Fundamentals of Algorithms

Discrete Inverse Problems

Insight and Algorithms

Per Christian Hansen

siam

Per Christian Hansen
Technical University of Denmark
Lyngby, Denmark

Discrete Inverse Problems

Insight and Algorithms

siam®
Society for Industrial and Applied Mathematics
Philadelphia

10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data:

Hansen, Per Christian.
Discrete inverse problems : insight and algorithms / Per Christian Hansen.
p. cm. -- (Fundamentals of algorithms series)
Includes bibliographical references and index.
ISBN 978-0-898716-96-2
1. Inverse problems (Differential equations) I. Title.
QA371.H365 2010
515'.357--dc22
2009047057

Discrete Inverse Problems

Fundamentals of Algorithms

Editor-in-Chief: Nicholas J. Higham, University of Manchester

The SIAM series on Fundamentals of Algorithms is a collection of short user-oriented books on state-of-the-art numerical methods. Written by experts, the books provide readers with sufficient knowledge to choose an appropriate method for an application and to understand the method's strengths and limitations. The books cover a range of topics drawn from numerical analysis and scientific computing. The intended audiences are researchers and practitioners using the methods and upper level undergraduates in mathematics, engineering, and computational science.

Books in this series not only provide the mathematical background for a method or class of methods used in solving a specific problem but also explain how the method can be developed into an algorithm and translated into software. The books describe the range of applicability of a method and give guidance on troubleshooting solvers and interpreting results. The theory is presented at a level accessible to the practitioner. MATLAB® software is the preferred language for codes presented since it can be used across a wide variety of platforms and is an excellent environment for prototyping, testing, and problem solving.

The series is intended to provide guides to numerical algorithms that are readily accessible, contain practical advice not easily found elsewhere, and include understandable codes that implement the algorithms.

Series Volumes

Hansen, P. C., *Discrete Inverse Problems: Insight and Algorithms*
Modersitzki, J., *FAIR: Flexible Algorithms for Image Registration*
Chan, R. H.-F. and Jin, X.-Q., *An Introduction to Iterative Toeplitz Solvers*
Eldén, L., *Matrix Methods in Data Mining and Pattern Recognition*
Hansen, P. C., Nagy, J. G., and O'Leary, D. P., *Deblurring Images: Matrices, Spectra, and Filtering*
Davis, T. A., *Direct Methods for Sparse Linear Systems*
Kelley, C. T., *Solving Nonlinear Equations with Newton's Method*

Preface

Inverse problems are mathematical problems that arise when our goal is to recover "interior" or "hidden" information from "outside"—or otherwise available—noisy data. For example, an inverse problem arises when we reconstruct a two-dimensional (2D) or three-dimensional (3D) medical image from tomography data, or when we reconstruct a sharper image from a blurred one. When we solve an inverse problem, we compute the source that gives rise to some observed data, using a mathematical model for the relation between the source and the data.

Inverse problems arise in many technical and scientific areas, such as medical and geophysical imaging, electromagnetic scattering, and nondestructive testing. Image deblurring arises, e.g., in astronomy or in biometric applications that involve fingerprint or iris recognition. The underlying mathematics is rich and well developed, and there are many books devoted to the subject of inverse (and ill-posed) problems.

So why yet another book? My experience from teaching this subject to engineering graduate students is that there is a need for a textbook that covers the basic subjects and also focuses on the computational aspects. Moreover, I believe that practical computational experience is important for understanding applied mathematics, and therefore the textbook should include a number of tutorial exercises to give the reader hands-on experience with the difficulties and challenges associated with the treatment of inverse problems.

The title of the book reflects this point of view: our *insight* about inverse problems must go hand-in-hand with our *algorithms* for solving these problems. Solving an inverse problem is rarely a matter of just picking an algorithm from a textbook, a research paper, or a software package. My experience is that each new inverse problem has its own features and peculiarities, which must be understood before one can decide on an algorithm (or, sometimes, develop a new one).

The present book is intended as a quite gentle introduction to a field characterized by advanced mathematics and sophisticated numerical methods. The book does not pretend to tell the whole story, to give all the details, or to survey all the important methods and techniques. The aim is to provide the reader with enough background in mathematics and numerical methods to understand the basic difficulties associated with linear inverse problems, to analyze the influence of measurement and approximation errors, and to design practical algorithms for computing regularized/stabilized solutions to these problems. Provided with this insight, the reader will be able to start reading the more advanced literature on the subject; indeed, anyone who wants to work in the area of linear inverse problems is advised to also consult some of the many well-written books on the subject, such as [3], [8], [14], [23], [24], [32], [64], [74], [76].

