN, AND A. SAAD, *The T-SCAN technology: electrical impedance as a diagnostic tool for breast cancer detection*, Physiological measurement, 22 (2001), pp. 1–8.
- [14] K. ASTALA, T. IWANIEC, AND G. MARTIN, *Elliptic partial differential equations and quasiconformal mappings in the plane*, vol. 48 of Princeton Mathematical Series, Princeton University Press, Princeton, NJ, 2009.
- [15] K. ASTALA, M. LASSAS, AND L. PÄIVÄRINTA, *The borderlines of the invisibility and visibility for Calderón’s inverse problem*, arXiv:1109.2749v1, (2012).
- [16] K. ASTALA, J. MUELLER, L. PÄIVÄRINTA, A. PERÄMÄKI, AND S. SILTANEN, *Direct electrical impedance tomography for nonsmooth conductivities*, Inverse Problems and Imaging, 5 (2011), pp. 531–549.
- [17] K. ASTALA, J. MUELLER, L. PÄIVÄRINTA, AND S. SILTANEN, *Numerical computation of complex geometrical optics solutions to the conductivity equation*, Applied and Computational Harmonic Analysis, 29 (2010), pp. 391–403.
- [18] K. ASTALA AND L. PÄIVÄRINTA, *A boundary integral equation for Calderón’s inverse conductivity problem*, in Proc. 7th Internat. Conference on Harmonic Analysis, Collectanea Mathematica, 2006.
- [19] K. ASTALA AND L. PÄIVÄRINTA, *Calderón’s inverse conductivity problem in the plane*, Annals of Mathematics, 163 (2006), pp. 265–299.

- [20] K. ASTALA, L. PÄIVÄRINTA, AND M. LASSAS, *Calderón's inverse problem for anisotropic conductivity in the plane*, Comm. Partial Differential Equations, 30 (2005), pp. 207–224.
- [21] R. ASTER, C. THURBER, AND B. BORCHERS, *Parameter estimation and inverse problems*, vol. 90, Elsevier Academic Press, 2005.
- [22] S. BAAZIZADEH, D. BROOKS, AND D. ISAACSON, *A 3-D boundary element solution to the forward problem of electrical impedance tomography*, in Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE, vol. 1, sept. 2004, pp. 960 –963.
- [23] S. BAAZIZADEH, D. BROOKS, D. ISAACSON, AND J. NEWELL, *Electrode boundary conditions and experimental validation for bem-based EIT forward and inverse solutions*, Medical Imaging, IEEE Transactions on, 25 (2006), pp. 1180 –1188.
- [24] A. P. BAGSHAW, A. D. LISTON, R. H. BAYFORD, A. TIZZARD, A. P. GIBSON, A. TIDSWELL, M. K. SPARKES, H. DEHGHANI, C. D. BINNIE, AND D. S. HOLDER, *Electrical impedance tomography of human brain function using reconstruction algorithms based on the finite element method*, NeuroImage, 20 (2003), pp. 752 – 764.
- [25] J. A. BAKER AND J. Y. LO, *Breast tomosynthesis: State-of-the-art and review of the literature*, Academic Radiology, 18 (2011), pp. 1298 – 1310.
- [26] N. V. BANICHUK AND P. J. NEITTAANMÄKI, *Structural optimization with uncertainties*, vol. 162 of Solid Mechanics and its Applications, Springer, Dordrecht, 2010.
- [27] D. C. BARBER AND B. H. BROWN, *Progress in electrical impedance tomography*, in Inverse problems in partial differential equations (Arcata, CA, 1989), SIAM, Philadelphia, PA, 1990, pp. 151–164.
- [28] J. A. BARCELÓ, T. BARCELÓ, AND A. RUIZ, *Stability of the inverse conductivity problem in the plane for less regular conductivities*, Journal of Differential Equations, 173 (2001), pp. 231–270.
- [29] T. BARCELÓ, D. FARACO, AND A. RUIZ, *Stability of Calderón inverse conductivity problem in the plane*, Journal de Mathématiques Pures et Appliquées, 88 (2007), pp. 522–556.

- [30] D. BAROUDI AND E. SOMERSALO, *Gas temperature mapping using impedance tomography*, Inverse Problems, 13 (1997), p. 1177.
- [31] J. BARZILAI AND J. BORWEIN, *Two-point step size gradient methods*, IMA Journal of Numerical Analysis, 8 (1988), p. 141.
- [32] X. L. BATTLE, G. S. CUNNINGHAM, AND K. M. HANSON, *3D tomographic reconstruction using geometrical models*, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, K. M. Hanson, ed., vol. 3034 of Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Apr. 1997, pp. 346–357.
- [33] X. L. BATTLE, K. M. HANSON, AND G. S. CUNNINGHAM, *Tomographic reconstruction using 3D deformable models*, Physics in Medicine & Biology, 43 (1998), pp. 983–990.
- [34] R. BEALS AND R. R. COIFMAN, *Scattering, transformations spectrales et équations d'évolution non linéaires*, in Goulaouic-Meyer-Schwartz Seminar, 1980–1981, École Polytech., Palaiseau, 1981, pp. Exp. No. XXII,10.
- [35] ———, *Scattering, transformations spectrales et équations d'évolution non linéaire II*, in Goulaouic-Meyer-Schwartz Seminar, 1981–1982, École Polytech., Palaiseau, 1982, pp. Exp. No. XXI, 9.
- [36] ———, *Multidimensional inverse scatterings and nonlinear partial differential equations*, in Pseudodifferential operators and applications (Notre Dame, Ind., 1984), Amer. Math. Soc., Providence, RI, 1985, pp. 45–70.
- [37] ———, *The D-bar approach to inverse scattering and nonlinear evolutions*, Phys. D, 18 (1986), pp. 242–249. Solitons and coherent structures (Santa Barbara, Calif., 1985).
- [38] ———, *Linear spectral problems, nonlinear equations and the $\bar{\partial}$ -method*, Inverse Problems, 5 (1989), pp. 87–130.
- [39] J. BEAR, *Dynamics of Fluids in Porous Media*, Elsevier, New York, 1972.
- [40] J. BECK, B. BLACKWELL, AND S. CLAIR, *Inverse Heat Conduction, Ill-Posed Problems*, Wiley-Interscience, New York, 1985.

- [41] L. BEILINA AND M. KLIBANOV, *Approximate Global Convergence and Adaptivity for Coefficient Inverse Problems*, Springer, 2012.
- [42] M. BENNETT AND R. WILLIAMS, *Monitoring the operation of an oil/water separator using impedance tomography*, Minerals Engineering, 17 (2004), pp. 605 – 614.
- [43] Y. BENVENISTE AND T. MILOH, *Neutral inhomogeneities in conduction phenomena*, Journal of the Mechanics and Physics of Solids, 47 (1999), pp. 1873 – 1892.
- [44] E. BERETTA AND E. FRANCINI, *Lipschitz stability for the electrical impedance tomography problem: the complex case*, ArXiv e-prints, (2010).
- [45] J. BIAN, X. HAN, E. Y. SIDKY, G. CAO, J. LU, O. ZHOU, AND X. PAN, *Investigation of sparse data mouse imaging using micro-CT with a carbon-nanotube-based X-ray source*, Tsinghua Sci Technol, 15 (2010), pp. 74–78.
- [46] J. BIAN, J. H. SIEWERDSEN, X. HAN, E. Y. SIDKY, J. L. PRINCE, C. A. PELIZZARI, AND X. PAN, *Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam ct*, Physics in Medicine and Biology, 55 (2010), p. 6575.
- [47] J. BIKOWSKI, K. KNUDSEN, AND J. L. MUELLER, *Direct numerical reconstruction of conductivities in three dimensions using scattering transforms*, Inverse Problems, 27 (2011).
- [48] J. BIKOWSKI AND J. MUELLER, *2D EIT reconstructions using Calderón’s method*, Inverse Problems and Imaging, 2 (2008), pp. 43–61.
- [49] R. BLUE, *Real-time three-dimensional electrical impedance tomography*, PhD thesis, Rensselaer Polytechnic Institute, Troy, NY, 1997.
- [50] M. BOITI, J. P. LEON, M. MANNA, AND F. PEMPINELLI, *On a spectral transform of a KdV-like equation related to the Schrödinger operator in the plane*, Inverse Problems, 3 (1987), pp. 25–36.
- [51] K. BOONE, A. M. LEWIS, AND D. S. HOLDER, *Imaging of cortical spreading depression by EIT : implications for localization of epileptic foci*, Physiological Measurement, 15 (1994), p. A189.

- [52] L. BORCEA, *Electrical impedance tomography*, Inverse Problems, 18 (2002), pp. 99–136.
- [53] G. BOVERMAN, D. ISAACSON, T.-J. KAO, SAULNIER, G. J., AND J. C. NEWELL, *Methods for direct image reconstruction for EIT in two and three dimensions*, in Proceedings of the 2008 Electrical Impedance Tomography Conference, Dartmouth College, in Hanover, New Hampshire, USA, June 16 to 18 2008.
- [54] G. BOVERMAN, T.-J. KAO, D. ISAACSON, AND G. J. SAULNIER, *An implementation of Calderón’s method for 3-D limited view EIT*, IEEE Transactions on Medical Imaging, 1 (2008), pp. 1–10.
- [55] H. BREZIS, *Functional analysis, Sobolev spaces and partial differential equations*, Springer, New York, 2011.
- [56] B. BROWN, D. BARBER, A. MORICE, AND A. LEATHARD, *Cardiac and respiratory related electrical impedance changes in the human thorax*, IEEE Transactions on Biomedical Engineering, 41 (1994), pp. 729–734.
- [57] B. H. BROWN, *Electrical impedance tomography (EIT): a review*, Journal of Medical Engineering & Technology Number 3, (May/June 2003), pages 9710, 27 (2003), pp. 97–108.
- [58] R. M. BROWN AND G. UHLMANN, *Uniqueness in the inverse conductivity problem for nonsmooth conductivities in two dimensions*, Communications in Partial Differential Equations, 22 (1997), pp. 1009–1027.
- [59] M. BRÜHL AND M. HANKE, *Numerical implementation of two non-iterative methods for locating inclusions by impedance tomography*, Inverse Problems, 16 (2000), pp. 1029–1042.
- [60] J.-F. CAI, S. OSHER, AND Z. SHEN, *Linearized Bregman iterations for frame-based image deblurring*, SIAM J. Imaging Sci., 2 (2009), pp. 226–252.
- [61] F. CAKONI AND D. COLTON, *Qualitative methods in inverse scattering theory*, Interaction of Mechanics and Mathematics, Springer-Verlag, Berlin, 2006. An introduction.

- [62] F. CAKONI, D. COLTON, AND P. MONK, *The linear sampling method in inverse electromagnetic scattering*, vol. 80 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.
- [63] A.-P. CALDERÓN, *On an inverse boundary value problem*, in Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980), Soc. Brasil. Mat., Rio de Janeiro, 1980, pp. 65–73.
- [64] D. CALVETTI, G. LANDI, L. REICHEL, AND F. SGALLARI, *Non-negativity and iterative methods for ill-posed problems*, Inverse Problems, 20 (2004), pp. 1747–1758.
- [65] D. CALVETTI AND E. SOMERSALO, *Priorconditioners for linear systems*, Inverse Problems, 21 (2005), pp. 1397–1418.
- [66] D. CALVETTI AND E. SOMERSALO, *Introduction to Bayesian scientific computing: ten lectures on subjective computing*, vol. 2, Springer, 2007.
- [67] D. CALVETTI AND E. SOMERSALO, *Microlocal sequential regularization in imaging*, Inverse Probl. Imaging, 1 (2007), pp. 1–11.
- [68] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489–509.
- [69] A. CEDERLUND, M. KALKE, AND U. WELANDER, *Volumetric tomography-a new tomographic technique for panoramic units*, Dentomaxillofacial Radiology, 38 (2009), pp. 104–11.
- [70] K. CHADAN, D. COLTON, L. PÄIVÄRINTA, AND W. RUNDELL, *An introduction to inverse scattering and inverse spectral problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. With a foreword by Margaret Cheney.
- [71] A. CHAMBOLLE, S. E. LEVINE, AND B. J. LUCIER, *An upwind finite-difference method for total variation-based image smoothing*, SIAM J. Imaging Sci., 4 (2011), pp. 277–299.
- [72] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

- [73] T. F. CHAN AND K. CHEN, *On a nonlinear multigrid algorithm with primal relaxation for the image total variation minimisation*, Numer. Algorithms, 41 (2006), pp. 387–411.
- [74] T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.
- [75] T. F. CHAN AND J. SHEN, *Image processing and analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. Variational, PDE, wavelet, and stochastic methods.
- [76] C. CHAUX, P. L. COMBETTES, J.-C. PESQUET, AND V. R. WAJS, *A variational formulation for frame-based inverse problems*, Inverse Problems, 23 (2007), pp. 1495–1518.
- [77] M. CHENEY AND D. ISAACSON, *Distinguishability in impedance imaging*, IEEE Transactions on Biomedical Engineering, 39 (1992), pp. 852–860.
- [78] ———, *Issues in electrical impedance imaging*, IEEE Computational Science and Engineering, 2 (1995), pp. 53–62.
- [79] M. CHENEY, D. ISAACSON, J. NEWELL, S. SIMSKE, AND J. GOBLE, *Noser: An algorithm for solving the inverse conductivity problem*, International Journal of Imaging Systems and Technology, 2 (1990), pp. 66–75.
- [80] M. CHENEY, D. ISAACSON, AND J. C. NEWELL, *Electrical impedance tomography*, SIAM Review, 41 (1999), pp. 85–101.
- [81] K. CHENG, D. ISAACSON, J. NEWELL, AND D. GISSER, *Electrode models for electric current computed tomography*, IEEE Transactions on Biomedical Engineering, 36 (1989), pp. 918–924.
- [82] V. CHEREPENIN, A. KARPOV, A. KORJENEVSKY, V. KORNENKO, Y. KULTIASOV, M. OCHAPKIN, O. TROCHANOV, AND J. MEISTER, *Three-dimensional EIT imaging of breast tissues: system design and clinical testing*, IEEE Transactions on Medical Imaging, 21 (2002), pp. 662–667.
- [83] V. CHEREPENIN, A. KARPOV, A. KORJENEVSKY, V. KORNENKO, A. MAZALETSKAYA, D. MAZOUROV, AND D. MEISTER, *A 3D elec-*

- trical impedance tomography (EIT) system for breast cancer detection, *Physiological Measurement*, 22 (2001), pp. 9–18.
- [84] K. CHOI, J. WANG, L. ZHU, T.-S. SUH, S. BOYD, AND L. XING, *Compressed sensing based cone-beam computed tomography reconstruction with a first-order method*, *Medical Physics*, 37 (2010), pp. 5113–5125.
- [85] C. CHUI, *Wavelets: a mathematical tool for signal analysis*, SIAM, Philadelphia, 1997.
- [86] A. CLOP, D. FARACO, AND A. RUIZ, *Integral stability of Calderón inverse conductivity problem in the plane*, *Inverse Problems and Imaging*, 4 (2010), pp. 49–91. Submitted on 2008.
- [87] C. F. COLEMAN AND J. R. McLAUGHLIN, *Solution of the inverse spectral problem for an impedance with integrable derivative. I, II*, *Comm. Pure Appl. Math.*, 46 (1993), pp. 145–184, 185–212.
- [88] D. COLTON AND A. KIRSCH, *An approximation problem in inverse scattering problems theory*, *Applicable Analysis*, 41 (1991), pp. 23–32.
- [89] ——, *A simple method for solving inverse scattering problems in the resonance region*, *Inverse Problems*, 12 (1996), pp. 383–393.
- [90] D. COLTON AND R. KRESS, *Inverse acoustic and electromagnetic scattering theory*, Springer-Verlag, Berlin, second ed., 1998.
- [91] D. COLTON AND P. MONK, *The inverse scattering problem for acoustic waves in an inhomogeneous medium*, *Quart. J. Mech. Appl. Math.*, 41 (1988), pp. 97–125.
- [92] ——, *A new method for solving the inverse scattering problem for acoustic waves in an inhomogeneous medium*, *Inverse Problems*, 5 (1989), pp. 1013–1026.
- [93] ——, *A new method for solving the inverse scattering problem for acoustic waves in an inhomogeneous medium II*, *Inverse Problems*, 6 (1990), pp. 935–947.
- [94] ——, *A comparison of two methods for solving the inverse scattering problem for acoustic waves in an inhomogeneous medium*, *J. Comp. Appl. Math.*, 42 (1992), pp. 5–16.

- [95] ——, *The detection and monitoring of leukemia using electromagnetic waves: Mathematical theory*, Inverse Problems, 10 (1994), pp. 1235–1251.
- [96] ——, *The detection and monitoring of leukemia using electromagnetic waves: Numerical analysis*, Inverse Problems, 11 (1995), pp. 329–342.
- [97] ——, *A linear sampling method for the detection of leukemia using microwaves*, SIAM J. Appl. Math., 58 (1998), pp. 926–941 (electronic).
- [98] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200 (electronic).
- [99] A. M. CORMACK, *Representation of a function by its line integrals, with some radiological applications I*, Journal of Applied Physics, 34 (1963), pp. 2722–2727.
- [100] H. CORNEAN, K. KNUDSEN, AND S. SILTANEN, *Towards a d-bar reconstruction method for three-dimensional EIT*, Journal of Inverse and Ill-Posed Problems, 14 (2006), pp. 111–134.
- [101] E. COSTA, C. CHAVES, S. GOMES, M. BERALDO, M. VOLPE, M. TUCCI, I. SCHETTINO, S. BOHM, C. CARVALHO, H. TANAKA, L. R.G., AND M. AMATO, *Real-time detection of pneumothorax using electrical impedance tomography*, Critical Care Medicine, 36 (2008), pp. 1230–1238.
- [102] E. COSTA, R. LIMA, AND M. AMATO, *Electrical impedance tomography*, Current Opinion in Critical Care, 15 (2009), pp. 18–24.
- [103] J. DAHL, P. C. HANSEN, S. H. JENSEN, AND T. L. JENSEN, *Algorithms and software for total variation image reconstruction via first-order methods*, Numer. Algorithms, 53 (2010), pp. 67–92.
- [104] Y.-H. DAI AND R. FLETCHER, *Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming*, Numerische Mathematik, 100 (2005), pp. 21–47.
- [105] W. DAILY AND A. RAMIREZ, *Electrical resistance tomography during in-situ trichloroethylene remediation at the Savannah River Site*, Journal of Applied Geophysics, 33 (1995), pp. 239–249.

- [106] W. DAILY, A. RAMIREZ, AND R. JOHNSON, *Electrical impedance tomography of a perchloroethylene release*, Journal of Environmental & Engineering Geophysics, 2 (1998), pp. 189–201.
- [107] W. DAILY, A. RAMIREZ, D. LABRECQUE, AND J. NITAO, *Electrical resistivity tomography of vadose water movement*, Water Resources Research, 28 (1992), pp. 1429–1442.
- [108] R. A. DANFORTH AND D. E. CLARK, *Effective dose from radiation absorbed during a panoramic examination with a new generation machine*, Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology, 89 (2000), pp. 236 – 243.
- [109] I. DAUBECHIES, *Orthonormal bases of compactly supported wavelets*, Communications on pure and applied mathematics, 41 (1988), pp. 909–996.
- [110] I. DAUBECHIES, *Ten lectures on wavelets (Ninth printing, 2006)*, vol. 61 of CBMS-NSF Regional conference series in applied mathematics, SIAM, 2006.
- [111] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on pure and applied mathematics, 57 (2004), pp. 1413–1457.
- [112] M. E. DAVISON, *The ill-conditioned nature of the limited-angle tomography problem*, SIAM Journal on Applied Mathematics, 43 (1983), pp. 428–448.
- [113] G. DE MARSILY, *Quantitative Hydrogeology*, Academic Press, 1986.
- [114] M. DEANGELO AND J. L. MUELLER, *2d D-bar reconstructions of human chest and tank data using an improved approximation to the scattering transform*, Physiological Measurement, 31 (2010), pp. 221–232.
- [115] A. H. DELANEY AND Y. BRESLER, *Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography*, IEEE Transactions on Image Processing, 7 (1998), pp. 204–221.

- [116] F. DELBARY, P. C. HANSEN, AND K. KNUDSEN, *Electrical impedance tomography: 3D reconstructions using scattering transforms*, Applicable Analysis, 0 (0), pp. 1–19.
- [117] F. DELBARY, H. P.C., AND K. KNUDSEN, *A direct numerical reconstruction algorithm for the 3D Calderón problem*, in Journal of Physics: Conference Series, vol. 290, IOP Publishing, 2011, p. 0120003.
- [118] J. DENNIS AND R. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliff, NJ, 1983.
- [119] Z. DES PLANTES, *Eine neue Methode zur Differenzierung in der Röntgenographie (Planigraphies)*, Acta Radiologica [Old Series], 13 (1932), pp. 182–192.
- [120] F. DIEKMANN AND U. BICK, *Tomosynthesis and contrast-enhanced digital mammography: recent advances in digital mammography*, European Radiology, 17 (2007), pp. 3086–3092.
- [121] S. DO, W. KARL, M. KALRA, T. BRADY, AND H. PIEN, *A variational approach for reconstructing low dose images in clinical helical ct*, in Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on, april 2010, pp. 784 –787.
- [122] J. T. DOBBINS AND III, *Tomosynthesis imaging: At a translational crossroads*, Medical Physics, 36 (2009), pp. 1956–1967.
- [123] J. T. I. DOBBINS, H. P. MCADAMS, D. J. GODFREY, AND C. M. LI, *Digital tomosynthesis of the chest*, Journal of Thoracic Imaging, 23 (2008), pp. 86–92.
- [124] D. DOBSON AND F. SANTOSA, *Resolution and stability analysis of an inverse problem in electrical impedance tomography: dependence on the input current patterns*, SIAM Journal on Applied Mathematics, 54 (1994), pp. 1542–1560.
- [125] D. DOBSON AND C. VOGEL, *Convergence of an iterative method for total variation denoising*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1779–1791.

- [126] D. C. DOBSON AND F. SANTOSA, *Recovery of blocky images from noisy and blurred data*, SIAM Journal on Applied Mathematics, 56 (1996), pp. 1181–1198.
- [127] D. L. DONOHO, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.
- [128] X. DUAN, L. ZHANG, Y. XING, Z. CHEN, AND J. CHENG, *Few-view projection reconstruction with an iterative reconstruction-reprojection algorithm and TV constraint*, Nuclear Science, IEEE Transactions on, 56 (2009), pp. 1377 –1382.
- [129] P. EDIC, G. SAULNIER, J. NEWELL, AND D. ISAACSON, *A real-time electrical impedance tomograph*, IEEE Transactions on Biomedical Engineering, 42 (1995), pp. 849–859.
- [130] T. EIROLA, M. HUHTANEN, AND J. VON PFALER, *Solution methods for \mathbb{R} -linear problems in \mathbb{C}^n* , SIAM Journal on Matrix Analysis and Applications, 25 (2003), pp. 804–828.
- [131] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of inverse problems*, Kluwer Academic Publishers, 1996.
- [132] C. L. EPSTEIN, *Introduction to the mathematics of medical imaging*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2008.
- [133] C. L. EPSTEIN AND J. SCHOTLAND, *The bad truth about Laplace's transform*, SIAM Rev., 50 (2008), pp. 504–520.
- [134] E. ESSER, X. ZHANG, AND T. F. CHAN, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci., 3 (2010), pp. 1015–1046.
- [135] L. C. EVANS, *Partial differential equations*, American Mathematical Society, Providence, RI, 1998.
- [136] L. FABRIZI, R. YERWORTH, A. MC EWAN, O. GILAD, R. BAYFORD, AND D. S. HOLDER, *A method for removing artefacts from continuous eeg recordings during functional electrical impedance tomography for the detection of epileptic seizures*, Physiological Measurement, 31 (2010), p. S57.

- [137] L. D. FADDEEV, *Increasing solutions of the Schrödinger equation*, Soviet Physics Doklady, 10 (1966), pp. 1033–1035.
- [138] D. FANELLI AND O. KTEM, *Electron tomography: a short overview with an emphasis on the absorption potential model for the forward problem*, Inverse Problems, 24 (2008), p. 013001.
- [139] A. FARIDANI, *Introduction to the mathematics of computed tomography*, in Inside out: inverse problems and applications, vol. 47 of Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, Cambridge, 2003, pp. 1–46.
- [140] H. FENG, W. KARL, AND D. CASTANON, *A curve evolution approach to object-based tomographic reconstruction*, Image Processing, IEEE Transactions on, 12 (2003), pp. 44 – 57.
- [141] X. FENG AND A. PROHL, *Analysis of total variation flow and its finite element approximations*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 533–556.
- [142] A. V. FIACCO AND G. P. MCCORMICK, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, New York, 1968.
- [143] A. S. FOKAS AND M. J. ABLOWITZ, *Method of solution for a class of multidimensional nonlinear evolution equations*, Phys. Rev. Lett., 51 (1983), pp. 7–10.
- [144] ———, *Nonlinear Phenomena*, Springer, Berlin, 1984, ch. The inverse scattering transform for multidimensional 2 + 1 problems.
- [145] M. FORNASIER, A. LANGER, AND C.-B. SCHÖNLIEB, *A convergent overlapping domain decomposition method for total variation minimization*, Numer. Math., 116 (2010), pp. 645–685.
- [146] M. FORNASIER AND C.-B. SCHÖNLIEB, *Subspace correction methods for total variation and l_1 -minimization*, SIAM J. Numer. Anal., 47 (2009), pp. 3397–3428.
- [147] E. FRANCINI, *Recovering a complex coefficient in a planar domain from Dirichlet-to-Neumann map*, Inverse Problems, 16 (2000), pp. 107–119.