The focus of the book is on linear inverse problems in the form of Fredholm integral equations of the first kind. The presentation starts with a summary of the most important properties of linear inverse problems in the continuous setting. Then we briefly discuss discretization methods and describe how many of the properties of the integral equation directly carry over to the discretized system—in the form of a linear (perhaps overdetermined) system of equations. The next chapter is devoted to simple regularization methods for computing regularized solutions in the form of filtered spectral expansions; this is an important class of methods which clearly illustrates the basic ideas of regularization. Since no regularization algorithm is complete without a method for choosing the regularization parameter, we also include a discussion of some state-of-the-art parameter choice methods. We conclude with a chapter on iterative methods for large-scale problems, a chapter with a some real-world problems, and a chapter on a more general class of regularization methods. Sections and exercises marked with a * denote more advanced material that can be skipped in a basic course.

At the end of each section we give a number of exercises, most of them involving numerical experiments with the MATLAB package *Regularization Tools* [31], [33], which further illustrate the concepts and methods discussed in the corresponding section. The package is available from Netlib at http://www.netlib.org/numeralgo and from the MATLAB Central File Exchange at http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=52. It must be emphasized that the package is mainly intended for teaching and experimenting with small-scale inverse problems, and therefore the package is not designed to be efficient for large problems.

Acknowledgments. This tutorial grew out of a series of lectures given at the Fifth Winter School in Computational Mathematics in Geilo, Norway, in February of 2005. The atmosphere there was very positive, and I enjoyed the chance to teach one of my favorite subjects to a dedicated audience. The presentation is based on many years of experience from numerous collaborations, too many to mention here, and I thank everyone I worked with for inspiration and motivation. In particular, I thank Zdeněk Strakoš for insightful discussions about regularizing iterations, Maurizio Fedi for sharing his insight in potential field inversion, Søren Holdt Jensen for introducing me to all kinds of noise, Jim Nagy for showing me that structure is everything in image deblurring, and Bill Lionheart for all kinds of thoughts on inverse problems. I also thank Ann Manning Allen, Elizabeth Greenspan, Nancy Griscom, and Sara Murphy from SIAM for their competent handling of this book.

Per Christian Hansen
Lyngby, 2009

List of Symbols

Symbol	Quantity	Dimension
A	coefficient matrix	$m \times n$
A_k	TSVD matrix	$m \times n$
$A_k^\dagger$	pseudoinverse of A_k	$n \times m$
b	right-hand side	m
B_k	lower bidiagonal matrix	$(k{+}1) \times k$
c_λ	curvature of Tikhonov L-curve in lin-lin scale	scalar
$\hat{c}_\lambda$	ditto in log-log scale	scalar
e	noise component in right-hand side	m
E	error in quadrature or expansion method	scalar
f	solution function (integral equation)	
g	right-hand side function (integral equation)	
G	GCV function	
h	grid spacing	scalar
I	identity matrix	
k	truncation parameter (TSVD)	integer
k_η	transition index	integer
K	kernel function (integral equation)	
$\mathcal{K}_k$	Krylov subspace of dimension k	
ℓ_i	eigenvalue of kernel	scalar
L	regularization matrix	$p \times n$
L_1	discrete 1. order derivative	$(n{-}1) \times n$
L_2	discrete 2. order derivative	$(n{-}2) \times n$
m, n	matrix dimensions, $m \geq n$	scalars
s, t	independent variables (integral equation)	
u_i	left singular function or vector	m
U	left singular matrix	$m \times n$
v_i	right singular function or vector	n
V	right singular matrix	$n \times n$
w_k	basis vectors of projection method; also CGLS and Lanczos vectors	n
W_k	matrix of basis vectors	$n \times k$