- [148] D. FREIMARK, M. ARAD, R. SOKOLOVER, S. ZLOCHIVER, AND S. ABOUD, *Monitoring lung fluid content in chf patients under intravenous diuretics treatment using bio-impedance measurements*, Physiological Measurement, 28 (2007), pp. S269–S277.
- [149] I. FRERICHS, *Electrical impedance tomography (EIT) in applications related to lung and ventilation: a review of experimental and clinical activities*, Physiological Measurement, 21 (2000), pp. R1–R21.
- [150] I. FRERICHS, J. HINZ, P. HERRMANN, G. WEISSER, G. HAHN, T. DUDYKEVYCH, M. QUINTEL, AND G. HELLIGE, *Detection of local lung air content by electrical impedance tomography compared with electron beam ct*, Journal of Applied Physiology, 93 (2002), p. 660.
- [151] I. FRERICHS, S. PULLETTZ, G. ELKE, F. REIFFERSCHEID, D. SCHADLER, J. SCHOLZ, AND W. N., *Assessment of changes in distribution of lung perfusion by electrical impedance tomography*, Respiration, 77 (2009), pp. 282–291.
- [152] I. FRERICHS, G. SCHMITZ, S. PULLETTZ, D. SCHADLER, G. ZICK, J. SCHOLZ, AND N. WEILER, *Reproducibility of regional lung ventilation distribution determined by electrical impedance tomography during mechanical ventilation*, Physiological Measurement, 28 (2007), pp. S261–S267.
- [153] H. U. FREY, S. FREY, D. LARSON, T. NYGRÉN, AND J. SEMETER, *Science Closure and Enabling Technologies for Constellation Class Missions*, University of California Printing Services, University of California at Berkeley and NASA, Goddard Space Flight Center, 1998, ch. Tomographic methods for magnetospheric applications, pp. 72–77.
- [154] L. FUKS, M. CHENEY, D. ISAACSON, D. GISSE, AND J. NEWELL, *Detection and imaging of electric conductivity and permittivity at low frequency*, IEEE Transactions on Biomedical Engineering, 38 (1991), pp. 1106–1110.
- [155] I. M. GELFAND AND B. M. LEVITAN, *On identification of the differential expression via its spectral function*, Izv. Akad. Nauk SSSR (in russian), Sek. Matem, 15 (1951), pp. 309–360.
- [156] A. P. GIBSON, J. C. HEBDEN, AND S. R. ARRIDGE, *Recent advances in diffuse optical imaging*, Physics in Medicine and Biology, 50 (2005), pp. R1–R43.

- [157] D. GILBARG AND N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, Springer, 1977.
- [158] D. GISSER, D. ISAACSON, AND J. NEWELL, *Electric current computed tomography and eigenvalues*, SIAM Journal on Applied Mathematics, 50 (1990), pp. 1623–1634.
- [159] G. GLADWELL, *Inverse problems in vibration*, Martinus Hjhoff, Dordrecht, 1986.
- [160] G. GLADWELL, S. DODS, AND S. CHAUDHURI, *Nonuniform transmission-line synthesis using inverse eigenvalue analysis*, IEEE Trans. Circuits and Systems, 35 (1988), pp. 659–665.
- [161] J. GOBLE, M. CHENEY, AND D. ISAACSON, *Electrical impedance tomography in three dimensions*, The Applied Computational Electromagnetics Society Journal, 7 (1992), pp. 128–147.
- [162] D. GOLDFARB AND W. YIN, *Second-order cone programming methods for total variation-based image restoration*, SIAM J. Sci. Comput., 27 (2005), pp. 622–645 (electronic).
- [163] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
- [164] G. H. GOLUB, M. HEATH, AND G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.
- [165] T. GOMI AND H. HIRANO, *Clinical potential of digital linear tomosynthesis imaging of total joint arthroplasty*, Journal of Digital Imaging, 21 (2008), pp. 312–322. 10.1007/s10278-007-9040-9.
- [166] C. GORDON, D. L. WEBB, AND S. WOLPERT, *One cannot hear the shape of a drum*, Bulletin of the American Mathematical Society, 27 (1992), p. 134138.
- [167] J. GOTTLIEB AND P. DUChATEAU, *Parameter identification and inverse problems in hydrology, geology, and ecology*, Kluwer, 1996.
- [168] D. GRANT, *Tomosynthesis: A three-dimensional radiographic imaging technique*, IEEE Transactions on Biomedical Engineering, 19 (1972), pp. 20–28.

- [169] M. GRASMAIR, M. HALTMEIER, AND O. SCHERZER., *Sparse regularization with l^q penalty term*, Inverse Problems, 24 (2008), p. 055020 (13pp).
- [170] M. GRASMAIR, M. HALTMEIER, AND O. SCHERZER, *Necessary and sufficient conditions for linear convergence of l^1 -regularization*, Communications on Pure and Applied Mathematics, 64 (2011), pp. 161–182.
- [171] A. GREENLEAF, Y. KURYLEV, M. LASSAS, AND G. UHLMANN, *Full-wave invisibility of active devices at all frequencies*, Communications in Mathematical Physics, 275 (2007), pp. 749–789.
- [172] ——, *Cloaking devices, electromagnetic wormholes, and transformation optics*, SIAM Review, 51 (2009), pp. 3–33.
- [173] A. GREENLEAF, Y. KURYLEV, M. LASSAS, AND G. UHLMANN, *Invisibility and inverse problems*, Bulletin of the American Mathematical Society, 46 (2009), pp. 55–79.
- [174] A. GREENLEAF, M. LASSAS, AND G. UHLMANN, *Anisotropic conductivities that cannot be detected by EIT*, Physiological Measurement, 24 (2003), p. 413.
- [175] ——, *The Calderón problem for conormal potentials i: Global uniqueness and reconstruction*, Communications on pure and applied mathematics, 56 (2003), pp. 328–352.
- [176] ——, *On nonuniqueness for Calderón’s inverse problem*, Mathematical Research Letters, 10 (2003), pp. 685–694.
- [177] A. GREENLEAF AND G. UHLMANN, *Nonlocal inversion formulas for the X-ray transform*, Duke Math. J., 58 (1989), pp. 205–240.
- [178] P. G. GRINEVICH AND R. G. NOVIKOV, *Transparent potentials at fixed energy in dimension two. Fixed-energy dispersion relations for the fast decaying potentials*, Communications in Mathematical Physics, 174 (1995), pp. 409–446.
- [179] R. A. GROENHUIS, R. L. WEBBER, AND U. E. RUTTIMANN, *Computerized tomosynthesis of dental tissues*, Oral Surgery, Oral Medicine, Oral Pathology, 56 (1983), pp. 206 – 214.

- [180] R. GUENTHER AND J. W. LEE, *Partial Differential Equations of Mathematical Physics and Integral Equations*, Dover, 1988.
- [181] G. HAHN, I. SIPINKOVA, F. BAISCH, AND G. HELLIGE, *Changes in the thoracic impedance distribution under different ventilatory conditions*, Physiological Measurement, 16 (1995), p. A161.
- [182] E. T. HALE, W. YIN, AND Y. ZHANG, *Fixed-point continuation for l_1 -minimization: methodology and convergence*, SIAM J. Optim., 19 (2008), pp. 1107–1130.
- [183] ———, *Fixed-point continuation applied to compressed sensing: implementation and numerical experiments*, J. Comput. Math., 28 (2010), pp. 170–194.
- [184] D. HALLIKAINEN, *History of panoramic radiography*, Acta Radiologica, 37 (1996), pp. 441–445.
- [185] K. HÄMÄLÄINEN, A. KALLONEN, V. KOLEHMAINEN, M. LASSAS, E. NIEMI, K. NIINIMÄKI, AND S. SILTANEN, *Total variation regularized tomography with automatic parameter choice*, Unpublished manuscript, (2012).
- [186] K. HÄMÄLÄINEN, A. KALLONEN, V. KOLEHMAINEN, M. LASSAS, K. NIINIMÄKI, AND S. SILTANEN, *Sparsity-promoting multiresolution tomography*, Unpublished manuscript, (2012).
- [187] U. HÄMARIK, R. PALM, AND T. RAUS, *A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level*, Journal of Computational and Applied Mathematics, (2011).
- [188] S. HAMILTON, C. HERRERA, J. L. MUELLER, AND A. VON HERRMANN, *A direct D-bar reconstruction algorithm for recovering a complex conductivity in 2-D*, Arxiv preprint, (2012).
- [189] M. HANKE, *Limitations of the l-curve method in ill-posed problems*, BIT, 36 (1996), pp. 287–301.
- [190] M. HANKE, *Iterative regularization techniques in image reconstruction*, in Surveys on solution methods for inverse problems, Springer, Vienna, 2000, pp. 35–52.

- [191] M. HANKE AND M. BRÜHL, *Recent progress in electrical impedance tomography*, Inverse Problems, 19 (2003), pp. S65–S90. Special section on imaging.
- [192] M. HANKE, J. NAGY, AND R. PLEMMONS, *Preconditioned iterative regularization for ill-posed problems*, in Numerical linear algebra (Kent, OH, 1992), de Gruyter, Berlin, 1993, pp. 141–163.
- [193] P. C. HANSEN, *Rank-deficient and discrete ill-posed problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Numerical aspects of linear inversion.
- [194] ———, *Discrete inverse problems*, vol. 7 of Fundamentals of Algorithms, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. Insight and algorithms.
- [195] P. C. HANSEN, J. G. NAGY, AND D. P. O’LEARY, *Deblurring images*, vol. 3 of Fundamentals of Algorithms, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Matrices, spectra, and filtering.
- [196] K. HANSON, G. CUNNINGHAM, J. JENNINGS, G.R., AND D. WOLF, *Tomographic reconstruction based on flexible geometric models*, in Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, vol. 2, nov 1994, pp. 145 –147 vol.2.
- [197] K. M. HANSON, G. S. CUNNINGHAM, AND R. J. MCKEE, *Uncertainties in tomographic reconstructions based on deformable models*, in Medical Imaging: Image processing. Proceedings of SPIE, K. M. Hanson, ed., vol. 3034, 1997, pp. 276–286.
- [198] B. HARRACH AND J. K. SEO, *Detecting inclusions in electrical impedance tomography without reference measurements*, SIAM J. Appl. Math., 69 (2009), pp. 1662–1681.
- [199] A. HASANOV AND J. MUELLER, *A numerical method for backward parabolic problems with non-selfadjoint elliptic operators*, Applied Numerical Mathematics, 37 (2001), pp. 55–78.
- [200] J. HASLINGER AND P. NEITTAANMÄKI, *Finite element approximation for optimal shape, material and topology design*, John Wiley & Sons Ltd., Chichester, second ed., 1996.

- [201] A. HAUPTMANN, K. HÄMÄLÄINEN, A. KALLONEN, E. NIEMI, AND S. SILTANEN, *Total variation regularization for X-ray tomography*, (2012).
- [202] J. C. HEBDEN, A. GIBSON, T. AUSTIN, R. M. YUSOF, N. EVERDELL, D. T. DELPY, S. R. ARRIDGE, J. H. MEEK, AND J. S. WYATT, *Imaging changes in blood volume and oxygenation in the newborn infant brain using three-dimensional optical tomography*, Physics in Medicine and Biology, 49 (2004), p. 1117.
- [203] L. M. HEIKKINEN, J. KOURUNEN, T. SAVOLAINEN, P. J. VAUHKONEN, J. P. KAIPIO, AND M. VAUHKONEN, *Real time three-dimensional electrical impedance tomography applied in multiphase flow imaging*, Measurement Science and Technology, 17 (2006), p. 2083.
- [204] J. HEINO AND E. SOMERSALO, *Estimation of optical absorption in anisotropic background*, Inverse Problems, 18 (2002), pp. 559–573.
- [205] G. T. HERMAN AND R. DAVIDI, *Image reconstruction from a small number of projections*, Inverse Problems, 24 (2008), p. 045011.
- [206] G. T. HERMAN AND A. KUBA, *Discrete Tomography: Foundations, Algorithms, and Applications*, Birkhäuser, 1999.
- [207] ——, *Advances in discrete tomography and its applications*, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2007.
- [208] E. M. C. HILLMAN, J. C. HEBDEN, M. SCHWEIGER, H. DEHGHANI, F. E. W. SCHMIDT, D. T. DELPY, AND S. R. ARRIDGE, *Time resolved optical tomography of the human forearm*, Physics in Medicine and Biology, 46 (2001), p. 1117.
- [209] M. HINTERMÜLLER AND K. KUNISCH, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311–1333.
- [210] S.-E. HJELT, *Pragmatic Inversion of Geophysical data*, no. 39 in Lecture Notes in Earth Sciences, Springer-Verlag, 1992.
- [211] H. HOCHSTADT, *On determination of the density of a vibrating string from spectral data*, J. Math. Anal. Appl., 55 (1976), pp. 673–685.

- [212] B. HOFMANN, *Regularization for Applied Inverse and Ill-Posed Problems*, Teubner, Stuttgart, 1986.
- [213] D. HOLDER, *Clinical and physiological applications of electrical impedance tomography*, UCL Press, London, 1993.
- [214] D. HOLDER, ed., *Electrical impedance tomography; methods, history, and applications*, IOP publishing Ltd., 2005.
- [215] P. HUA, J. WOO, J. WEBSTER, AND W. TOMPKINS, *Finite element modeling of electrode-skin contact impedance in electrical impedance tomography*, IEEE Transactions on Biomedical Engineering, 40 (1993), pp. 335–343.
- [216] M. HUHTANEN AND A. PERÄMÄKI, *Numerical solution of the R-linear Beltrami equation*, Mathematics of Computation, 81 (2012), pp. 387–397.
- [217] N. HYVÖNEN, M. KALKE, M. LASSAS, H. SETÄLÄ, AND S. SILTANEN, *Three-dimensional dental X-ray imaging by combination of panoramic and projection data*, Inverse Probl. Imaging, 4 (2010), pp. 257–271.
- [218] T. IDE, H. ISOZAKI, S. NAKATA, AND S. SILTANEN, *Local detection of three-dimensional inclusions in electrical impedance tomography*, Inverse problems, 26 (2010), p. 035001.
- [219] T. IDE, H. ISOZAKI, S. NAKATA, S. SILTANEN, AND G. UHLMANN, *Probing for electrical inclusions with complex spherical waves*, Communications on pure and applied mathematics, 60 (2007), pp. 1415–1442.
- [220] M. IKEHATA, *Reconstruction of the support function for inclusion from boundary measurements*, Journal of Inverse and Ill-Posed Problems, 8 (2000), pp. 367–378.
- [221] ——, *A regularized extraction formula in the enclosure method*, Inverse problems, 18 (2002), p. 435.
- [222] ——, *Mittag-Leffler's function and extracting from Cauchy data*, in Inverse Problems and Spectral Theory: Proceedings of the Workshop on Spectral Theory of Differential Operators and Inverse Problems, October 28–November 1, 2002, Research Institute for, vol. 348, Amer Mathematical Society, 2004, p. 41.

- [223] M. IKEHATA, E. NIEMI, AND S. SILTANEN, *Inverse obstacle scattering with limited-aperture data*, Inverse Problems and Imaging, 6 (2012), pp. 77–94.
- [224] M. IKEHATA AND T. OHE, *A numerical method for finding the convex hull of inclusions using the enclosure method*, in Electromagnetic Nondestructive Evaluation (VI), IOS Press, 2002.
- [225] M. IKEHATA AND T. OHE, *The enclosure method for an inverse crack problem and the MittagLeffler function*, Inverse Problems, 24 (2008), p. 015006.
- [226] M. IKEHATA AND S. SILTANEN, *Numerical method for finding the convex hull of an inclusion in conductivity from boundary measurements*, Inverse Problems, 16 (2000), pp. 1043–1052.
- [227] M. IKEHATA AND S. SILTANEN, *Electrical impedance tomography and Mittag-Leffler's function*, Inverse Problems, 20 (2004), pp. 1325–1348.
- [228] ——, *Numerical solution of the Cauchy problem for the stationary Schrödinger equation using faddeev's green function*, SIAM Journal on Applied Mathematics, (2004), pp. 1907–1932.
- [229] D. ISAACSON, *Distinguishability of conductivities by electric current computed tomography*, IEEE Transactions on Medical Imaging, 5 (1986), pp. 91–95.
- [230] D. ISAACSON AND M. CHENEY, *Current problems in impedance imaging*, in Inverse Problems in Partial Differential, D. Colton, R. Ewing, and W. Rundell, eds., SIAM, 1990, pp. 139–148.
- [231] ——, *Effects of measurement precision and finite numbers of electrodes on linear impedance imaging algorithms*, SIAM Journal on Applied Mathematics, 51 (1991), pp. 1705–1731.
- [232] D. ISAACSON, J. MUELLER, J. NEWELL, AND S. SILTANEN, *Imaging cardiac activity by the D-bar method for electrical impedance tomography*, Physiological Measurement, 27 (2006), pp. S43–S50.
- [233] D. ISAACSON, J. L. MUELLER, J. C. NEWELL, AND S. SILTANEN, *Reconstructions of chest phantoms by the D-bar method for electrical impedance tomography*, IEEE Transactions on Medical Imaging, 23 (2004), pp. 821–828.

- [234] E. ISAACSON AND E. TRUBOWITZ, *The inverse Sturm-Liouville problem I*, IEEE Transactions on Medical Imaging, 5 (1986), pp. 91–95.
- [235] V. ISAKOV, *Inverse Problems for Partial Differential Equations*, Springer, 1998.
- [236] T. JENSEN, J. JØRGENSEN, P. HANSEN, AND S. JENSEN, *Implementation of an optimal first-order method for strongly convex total variation regularization*, BIT Numerical Mathematics, (2011), pp. 1–28.
- [237] X. JIA, B. DONG, Y. LOU, AND S. B. JIANG, *GPU-based iterative cone-beam CT reconstruction using tight frame regularization*, Physics in Medicine and Biology, 56 (2011), p. 3787.
- [238] X. JIA, Y. LOU, R. LI, W. Y. SONG, AND S. B. JIANG, *GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation*, Medical Physics, 37 (2010), pp. 1757–1760.
- [239] F. JOHN, *Numerical solution of the equation of heat conduction for preceding times*, Annals of Mathematics, 40 (1955), pp. 129–142.
- [240] J. JORDANA, M. GASULLA, AND R. PALLÀS-ARENY, *Electrical resistance tomography to detect leaks from buried pipes*, Measurement Science and Technology, 12 (2001), pp. 1061–1068.
- [241] J. JOSSINET, *Variability of impeditivity in normal and pathological breast tissue*, Medical and Biological Engineering and Computing, 34 (1996), pp. 346–350.
- [242] ——, *The impeditivity of freshly excised human breast tissue*, Physiological Measurement, 19 (1998), pp. 61–75.
- [243] M. KAC, *Can one hear the shape of a drum?*, The American Mathematical Monthly, 73 (1966), pp. 1–23.
- [244] J. KAIPIO, A. SEPPÄNEN, E. SOMERSALO, AND H. HAARIO, *Statistical inversion approach for optimizing current patterns in EIT*, in Proceedings of 3rd World Congress on Industrial Process Tomography, 2003, pp. 683–688.
- [245] ——, *Posterior covariance related optimal current patterns in electrical impedance tomography*, Inverse Problems, 20 (2004), pp. 919–936.

- [246] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, vol. 160 of Applied Mathematical Sciences, Springer Verlag, 2004.
- [247] G. KAISER, *A friendly guide to Wavelets*, Birkhäuser, Boston, 1994.
- [248] A. C. KAK AND M. SLANEY, *Principles of computerized tomographic imaging*, IEEE Press, New York, 1988.
- [249] T. KAKO AND K. TOUDA, *Numerical approximation of Dirichlet-to-Neumann mapping and its application to voice generation problem*, in Domain decomposition methods in science and engineering, vol. 40 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 2005, pp. 51–65.
- [250] ———, *Numerical method for voice generation problem based on finite element method*, Journal of Computational Acoustics, 14 (2006), pp. 45–56.
- [251] M. KALKE, *Method for limited angle tomography*, 2010.
- [252] H. KALLASJOKI, K. PALOMÄKI, C. MAGI, P. ALKU, AND M. KURIMO, *Noise robust LVCSR feature extraction based on stabilized weighted linear prediction*, in International Conference on Speech and Computer, St. Petersburg, Russia, June 21-25, 2009, 2009.
- [253] B. KALTENBACHER, A. NEUBAUER, AND O. SCHERZER, *Iterative regularization methods for nonlinear ill-posed problems*, vol. 6, de Gruyter, 2008.
- [254] T.-J. KAO, G. BOVERMAN, B. S. KIM, D. ISAACSON, G. SAULNIER, J. NEWELL, M. CHOI, R. MOORE, AND D. KOPANS, *Regional admittivity spectra with tomosynthesis images for breast cancer detection: Preliminary patient study*, Medical Imaging, IEEE Transactions on, 27 (2008), pp. 1762 –1768.
- [255] T.-J. KAO, G. SAULNIER, H. XIA, C. TAMMA, J. NEWELL, AND D. ISAACSON, *A compensated radiolucent electrode array for combined EIT and mammography*, Physiological Measurement, 28 (2007), p. S291.
- [256] K. KARHUNEN, A. SEPPÄNEN, A. LEHIKOINEN, P. J. M. MONTEIRO, AND J. P. KAIPIO, *Electrical resistance tomography imaging of concrete*, Cement and Concrete Research, 40 (2010), pp. 137–145.

- [257] A. KATCHALOV, Y. KURYLEV, AND M. LASSAS, *Inverse boundary spectral problems*, vol. 123 of Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics, Chapman & Hall/CRC, Boca Raton, FL, 2001.
- [258] A. KATSEVICH, *A general scheme for constructing inversion algorithms for cone beam CT*, Int. J. Math. Math. Sci., (2003), pp. 1305–1321.
- [259] A. KATSEVICH AND M. KAPRALOV, *Filtered backprojection inversion of the cone beam transform for a general class of curves*, SIAM Journal on Applied Mathematics, 68 (2007), pp. 334–353.
- [260] A. KATSEVICH AND G. LAURITSCH, *Filtered backprojection algorithms for spiral cone beam CT*, in Sampling, wavelets, and tomography, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2004, pp. 255–287.
- [261] F. KEINERT, *Wavelets and Multiwavelets*, Chapman & Hall, 2004.
- [262] C. KELLEY, *Iterative methods for optimization*, Society for Industrial Mathematics, 1999.
- [263] A. KEMNA, A. BINLEY, A. RAMIREZ, AND W. DAILY, *Complex resistivity tomography for environmental applications*, Chemical Engineering Journal, 77 (2000), pp. 11–18.
- [264] A. KEMNA, B. KULESSA, AND H. VEREECKEN, *Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models*, Journal of Hydrology, 267 (2002), pp. 125 – 146.
- [265] M. KERKER, *Invisible bodies*, J. Opt. Soc. Am., 65 (1975), pp. 376–379.
- [266] T. KERNER, K. PAULSEN, A. HARTOV, S. SOHO, AND S. POPLACK, *Electrical impedance spectroscopy of the breast: clinical imaging results in 26 subjects*, IEEE Transactions on Medical Imaging, 21 (2002), pp. 638–645.
- [267] B. KIM, D. ISAACSON, H. XIA, T. KAO, J. NEWELL, AND G. SAULNIER, *A method for analyzing electrical impedance spectroscopy data from breast cancer patients*, Physiological measurement, 28 (2007), pp. S237–S246.

- [268] A. KIRSCH, *An introduction to the mathematical theory of inverse problems*, Springer-Verlag, second ed., 2011.
- [269] A. KIRSCH AND N. GRINBERG, *The factorization method for inverse problems*, Oxford University Press, USA, 2008.
- [270] E. KLANN, R. RAMLAU, AND L. REICHEL, *Wavelet-based multilevel methods for linear ill-posed problems*, BIT, 51 (2011), pp. 669–694.
- [271] E. KLANN, R. RAMLAU, AND W. RING, *A Mumford-Shah level-set approach for the inversion and segmentation of SPECT/CT data*, Inverse Probl. Imaging, 5 (2011), pp. 137–166.
- [272] D. KLOTZ, K. SEILER, H. MOSER, AND F. NEUMAIER, *Dispersivity and velocity relationship from laboratory and field experiments*, J. Hydrology, 45 (1980), pp. 169–184.
- [273] K. KNUDSEN, *A new direct method for reconstructing isotropic conductivities in the plane*, Physiological Measurement, 24 (2003), pp. 391–403.
- [274] K. KNUDSEN, M. LASSAS, J. MUELLER, AND S. SILTANEN, *D-bar method for electrical impedance tomography with discontinuous conductivities*, SIAM Journal on Applied Mathematics, 67 (2007), p. 893.
- [275] K. KNUDSEN, M. LASSAS, J. MUELLER, AND S. SILTANEN, *Reconstructions of piecewise constant conductivities by the D-bar method for electrical impedance tomography*, in Journal of Physics: Conference Series, vol. 124, IOP Publishing, 2008, p. 012029.
- [276] K. KNUDSEN, M. LASSAS, J. MUELLER, AND S. SILTANEN, *Regularized D-bar method for the inverse conductivity problem*, Inverse Problems and Imaging, 3 (2009), pp. 599–624.
- [277] K. KNUDSEN AND J. MUELLER, *The born approximation and Calderón’s method for reconstructions of conductivities in 3-D*, Discrete and Continuous Dynamical Systems, (2011), pp. 884–893.
- [278] K. KNUDSEN, J. MUELLER, AND S. SILTANEN, *Numerical solution method for the dbar-equation in the plane*, Journal of Computational Physics, 198 (2004), pp. 500–517.
- [279] K. KNUDSEN AND A. TAMASAN, *Reconstruction of less regular conductivities in the plane*, Communications in Partial Differential Equations, 29 (2004), pp. 361–381.

- [280] R. KOHN, H. SHEN, M. VOGELIUS, AND M. WEINSTEIN, *Cloaking via change of variables in electric impedance tomography*, Inverse Problems, 24 (2008), p. 015016.
- [281] R. V. KOHN AND M. VOGELIUS, *Identification of an unknown conductivity by means of measurements at the boundary*, in Inverse problems (New York, 1983), vol. 14 of SIAM-AMS Proc., Amer. Math. Soc., Providence, RI, 1984, pp. 113–123.
- [282] V. KOLEHMAINEN, M. LASSAS, K. NIINIMÄKI, AND S. SILTANEN, *Sparsity-promoting Bayesian inversion*, Inverse Problems, 28 (2012), p. 025005.
- [283] V. KOLEHMAINEN, M. LASSAS, AND P. OLA, *The inverse conductivity problem with an imperfectly known boundary*, SIAM Journal on Applied Mathematics, 66 (2005), pp. 365–383 (electronic).
- [284] ——, *Electrical impedance tomography problem with inaccurately known boundary and contact impedances*, Medical Imaging, IEEE Transactions on, 27 (2008), pp. 1404 –1414.
- [285] V. KOLEHMAINEN, M. LASSAS, AND P. OLA, *Calderón’s inverse problem with an imperfectly known boundary and reconstruction up to a conformal deformation*, SIAM J. Math. Anal., 42 (2010), pp. 1371–1381.
- [286] V. KOLEHMAINEN, M. LASSAS, P. OLA, AND S. SILTANEN, *Recovering boundary shape and conductivity in electrical impedance tomography*, Inverse Problems and Imaging, (2012).
- [287] V. KOLEHMAINEN, M. LASSAS, AND S. SILTANEN, *Limited data X-ray tomography using nonlinear evolution equations*, SIAM Journal on Scientific Computing, 30 (2008), pp. 1413–1429.
- [288] V. KOLEHMAINEN, S. SILTANEN, S. JÄRVENPÄÄ, J. KAIPIO, P. KOISTINEN, M. LASSAS, J. PIRTTILÄ, AND E. SOMERSALO, *Statistical inversion for medical X-ray tomography with few radiographs: II. application to dental radiology*, Physics in Medicine and Biology, 48 (2003), pp. 1465–1490.
- [289] V. KOLEHMAINEN, A. VANNE, S. SILTANEN, S. JÄRVENPÄÄ, J. KAIPIO, M. LASSAS, AND M. KALKE, *Parallelized Bayesian inversion for three-dimensional dental X-ray imaging*, IEEE Transactions on Medical Imaging, 25 (2006), pp. 218–228.

- [290] ———, *Bayesian inversion method for 3D dental X-ray imaging*, Elektrotechnik und Informationstechnik, 124 (2007), pp. 248–253.
- [291] R. KRESS, *On the numerical solution of a hypersingular integral equation in scattering theory*, J. Comput. Appl. Math., 61 (1995), pp. 345–360.
- [292] T. KUBO, P.-J. P. LIN, W. STILLER, M. TAKAHASHI, H.-U. KAUCZOR, Y. OHNO, AND H. HATABU, *Radiation dose reduction in chest CT: A review*, American Journal of Roentgenology, 190 (2008), pp. 335–343.
- [293] P. KUNST, A. NOORDEGRAAF, E. RAAIJMAKERS, J. BAKKER, A. GROENEVELD, P. POSTMUS, AND P. DE VRIES, *Electrical impedance tomography in the assessment of extravascular lung water in noncardiogenic acute respiratory failure**, Chest, 116 (1999), pp. 1695–1702.
- [294] J. KYBIC, T. BLU, AND M. A. UNSER, *Variational approach to tomographic reconstruction*, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, M. Sonka & K. M. Hanson, ed., vol. 4322 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, July 2001, pp. 30–39.
- [295] E. KYRÖLÄ, J. TAMMINEN, G. LEPPELMEIER, V. SOFIEVA, S. HAS-SINEN, J. BERTAUX, A. HAUCHECORNE, F. DALAUDIER, C. COT, O. KORABLEV, O. F. D'ANDON, G. BARROT, A. MANGIN, B. THÉODORE, M. GUIREL, F. ETANCHAUD, P. SNOEIJ, R. KOOPMAN, L. SAAVEDRA, R. FRAISSE, D. FUSSEN, AND F. VANHELLEMONT, *Gomos on envisat: an overview*, Advances in Space Research, 33 (2004), pp. 1020 – 1028.
- [296] H. LANGET, C. RIDDELL, Y. TROUSSET, A. TENENHAUS, E. LA-HALLE, G. FLEURY, AND N. PARAGIOS, *Compressed sensing based 3D tomographic reconstruction for rotational angiography*, in Medical Image Computing and Computer-Assisted Intervention MICCAI 2011, G. Fichtinger, A. Martel, and T. Peters, eds., vol. 6891 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2011, pp. 97–104.
- [297] M. L. LAPIDUS, *Can one hear the shape of a fractal drum? partial resolution of the weylberry conjecture*, in Geometric analysis and com-

- puter graphics (Berkeley, CA, 1988), vol. 17, Math. Sci. Res. Inst. Publ. (New York: Springer), 1991, p. 119126.
- [298] M. LASSAS, M. MATAICH, S. SILTANEN, AND E. SOMERSALO, *Wind velocity observation with a CW doppler radar*, Geoscience and Remote Sensing, IEEE Transactions on, 40 (2002), pp. 2427 – 2437.
 - [299] M. LASSAS, J. L. MUELLER, AND S. SILTANEN, *Mapping properties of the nonlinear Fourier transform in dimension two*, Communications in Partial Differential Equations, 32 (2007), pp. 591–610.
 - [300] M. LASSAS, J. L. MUELLER, S. SILTANEN, AND A. STAHEL, *The Novikov-Veselov Equation and the Inverse Scattering Method, Part I: Analysis*, ArXiv e-prints, (2011).
 - [301] ———, *The Novikov-Veselov Equation and the Inverse Scattering Method, Part II: Computation*, (2011).
 - [302] M. LASSAS, E. SAKSMAN, AND S. SILTANEN, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Problems and Imaging, 3 (2009), pp. 87–122.
 - [303] M. LASSAS AND S. SILTANEN, *Can one use total variation prior for edge-preserving Bayesian inversion?*, Inverse Problems, 20 (2004), pp. 1537–1564.
 - [304] M. LASSAS, M. TAYLOR, AND G. UHLMANN, *The Dirichlet-to-Neumann map for complete riemannian manifolds with boundary*, Communications in Analysis and Geometry, 11 (2003), pp. 207–222.
 - [305] M. LASSAS AND G. UHLMANN, *On determining a Riemannian manifold from the Dirichlet-to-Neumann map*, Ann. Sci. École Norm. Sup. (4), 34 (2001), pp. 771–787.
 - [306] R. LATTES AND J.-L. LIONS, *The Method of Quasi-Reversibility, Applications to Partial Differential Equations*, Elsevier, New York, 1969.
 - [307] A. LECHLEITER, *A regularization technique for the factorization method*, Inverse problems, 22 (2006), p. 1605.
 - [308] ———, *Factorization methods for photonics and rough surfaces*, KIT Scientific Publishing, 2008.

- [309] J. LEE AND G. UHLMANN, *Determining anisotropic real-analytic conductivities by boundary measurements*, Communications on Pure and Applied Mathematics, 42 (1989), pp. 1097–1112.
- [310] U. LEONHARDT, *Optical conformal mapping*, Science, 312 (2006), pp. 1777–1780.
- [311] Y. LI AND F. SANTOSA, *A computational algorithm for minimizing total variation in image restoration*, Image Processing, IEEE Transactions on, 5 (1996), pp. 987 –995.
- [312] H. LIAO, *A gradually unmasking method for limited data tomography*, in Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on, april 2007, pp. 820 –823.
- [313] H. LIAO AND G. SAPIRO, *Sparse representations for limited data tomography*, in Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on, may 2008, pp. 1375 – 1378.
- [314] L. H. LIEU AND L. A. VESE, *Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces*, Appl. Math. Optim., 58 (2008), pp. 167–193.
- [315] A. LINNINGER, S. BASATI, R. DAWE, AND R. PENN, *An impedance sensor to monitor and control cerebral ventricular volume*, Medical Engineering & Physics, 31 (2009), pp. 838 – 845.
- [316] L. LIU, *Stability Estimates for the Two-Dimensional Inverse Conductivity Problem*, PhD thesis, University of Rochester, 1997.
- [317] B. F. LOGAN, *The uncertainty principle in reconstructing functions from projections*, Duke Math. J., 42 (1975), pp. 661–706.
- [318] A. K. LOUIS, *Nonuniqueness in inverse Radon problems: the frequency distribution of the ghosts*, Math. Z., 185 (1984), pp. 429–440.
- [319] ——, *Orthogonal function series expansions and the null space of the Radon transform*, SIAM J. Math. Anal., 15 (1984), pp. 621–633.
- [320] A. K. LOUIS, *Incomplete data problems in X-ray computerized tomography I. singular value decomposition of the limited angle transform*, Numerische Mathematik, 48 (1986), pp. 251–262.

- [321] B. LOWE AND W. RUNDELL, *An inverse problem for a Sturm-Liouville operator*, Journal of Mathematical Analysis and Applications, 181 (1994), pp. 188–199.
- [322] J. LUDLOW, L. DAVIES-LUDLOW, S. BROOKS, AND W. HOWERTON, *Dosimetry of 3 CBCT devices for oral and maxillofacial radiology: CB Mercuray, NewTom 3G and i-CAT*, Dentomaxillofacial Radiology, 35 (2006), pp. 219–226.
- [323] D. G. LUENBERGER, *Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Inc., 2 ed., 1989.
- [324] P. MAASS, *The X-ray transform: singular value decomposition and resolution*, Inverse Problems, 3 (1987), pp. 729–741.
- [325] A. MALICH, T. FRITSCH, R. ANDERSON, T. BOEHM, M. G. FREESMEYER, M. FLECK, AND W. A. KAISER, *Electrical impedance scanning for classifying suspicious breast lesions: first results*, European Radiology, 10 (2000), pp. 1555–1561.
- [326] C. H. MCCOLLOUGH, A. N. PRIMAK, N. BRAUN, J. KOFLER, L. YU, AND J. CHRISTNER, *Strategies for reducing radiation dose in ct*, Radiol Clin North Am, 47 (2009), pp. 27–40.
- [327] J. R. MC LAUGHLIN, *Inverse spectral theory using nodal points as data—a uniqueness result*, J. Differential Equations, 73 (1988), pp. 354–362.
- [328] S. MEHROTRA, *On the implementation of a primal-dual interior point method*, SIAM Journal on Optimization, 2 (1992), pp. 575–601.
- [329] W. MENKE, *Geophysical data analysis: discrete inverse theory*, International geophysics series, Academic Press, 1989.
- [330] P. METHERALL, D. C. BARBER, AND R. H. SMALLWOOD, *Three dimensional electrical impedance tomography*, in Proc IX Int Conf Electrical Bio-Impedance, Heidelberg, Germany, 1995.
- [331] P. METHERALL, D. C. BARBER, R. H. SMALLWOOD, AND B. H. BROWN, *Three-dimensional electrical impedance tomography*, Nature, 380 (1996), pp. 509–512.
- [332] P. METHERALL, R. H. SMALLWOOD, AND D. C. BARBER, *Three dimensional electrical impedance tomography of the human thorax*, in Proc 18th Int Conf IEEE Eng Med Biol Society, 1996.

- [333] Y. MEYER, *Wavelets and operators*, vol. 1, Cambridge University Press, Cambridge, UK, 1995.
- [334] G. W. MILTON AND N.-A. P. NICOROVICI, *On the cloaking effects associated with anomalous localized resonance*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 462 (2006), pp. 3027–3059.
- [335] A. MOHAMMAD-DJAFARI AND K. SAUER, *Shape reconstruction in X-ray tomography from a small number of projections using deformable models*, in The 17th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt97), Boise, Idaho, USA, August 1997.
- [336] A. MOHAMMAD-DJAFARI AND C. SOUSSEN, *Discrete Tomography — Foundations, Algorithms and Applications*, Birkhäuser, 1999, ch. Reconstruction of Compact Homogeneous 3D Objects from their Projections, pp. 317–342. ISBN 0-8176-4101-7.
- [337] P. MONK, *Finite element methods for Maxwell's equations*, Oxford University Press, New York, 2003.
- [338] V. A. MOROZOV, *Methods for solving incorrectly posed problems*, Springer-Verlag, New York, 1984. Translated from the Russian by A. B. Aries, Translation edited by Z. Nashed.
- [339] J. P. MORUCCI, M. GRANIE, M. LEI, M. CHABERT, AND P. M. MARSILLI, *3D reconstruction in electrical impedance imaging using a direct sensitivity matrix approach*, Physiological Measurement, 16 (1995), pp. A123–A128.
- [340] J. MUELLER, D. ISAACSON, AND J. C. NEWELL, *A reconstruction algorithm for electrical impedance tomography data collected on rectangular electrode arrays*, IEEE Transactions on Biomedical Engineering, 49 (1999), pp. 1379–1386.
- [341] J. MUELLER AND S. SILTANEN, *Direct reconstructions of conductivities from boundary measurements*, SIAM Journal on Scientific Computing, 24 (2003), pp. 1232–1266.
- [342] J. L. MUELLER, D. ISAACSON, AND J. C. NEWELL, *Reconstruction of conductivity changes due to ventilation and perfusion from EIT data collected on a rectangular electrode array*, Physiological Measurement, 22 (2001), pp. 97–106.

- [343] E. MURPHY, *2-D D-bar Conductivity Reconstructions on Non-circular Domains*, PhD thesis, Colorado State University, Fort Collins, CO, 2007.
- [344] E. K. MURPHY AND J. L. MUELLER, *Effect of domain-shape modeling and measurement errors on the 2-d D-bar method for electrical impedance tomography*, IEEE Transactions on Medical Imaging, 28 (2009), pp. 1576–1584.
- [345] E. K. MURPHY, J. L. MUELLER, AND J. C. NEWELL, *Reconstructions of conductive and insulating targets using the D-bar method on an elliptical domain*, Physiological Measurement, 28 (2007), pp. S101–S144.
- [346] A. I. NACHMAN, *Reconstructions from boundary measurements*, Annals of Mathematics, 128 (1988), pp. 531–576.
- [347] A. I. NACHMAN, *Global uniqueness for a two-dimensional inverse boundary value problem*, University of Rochester, Dept. of Mathematics Preprint Series, (1993).
- [348] A. I. NACHMAN, *Global uniqueness for a two-dimensional inverse boundary value problem*, Annals of Mathematics, 143 (1996), pp. 71–96.
- [349] G. NAKAMURA, P. RONKANEN, S. S., AND T. K., *Recovering conductivity at the boundary in three-dimensional electrical impedance tomography*, Inverse Problems and Imaging, 5 (2011), pp. 485–510.
- [350] G. NAKAMURA, S. SILTANEN, K. TANUMA, AND S. WANG, *Numerical recovery of conductivity at the boundary from the localized Dirichlet to Neumann map*, Computing, 75 (2005), pp. 197–213.
- [351] F. NATTERER, *The mathematics of computerized tomography*, vol. 32 of SIAM Classics in Applied Mathematics, SIAM, 2001.
- [352] F. NATTERER AND F. WÜBBELING, *Mathematical Methods in Image Reconstruction*, Monographs on Mathematical Modeling and Computation, SIAM, 2001.
- [353] P. NEITTAANMÄKI, M. RUDNICKI, AND A. SAVINI, *Inverse Problems and Optimal Design in Electricity and Magnetism*, Monographs in Electrical and Electronic Engineering, Clarendon Press, 1996.

- [354] Y. NESTEROV, *Barrier subgradient method*, Math. Program., 127 (2011), pp. 31–56.
- [355] J. NEWELL, R. BLUE, D. ISAACSON, G. SAULNIER, AND A. ROSS, *Phasic three-dimensional impedance imaging of cardiac activity*, Physiological Measurement, 23 (2002), pp. 203–209.
- [356] R. NEWTON, *Inverse Schrödinger scattering in three dimensions*, Springer, 1989.
- [357] K. NIINIMÄKI, S. SILTANEN, AND V. KOLEHMAINEN, *Bayesian multiresolution method for local tomography in dental X-ray imaging*, Physics in Medicine and Biology, 52 (2007), pp. 6663–6678.
- [358] L. T. NIKLASON, B. T. CHRISTIAN, L. E. NIKLASON, D. B. KOPANS, D. E. CASTLEBERRY, B. H. OPSAHL-ONG, C. E. LANDBERG, P. J. SLANETZ, A. A. GIARDINO, R. MOORE, D. ALBAGLI, M. C. DEJULE, P. F. FITZGERALD, D. F. FOBARE, B. W. GIAMBATTISTA, R. F. KWASNICK, J. LIU, S. J. LUBOWSKI, G. E. POSSIN, J. F. RICHOTTE, C. Y. WEI, AND R. F. WIRTH, *Digital tomosynthesis in breast imaging.*, Radiology, 205 (1997), pp. 399–406.
- [359] A. NISSINEN, L. M. HEIKKINEN, AND J. P. KAIPIO, *The Bayesian approximation error approach for electrical impedance tomography—experimental results*, Measurement Science and Technology, 19 (2008).
- [360] A. NISSINEN, L. M. HEIKKINEN, V. KOLEHMAINEN, AND J. P. KAIPIO, *Compensation of errors due to discretization, domain truncation and unknown contact impedances in electrical impedance tomography*, Measurement Science and Technology, 20 (2009).
- [361] A. NISSINEN, V. KOLEHMAINEN, AND J. P. KAIPIO, *Compensation of errors due to incorrect model geometry in electrical impedance tomography*, Journal of Physics: Conference Series, 224 (2010).
- [362] ———, *Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography*, IEEE Transaction on Medical Imaging, 30 (2011), pp. 231–242.
- [363] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer Series in Operations Research, Springer Verlag, New York, 2 ed., 2006.

- [364] R. NOVIKOV, *A multidimensional inverse spectral problem for the equation $-\delta\psi + (v(x) - eu(x))\psi = 0$* , Functional Analysis and Its Applications, 22 (1988), pp. 263–272.
- [365] S. P. NOVIKOV AND A. P. VESELOV, *Two-dimensional Schrödinger operator: inverse scattering transform and evolutional equations*, Physica D, 18 (1986), pp. 267–273.
- [366] H. NUMATA, *Consideration of the parabolic radiography of the dental arch*, J. Shimazu Stud., 10 (1933), pp. 13–21.
- [367] S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489 (electronic).
- [368] S. OSHER AND R. FEDKIW, *Level set methods and dynamic implicit surfaces*, vol. 153 of Applied Mathematical Sciences, Springer-Verlag, New York, 2003.
- [369] S. OSHER, Y. MAO, B. DONG, AND W. YIN, *Fast linearized Bregman iteration for compressive sensing and sparse denoising*, Commun. Math. Sci., 8 (2010), pp. 93–111.
- [370] S. OSHER AND J. A. SETHIAN, *Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations*, Journal of Computational Physics, 79 (1988), pp. 12–49.
- [371] Y. PAATERO, *Suunnittelemastani uudesta hampaiden röntgenkuvaustekniikasta*, Suomen Hammaslääkäriseuran Toimituksia, 86 (1946), p. 37.
- [372] ——, *A new tomographical method for radiographing curved outer surfaces*, Acta Radiologica, 32 (1949), pp. 177–184.
- [373] X. PAN, E. Y. SIDKY, AND M. VANNIER, *Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?*, Inverse Problems, 25 (2009), p. 123009.
- [374] J. B. PENDRY, D. SCHURIG, AND D. R. SMITH, *Controlling electromagnetic fields*, Science, 312 (2006), pp. 1780–1782.
- [375] M. PERSSON, D. BONE, AND H. ELMQVIST, *Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography*, Physics in Medicine and Biology, 46 (2001), pp. 853–866.

- [376] T. PHAM, M. YUILL, C. DAKIN, AND A. SCHIBLER, *Regional ventilation distribution in the first 6 months of life*, European Respiratory Journal, 37 (2011), pp. 919–924.
- [377] C. POPA AND R. ZDUNEK, *Kaczmarz extended algorithm for tomographic image reconstruction from limited-data*, Math. Comput. Simul., 65 (2004), pp. 579–598.
- [378] J. PÖSCHEL AND E. TRUBOWITZ, *Inverse spectral theory*, Academic Press, New York, 1987.
- [379] R. POTTHAST, *Point sources and multipoles in inverse scattering theory*, CRC Press, 2001.
- [380] S. PRÖSSDORF AND J. SARANEN, *A fully discrete approximation method for the exterior Neumann problem of the Helmholtz equation*, Z. Anal. Anwendungen, 13 (1994), pp. 683–695.
- [381] E. QUINTO, *Singularities of the X-ray transform and limited data tomography in \mathbb{R}^2 and \mathbb{R}^3* , SIAM Journal on Mathematical Analysis, 24 (1993), pp. 1215–1215.
- [382] ———, *Local algorithms in exterior tomography*, Journal of Computational and Applied Mathematics, 199 (2007), pp. 141 – 148.
- [383] E. T. QUINTO, *Exterior and limited-angle tomography in non-destructive evaluation*, Inverse Problems, 14 (1998), pp. 339–353.
- [384] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996.
- [385] T. RAITIO, A. SUNI, J. YAMAGISHI, H. PULAKKA, J. NURMINEN, M. VAINIO, AND P. ALKU, *HMM-based speech synthesis utilizing glottal inverse filtering*, IEEE Transactions on Audio, Speech, and Language Processing, 19 (2011), pp. 153–165.
- [386] A. RAMIREZ, W. DAILY, D. LABREQUE, E. OWEN, AND D. CHESNUT, *Monitoring an underground steam injection process using electrical resistance tomography*, Water Resources Research, 29 (1993), pp. 73–88.
- [387] A. RAMIREZ, W. DAILY, D. LABREQUE, AND D. ROELANT, *Detection of leaks in underground storage tanks using electrical resistance method*, Journal of Environmental and Engineering Geophysics, 1 (1996), pp. 189–203.

- [388] R. RAMLAU AND W. RING, *A Mumford-Shah level-set approach for the inversion and segmentation of X-ray tomography data*, Journal of Computational Physics, 221 (2007), pp. 539–557.
- [389] A. RAMM, *Recovery of the potential from fixed energy scattering data*, Inverse Problems, 4 (1988), pp. 877–886.
- [390] A. RAMM, *Multidimensional Inverse Scattering Problems*, Longman-Wiley, New York, 1992.
- [391] M. RANTALA, S. VÄNSKÄ, S. JÄRVENPÄÄ, M. KALKE, M. LAS-SAS, J. MOBERG, AND S. SILTANEN, *Wavelet-based reconstruction for limited-angle X-ray tomography*, Medical Imaging, IEEE Transactions on, 25 (2006), pp. 210–217.
- [392] M. RAYDAN, *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM J. Optim., 7 (1997), pp. 26–33.
- [393] M. REED AND B. SIMON, *Methods of modern mathematical physics. Vol I: Functional analysis*, Academic Press Inc., New York, 1980.
- [394] C. REID, M. BETCKE, D. CHANA, AND R. SPELLER, *The development of a pseudo-3D imaging system (tomosynthesis) for security screening of passenger baggage*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 652 (2011), pp. 108 – 111.
- [395] M. RENARDY AND R. C. ROGERS, *An introduction to partial differential equations*, vol. 13 of Texts in Applied Mathematics, Springer-Verlag., New York, 1993.
- [396] M. ROBINI, A. LACHAL, AND I. MAGNIN, *A stochastic continuation approach to piecewise constant reconstruction*, Image Processing, IEEE Transactions on, 16 (2007), pp. 2576 –2589.
- [397] H. L. ROYDEN, *Real analysis*, Macmillan Publishing Company, New York, third ed., 1988.
- [398] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.
- [399] W. RUDIN, *Functional Analysis*, Tata McGraw-Hill, 1974.

- [400] A. R. RUUSKANEN, A. SEPPÄNEN, S. DUNCAN, E. SOMERSALO, AND J. P. KAIPIO, *Using process tomography as a sensor for optimal control*, Appl. Numer. Math., 56 (2006), pp. 37–54.
- [401] E. SAKSMAN, T. NYGRÉN, AND M. MARKKANEN, *Ionospheric structures invisible in satellite radiotomography*, Radio Science, 32 (1997), pp. 605–616.
- [402] J. C. SANTAMARINA AND D. FRATTA, *Discrete signals and inverse problems*, Wiley, 2005.
- [403] J. SARANEN AND G. VAINIKKO, *Periodic Integral and Pseudodifferential Equations with Numerical Approximation*, Springer, 2002.
- [404] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, AND F. LENZEN, *Variational methods in imaging*, vol. 167 of Applied Mathematical Sciences, Springer Verlag, 2009.
- [405] D. SCHURIG, J. J. MOCK, B. J. JUSTICE, S. A. CUMMER, J. B. PENDRY, A. F. STARR, AND D. R. SMITH, *Metamaterial electromagnetic cloak at microwave frequencies*, Science, 314 (2006), pp. 977–980.
- [406] T. SCHUSTER, *The method of approximate inverse: theory and applications*, vol. 1906 of Lecture Notes in Mathematics, Springer, Berlin, 2007.
- [407] A. SEPPÄNEN, M. VAUHKONEN, P. J. VAUHKONEN, E. SOMERSALO, AND J. P. KAIPIO, *State estimation with fluid dynamical evolution models in process tomography—an application to impedance tomography*, Inverse Problems, 17 (2001), pp. 467–483.
- [408] K. SHUNG, M. SMITH, AND B. TSUI, *Principles of Medical Imaging*, Academic Press, San Diego, 1992.
- [409] E. SIDKY, C. KAO, AND X. PAN, *Effect of the data constraint on few-view, fan-beam CT image reconstruction by TV minimization*, in Nuclear Science Symposium Conference Record, 2006. IEEE, vol. 4, IEEE, 2006, pp. 2296–2298.
- [410] E. Y. SIDKY, C.-M. KAO, AND P. XIAOCHUAN, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct*, Journal of X-Ray Science and Technology, 14 (2006), pp. 119–139.

- [411] E. Y. SIDKY AND X. PAN, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Physics in Medicine and Biology, 53 (2008), p. 4777.
- [412] S. SILTANEN, *Electrical impedance tomography and Faddeev Green's functions*, Annales Academiae Scientiarum Fennicae Mathematica Dissertationes, 121 (1999), p. 56. Dissertation, Helsinki University of Technology, Espoo, 1999.
- [413] S. SILTANEN, V. KOLEHMAINEN, S. JÄRVENPÄÄ, J. P. KAIPIO, P. KOISTINEN, M. LASSAS, J. PIRTILÄ, AND E. SOMERSALO, *Statistical inversion for medical X-ray tomography with few radiographs: I. general theory*, Physics in medicine and biology, 48 (2003), pp. 1437–1463.
- [414] S. SILTANEN, J. MUELLER, AND D. ISAACSON, *An implementation of the reconstruction algorithm of A. Nachman for the 2-D inverse conductivity problem*, Inverse Problems, 16 (2000), pp. 681–699.
- [415] S. SILTANEN AND J. P. TAMMINEN, *Reconstructing conductivities with boundary corrected D-bar method*, Journal of Inverse and Ill-posed Problems, (2012).
- [416] V. SINGH, L. MUKHERJEE, P. DINU, J. XU, AND K. HOFFMANN, *Limited view CT reconstruction and segmentation via constrained metric labeling*, Computer Vision and Image Understanding, 112 (2008), pp. 67–80.
- [417] H. SMIT, A. VONK NOORDEGRAAF, J. MARCUS, A. BOONSTRA, P. VRIES, AND P. POSTMUS, *Determinants of pulmonary perfusion measured by electrical impedance tomography*, European journal of applied physiology, 92 (2004), pp. 45–49.
- [418] K. SMITH AND F. KEINERT, *Mathematical foundations of computed tomography*, Applied Optics, 24 (1985), pp. 3950–3957.
- [419] K. T. SMITH, D. C. SOLMON, AND S. L. WAGNER, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bulletin of the AMS, 83 (1977), pp. 1227–1270.
- [420] F. SMITHIES, *The eigenvalues and singular values of integral equations*, Proc. London Math. Soc., (1937), pp. 255–279.

- [421] E. SOMERSALO, M. CHENEY, AND D. ISAACSON, *Existence and uniqueness for electrode models for electric current computed tomography*, SIAM Journal on Applied Mathematics, 52 (1992), pp. 1023–1040.
- [422] M. SONHDI, *A survey of the vocal tract inverse problem: Theory, computations and experiments*, in Medical Imaging: Image processing. Proceedings of SPIE, F. Santosa, Y.-H. Pao, W. Symes, and C. Holland, eds., SIAM, 1984, pp. 1–19.
- [423] C. SOUSSEN AND J. IDIER, *Reconstruction of three-dimensional localized objects from limited angle X-ray projections: an approach based on sparsity and multigrid image representation*, Journal of Electronic Imaging, 17 (2008), p. 033011.
- [424] C. SOUSSEN AND A. MOHAMMAD-DJAFARI, *Polygonal and polyhedral contour reconstruction in computed tomography*, Image Processing, IEEE Transactions on, 13 (2004), pp. 1507 –1523.
- [425] S. STEFANESCO, C. SCHLUMBERGER, AND M. SCHLUMBERGER, *Sur la distribution électrique potentielle autour d'une prise de terre ponctuelle dans un terrain à couches horizontales, homogènes et isotropes*, J. Physics and Radium Ser., 7 (1930), pp. 132–140.
- [426] W. STRANG AND G. FIX, *Analysis of the finite element method*, Prentice-Hall series in automatic computation, Prentice-Hall, New Jersey, 1973.
- [427] N. SUN, *Mathematical Modeling of Groundwater Pollution*, Springer, New York, 1996.
- [428] Z. SUN AND G. UHLMANN, *Anisotropic inverse problems in two dimensions*, Inverse Problems, 19 (2003), pp. 1001–1010.
- [429] J. SYLVESTER, *An anisotropic inverse boundary value problem*, Communications on Pure and Applied Mathematics, 43 (1990), pp. 201–232.
- [430] J. SYLVESTER AND G. UHLMANN, *A global uniqueness theorem for an inverse boundary value problem*, Annals of Mathematics, 125 (1987), pp. 153–169.
- [431] K. C. TAM AND V. PEREZ-MENDEZ, *Tomographical imaging with limited-angle input*, J. Opt. Soc. Amer., 71 (1981), pp. 582–592.

- [432] J. TANG, B. E. NETT, AND G.-H. CHEN, *Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms*, Physics in Medicine and Biology, 54 (2009), p. 5781.
- [433] A. TARANTOLA, *Inverse problem theory and methods for model parameter estimation*, Society for Industrial Mathematics, 2005.
- [434] T. TARVAINEN, M. VAUHKONEN, V. KOLEHMAINEN, AND J. P. KAIPIO, *Finite element model for the coupled radiative transfer equation and diffusion approximation*, Internat. J. Numer. Methods Engrg., 65 (2006), pp. 383–405.
- [435] Z. TIAN, X. JIA, K. YUAN, T. PAN, AND S. B. JIANG, *Low-dose CT reconstruction via edge-preserving total variation regularization*, Physics in Medicine and Biology, 56 (2011), p. 5949.
- [436] R. TIBSHIRANI, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.
- [437] T. TIDSWELL, A. GIBSON, R. BAYFORD, AND H. D.S., *Three-dimensional electrical impedance tomography of human brain activity*, NeuroImage, 13 (2001), pp. 283–294.
- [438] A. N. TIKHONOV AND V. Y. ARSENIN, *Solutions of ill-posed problems*, V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.
- [439] Y. TING LIN, A. ORTEGA, AND A. DIMAKIS, *Sparse recovery for discrete tomography*, in Image Processing (ICIP), 2010 17th IEEE International Conference on, sept. 2010, pp. 4181 –4184.
- [440] A. TINGBERG, *X-ray tomosynthesis: a review of its use for breast and chest imaging*, Radiat. Prot. Dosimetry, 139 (2010), pp. 100–107.
- [441] H. TRIEBEL, *Function spaces and wavelets on domains*, vol. 7 of Tracts in Mathematics, European Mathematical Society, 2008.
- [442] T. TSAI, *The Schrödinger operator in the plane*, Inverse Problems, 9 (1993), pp. 763–787.
- [443] T. TSAI, *The associated evolution equations of the Schrödinger operator in the plane*, Inverse Problems, 10 (1994), p. 1419.

- [444] V. TSAPAKI, J. E. ALDRICH, R. SHARMA, M. A. STANISZEWSKA, A. KRISANACHINDA, M. REHANI, A. HUFTON, C. TRIANTOPOULOU, P. N. MANIATIS, J. PAPAILIOU, AND M. PROKOP, *Dose reduction in CT while maintaining diagnostic confidence: Diagnostic reference levels at routine head, chest, and abdominal CT—IAEA-coordinated research project*, Radiology, 240 (September 2006), pp. 828–834.
- [445] G. UHLMANN, ed., *Inside out: inverse problems and applications*, vol. 47 of Mathematical Sciences Research Institute Publications, Cambridge University Press, Cambridge, 2003.
- [446] G. UHLMANN, *Electrical impedance tomography and Calderón’s problem*, Inverse Problems, 25 (2009), p. 123011.
- [447] G. UHLMANN AND J.-N. WANG, *Reconstructing discontinuities using complex geometrical optics solutions*, SIAM J. Appl. Math., 68 (2008), pp. 1026–1044.
- [448] G. VAINIKKO, *Multidimensional weakly singular integral equations*, vol. 1549 of Lecture Notes in Mathematics, Springer, 1993.
- [449] G. VAINIKKO, *Fast solvers of the Lippmann-Schwinger equation*, in Direct and inverse problems of mathematical physics (Newark, DE, 1997), vol. 5 of Int. Soc. Anal. Appl. Comput., Kluwer Acad. Publ., Dordrecht, 2000, pp. 423–440.
- [450] M. T. VAN GENUCHTEN AND J. PARKER, *Boundary conditions for displacement experiments through short laboratory soil columns*, Soil Sci. Soc. Am. J., 48 (1984), pp. 703–708.
- [451] R. J. VANDERBEI, *Linear programming: Foundations and Extensions*, vol. 114 of International Series in Operations Research & Management Science, Springer Verlag, New York, 3 ed., 2008.
- [452] S. VÄNSKÄ, M. LASSAS, AND S. SILTANEN, *Statistical X-ray tomography using empirical Besov priors*, International Journal of Tomography & Statistics, 11 (2009), pp. 3–32.
- [453] P. J. VAUHKONEN, *Image Reconstruction in Three-Dimensional Electrical Impedance Tomography*, PhD thesis, University of Kuopio, 2005.
- [454] P. J. VAUHKONEN, M. VAUHKONEN, T. SAVOLAINEN, AND J. P. KAIPIO, *Static three-dimensional electrical impedance tomography*, Annals of the New York Academy of Sciences, 873 (1999), pp. 472–481.

- [455] ——, *Three-dimensional electrical impedance tomography based on the complete electrode model*, IEEE Transactions on Biomedical Engineering, 46 (1999), pp. 1150–1160.
- [456] I. N. VEKUA, *Generalized Analytic Functions*, Pergamon Press, 1962.
- [457] L. A. VESE AND S. J. OSHER, *Image denoising and decomposition with total variation minimization and oscillatory functions*, J. Math. Imaging Vision, 20 (2004), pp. 7–18. Special issue on mathematics and image analysis.
- [458] V. G. VESELAGO, *The electrodynamics of substances with simultaneously negative values of ϵ and μ* , Soviet Physics Uspekhi, 10 (1968), p. 509.
- [459] A. P. VESELOV AND S. P. NOVIKOV, *Finite-zone, two-dimensional, potential Schrödinger operators, explicit formulas and evolution equations*, Sov. Math. Dokl, 30 (1984), pp. 558–591.
- [460] J. VICTORINO, J. BORGES, V. OKAMOTO, G. MATOS, M. TUCCI, M. CARAMEZ, H. TANAKA, F. SIPMANN, D. SANTOS, C. BARBAS, ET AL., *Imbalances in regional lung ventilation: a validation study on electrical impedance tomography*, American journal of respiratory and critical care medicine, 169 (2004), pp. 791–800.
- [461] C. VOGEL, *Computational methods for inverse problems*, no. 23 in Frontiers in Applied Mathematics, SIAM, 2002.
- [462] C. VOGEL AND M. OMAN, *Fast, robust total variation-based reconstruction of noisy, blurred images*, Image Processing, IEEE Transactions on, 7 (1998), pp. 813–824.
- [463] C. R. VOGEL, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems, 12 (1996), pp. 535–547.
- [464] A. VONK NOORDEGRAAF, T. J. C. FAES, A. JANSE, J. T. MARCUS, J. G. F. BRONZWAER, P. E. POSTMUS, , AND P. M. DE VRIES, *Non-invasive assessment of right ventricular diastolic function by electrical impedance tomography*, Chest, 111 (1997), p. 1222.1228.
- [465] A. VONK NOORDEGRAAF, T. J. C. FAES, A. JANSE, J. T. MARCUS, R. M. HEETHAAR, P. E. POSTMUS, , AND P. M. DE VRIES, *Improvement of cardiac imaging in electrical impedance tomography by*

- means of a new electrode configuration*, Physiological Measurement, 17 (1996), pp. 179–188.
- [466] A. VONK NOORDEGRAAF, P. W. KUNST, A. JANSE, J. T. MARCUS, P. E. POSTMUS, T. J. FAES, AND P. M. DE VRIES, *Pulmonary perfusion measured by means of electrical impedance tomography*, Physiological Measurement, 19 (1998), pp. 263–273.
- [467] Y. WANG AND S. MA, *Projected Barzilai-Borwein method for large-scale nonnegative image restoration*, Inverse Probl. Sci. Eng., 15 (2007), pp. 559–583.
- [468] Y. WANG, A. G. YAGOLA, AND C. YANG, eds., *Optimization and Regularization for Computational Inverse Problems and Applications*, Springer, first ed., 2011.
- [469] Y. WANG, J. YANG, W. YIN, AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272.
- [470] A. J. WARD AND J. B. PENDRY, *Refraction and geometry in maxwell's equations*, Journal of Modern Optics, 43 (1996), pp. 773–793.
- [471] R. WEBBER, R. HORTON, T. UNDERHILL, J. LUDLOW, AND D. TYNDALL, *Comparison of film, direct digital, and tuned-aperture computed tomography images to identify the location of crestal defects around endosseous titanium implants*, Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology, 81 (1996), pp. 480–490.
- [472] R. L. WEBBER, R. A. HORTON, D. A. TYNDALL, AND J. B. LUDLOW, *Tuned aperture computed tomography (TACT). theory and application for three-dimensional dento-alveolar imaging*, Dentomaxillofacial Radiology, 26 (1997), pp. 53–62.
- [473] R. L. WEBBER AND J. K. MESSURA, *An in vivo comparison of diagnostic information obtained from tuned-aperture computed tomography and conventional dental radiographic imaging modalities*, Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology, 88 (1999), pp. 239–247.

- [474] R. L. WEBBER, H. R. UNDERHILL, AND R. I. FREIMANIS, *A controlled evaluation of tuned-aperture computed tomography applied to digital spot mammography*, J. Digital Imaging, 13 (2000), pp. 90–97.
- [475] R. WEDER, *A rigorous analysis of high-order electromagnetic invisibility cloaks*, J. Phys. A, 41 (2008), pp. 065207, 21.
- [476] A. WEXLER, *Electrical impedance imaging in two and three dimensions*, Clinical Physics and Physiological Measurement, Suppl A, 9 (1988), pp. 29–33.
- [477] S. C. WHITE, *1992 assessment of radiation risk from dental radiography*, Dentomaxillofacial Radiology, 21 (1992), pp. 118–26.
- [478] S. WRIGHT, *Primal-Dual Interior-Point Methods*, vol. 54, SIAM, Philadelphia PA, 1997.
- [479] Q. WU AND F. FRICKE, *Determination of blocking locations and cross-sectional area in a duct by eigenfrequency shifts*, J. Acoustical Soc., 87 (1990), pp. 67–75.
- [480] T. WU, R. H. MOORE, E. A. RAFFERTY, AND D. B. KOPANS, *A comparison of reconstruction algorithms for breast tomosynthesis*, Medical Physics, 31 (2004), pp. 2636–2647.
- [481] C. XIE, S. HUANG, B. HOYLE, AND M. BECK, *Tomographic imaging of industrial process equipment—development of system model and image reconstruction algorithm for capacitive tomography*, Sensors & Their Applications, V (1991), pp. 203–208.
- [482] C. G. XIE, A. PLASKOWSKI, AND M. S. BECK, *8-electrode capacitance system for two-component flow identification*, IEEE Proc. A, 136 (1989), pp. 173–190.
- [483] T. YATES, J. C. HEBDEN, A. GIBSON, N. EVERDELL, S. R. ARIDGE, AND M. DOUEK, *Optical tomography of the breast using a multi-channel time-resolved imager*, Physics in Medicine and Biology, 50 (2005), p. 2503.
- [484] W. YIN, S. OSHER, D. GOLDFARB, AND J. DARBON, *Bregman iterative algorithms for l_1 -minimization with applications to compressed sensing*, SIAM J. Imaging Sci., 1 (2008), pp. 143–168.

- [485] S. YOON, A. R. PINEDA, AND R. FAHRIG, *Simultaneous segmentation and reconstruction: A level set method approach for limited view computed tomography*, Medical Physics, 37 (2010), pp. 2329–2340.
- [486] T. YORK, *Status of electrical tomography in industrial applications*, Journal of Electronic Imaging, 10 (2001), pp. 608–619.
- [487] T. YORK, J. DAVIDSON, L. MAZURKIEWICH, R. MANN, AND B. GRIEVE, *Towards process tomography for monitoring pressure filtration*, Sensors Journal, IEEE, 5 (2005), pp. 139 – 152.
- [488] K. YOSHIDA, *Functional analysis*, Springer, 1966.
- [489] D. F. YU AND J. A. FESSLER, *Edge-preserving tomographic reconstruction with nonlocal regularization*, IEEE Transactions on Medical Imaging, 21 (2002), pp. 159–173.
- [490] H. YU AND G. WANG, *Compressed sensing based interior tomography*, Physics in Medicine and Biology, 54 (2009), p. 2791.
- [491] ———, *A soft-threshold filtering approach for reconstruction from a limited number of projections*, Physics in Medicine and Biology, 55 (2010), p. 3905.
- [492] L. YU, X. LIU, S. LENG, J. M. KOFLER, J. C. RAMIREZ-GIRALDO, M. QU, J. CHRISTNER, J. G. FLETCHER, AND C. H. MCCOLLOUGH, *Radiation dose reduction in computed tomography: techniques and future perspective*, Imaging in Medicine, 1 (2009), pp. 65–84.
- [493] X. ZHANG, M. BURGER, AND S. OSHER, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.
- [494] Y. ZHANG, H.-P. CHAN, B. SAHINER, J. WEI, M. M. GOODSITT, L. M. HADJIISKI, J. GE, AND C. ZHOU, *A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis*, Medical Physics, 33 (2006), pp. 3781–3795.
- [495] T. ZHOU, *Reconstructing electromagnetic obstacles by the enclosure method*, Inverse Problems and Imaging, 4 (2010), pp. 547–569.
- [496] C. M. ZIEGLER, M. FRANETZKI, T. DENIG, J. MÜHLING, AND S. HASSFELD, *Digital tomosynthesis-experiences with a new imaging device for the dental field*, Clinical Oral Investigations, 7 (2003), pp. 41–45. 10.1007/s00784-003-0195-6.

- [497] E. ZIMMERMANN, A. KEMNA, J. BERWIX, W. GLAAS, H. M. MÜNCH, AND J. A. HUISMAN, *A high-accuracy impedance spectrometer for measuring sediments with low polarizability*, Measurement Science and Technology, 19 (2008), p. 105603.
- [498] Y. ZOU AND Z. GUO, *A review of electrical impedance techniques for breast cancer detection*, Medical Engineering & Physics, 25 (2003), pp. 79–90.

Index

- acoustic scattering, 193
- Alessandrini's identity, 292
- anisotropic conductivity, 236
- Beltrami equation, 279, 346
- Beurling transform, 280, 282
- blurring, 21, 26
- Born approximation, 196, 330, 336, 337, 367, 369
- boundary data
 - Dirichlet, 211, 215, 242, 294, 360
 - Neumann, 211, 212, 218, 244
 - Robin, 223, 247
- boundary integral equation, 294, 301, 331, 345, 348, 351, 355
- breast cancer, 208, 234
- Brownian motion, 28
- calculus
 - of $\bar{\partial}$ operator, 264
- Calderón, 206, 232, 234, 260, 341, 342, 370
- camera
 - digital, 174
- Cauchy integral formula
 - generalized, 265
- Central Slice Theorem, 40, 41
- circulant matrix, 23
- cloaking, 236, 237
- complex-linear, 305, 309
- condition number, 72
- conductivity equation, 191, 211, 242, 260, 266, 279, 294
- conjugate direction method, 106
- conjugate gradient method, 106, 108
- contact impedance, 223, 247
- convection-diffusion equation, 176
- convolution, 19, 73, 190, 267, 272, 274, 282, 296, 306, 308, 341
 - matrix, 23, 80
 - operator, 268
- current pattern, 223
 - adjacent, 224, 227
 - Fourier, 224
 - optimal, 207, 231
 - pairwise, 224, 229, 329
 - Skip 3, 225, 228
 - trigonometric, 224–226, 328, 329
 - Walsh, 224
- cut-off
 - frequency, 288, 299, 355
 - function, 306, 342, 344
- data
 - noise-free, 187
- deconvolution
 - blind, 190
- degree of ill-posedness, 64, 75
- density, 193, 198, 199
- diffusion, 27
 - coefficient, 29
- diffusion equation, 202
- dispersion coefficient, 191
- drum, Can you hear the shape of, 196
- ECG, 207

- EEG, 208
- eigenfunction
 - DN map, 240, 368
 - ND map, 230, 231
- eigenvalue
 - DN map, 240, 368, 369
 - ND map, 230, 231
 - problem, 197–199
- electrode model
 - complete, 223, 247
 - continuum, 211, 217, 221, 239
 - gap, 221
 - shunt, 222
- exceptional points, 268
- existence
 - lack of, 184, 186
- far field pattern, 195, 372
- Fast Fourier Transform (FFT), 276
- Fast Fourier Transform (FFT), 24, 274, 283, 308, 310
- flowchart
 - idealized D-bar reconstruction method, 290
 - regularized D-bar reconstruction method, 300
- Fourier symbol, 269
- Fréchet differential, 261
- Fredholm
 - equation
 - first kind, 56, 60, 62, 63
 - second kind, 267, 296, 375
 - operator, 268
- fundamental solution, 31, 265, 267, 268, 270, 373
- Galerkin method, 56, 66
- GMRES, 275, 276, 281, 282, 309, 310, 352, 370, 375, 405
- Green's function, 265
- Faddeev, 270, 271, 294, 297, 303, 332
- for the Laplacian, 268, 270, 271, 302
- heat equation
 - backward, 176
- indicator function, 360, 366
- inverse crime, 19, 24, 26, 34, 50, 173, 189
- kernel
 - Hilbert-Schmidt, 60
 - weakly singular, 62
- L-curve method, 85, 97, 98, 118, 136
- Laplace
 - equation, generalized, 211, 215
 - transform, 175
- least-squares, 48, 75, 91, 112, 177
- linear operators, 59, 301
 - compact, 59, 60, 62
- injective and bounded, 68
- unbounded, 60
- linearized problem, 340
- EIT, 261, 370
- Lippmann-Schwinger equation, 195, 196, 267, 272, 294–296, 304, 305
- low-pass filter, 287, 330, 343, 344, 353
- nonlinear regularization, 314
- low-pass transport matrix, 345, 353
- minimization
 - least-squares, 177
- minimization problem, 87, 94, 109, 110, 112, 113, 119, 120, 130, 131, 135, 154, 178, 188
- nonlinear, 188, 287

- quadratic, 109, 131
minimum
 global, 188
 local, 188, 189
mollifier, 341
Morozov discrepancy principle, 97

Neumann problem, 217, 244
non-uniqueness, 54, 71, 77, 183

optical tomography, 200
optimization, 188

Panoramic (dental) imaging, 160, 163, 165
parabolic problem
 backward, 176, 190
parallelization, 189
perfusion, 329, 332, 333, 338
permittivity, 206, 210, 234, 238, 339
photograph, 19, 154, 174
Picard condition, 65
point spread function, 20, 26, 174
 discrete, 22
power, 213
power method, 231
preconditioning, 109
pseudoinverse, 77
pulmonary imaging, 208, 234

quadratic
 form, 260, 261
 functional, 90, 121
 problem, 106, 109
 programming, 111, 115, 119, 130, 131, 151, 154

reaction-diffusion equation, 190
real-linear, 305, 309, 350
refractive index, 194
regularization, 67
admissible parameter, 68, 186, 298, 299
direct, 188
iterative, 188
nonlinear, 298
nonlinear low-pass filter, 313, 315
parameter, 85, 287, 368
strategy, 68, 182, 186–188, 298, 299
Tikhonov, 7, 8, 17, 85, 111, 113, 135, 137, 144, 146, 147, 149, 156, 169, 188
regularization method
 nonlinear, 361
regularized inversion, 67, 75, 187, 297
relative error, 79

S-curve method, 118, 155
scattered field, 194
Schrödinger equation, 266, 289, 367
shunting, 222
signal-to-noise ratio, 229
singular
 functions, 63
 values, 63, 70, 73, 76, 80, 87
 vectors, 78, 80, 82, 84, 87
singular values
 generalized, 100
sinogram, 13, 48, 49, 143, 147
solute transport, 190, 191
Sommerfeld radiation condition, 195
sound speed, 192, 193
sound-hard object, 194, 372
sound-soft object, 194
space
 Banach, 59
 Besov, 8, 125, 129, 130, 138, 147, 154, 155
 Hilbert, 60

Sobolev, 212, 226, 242, 244, 267,
393, 396, 398, 399
sparsity, 8, 97, 117, 125, 130, 131,
138, 154, 155
stability, 54, 322
 conditional, 69, 184, 185
 criterion, 32
 final time, 177
 numerical, 77
stacked form, 91
Sturm-Liouville problem, 199

tangential derivative map, 349, 350
transformation optics, 236
transport equation, 202
transport matrix, 347

ultrasound, 192

variational form, 248
vocal tract, 190, 198

wave equation, 193, 197
weak form
 of DN map, 213
wedge product, 265
well-posed, 54, 59