Symbol	Quantity	Dimension
x	"naive" solution	n
x_k, x_λ	TSVD and Tikhonov solutions	n
$x^{[k]}$	Landweber, Cimmino, ART iteration vector	n
$x^{(k)}$	solution computed via projection; also CGLS solution	n
$x_\lambda^{(k)}$	regularized solution via projection	n
$y^{(k)}$	solution to projected problem	k
α	relative decay of SVD coefficients	scalar
α_k, β_k	from bidiagonalization algorithm	scalars
$\bar{\alpha}_k$, $\bar{\beta}_k$	from CGLS algorithm	scalars
γ, γ_k, $\hat{\gamma}_k$, γ_λ	constants in perturbation bounds	scalars
Γ	Gamma function	
δ	upper bound on solution norm also delta function	scalar
ΔA	matrix perturbation	$m \times n$
Δb	right-hand side perturbation	m
ε	upper bound on residual norm	scalar
ε_k, ε_λ	TSVD and Tikhonov regularization errors	scalars
ζ_j	expansion coefficient	scalar
η	standard deviation for noise	scalar
κ_k, κ_λ	TSVD and Tikhonov condition numbers	scalars
λ	regularization parameter (Tikhonov)	scalar
μ_i	singular value of kernel	scalar
μ	safety factor (in various methods)	scalar
ξ, $\hat{\xi}$	solution norm squared, log of ditto	scalars
ρ, $\hat{\rho}$	residual norm squared, log of ditto	scalars
σ_i	singular value of matrix	scalar
Σ	diagonal matrix with singular values	$n \times n$
τ	threshold in SSVD method	scalar
ϕ_i, ψ_i	basis functions	
φ_i	generic filter factor	scalar
$\varphi_i^{[k]}$	filter factor for an iterative method	scalar
$\varphi_i^{[\lambda]}$	Tikhonov filter factor	scalar
$\Phi^{[\cdot]}$	diagonal matrix of filter factors	$n \times n$
Ψ	matrix used to generate colored noise	$m \times m$
χ_i	"top hat" ⎍ function	
ω_j	quadrature weight	scalar
$\langle \cdot, \cdot \rangle$	inner product	
$\\|\cdot\\|_2$, $\\|\cdot\\|_F$	2-norm and Frobenius norm	
$\mathrm{cond}(\cdot)$	condition number	
$\mathrm{Cov}(\cdot)$	covariance matrix	
$\mathcal{E}(\cdot)$	expected value	
$\mathrm{span}\{\cdots\}$	subspace spanned by vectors	
$\widetilde{\square}$	perturbed version of $\square$	

Contents

Chapter 1

Introduction and Motivation

If you have acquired this book, perhaps you do not need a motivation for studying the numerical treatment of inverse problems. Still, it is preferable to start with a few examples of the use of linear inverse problems. One example is that of computing the magnetization inside the volcano Mt. Vesuvius (near Naples in Italy) from measurements of the magnetic field above the volcano—a safe way to monitor the internal activities. Figure 1.1 below shows a computer simulation of this situation; the left figure shows the measured data on the surface of the volcano, and the right figure shows a reconstruction of the internal magnetization. Another example is the computation of a sharper image from a blurred one, using a mathematical model of the point spread function that describes the blurring process; see Figure 1.2 below.

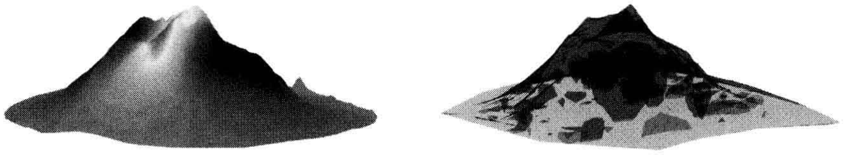

Figure 1.1. *Left: Simulated measurements of the magnetic field on the surface of Mt. Vesuvius. Right: Reconstruction of the magnetization inside the volcano.*

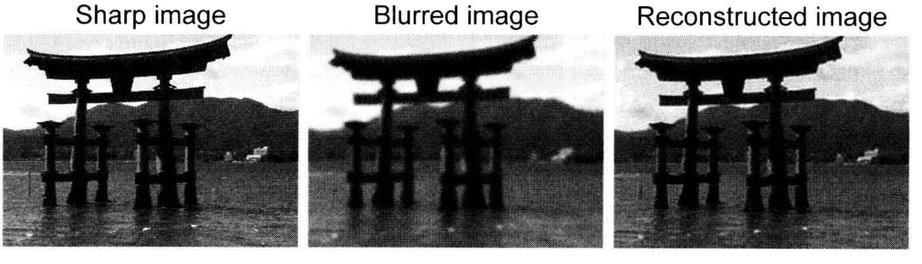

Figure 1.2. *Reconstruction of a sharper image from a blurred one.*

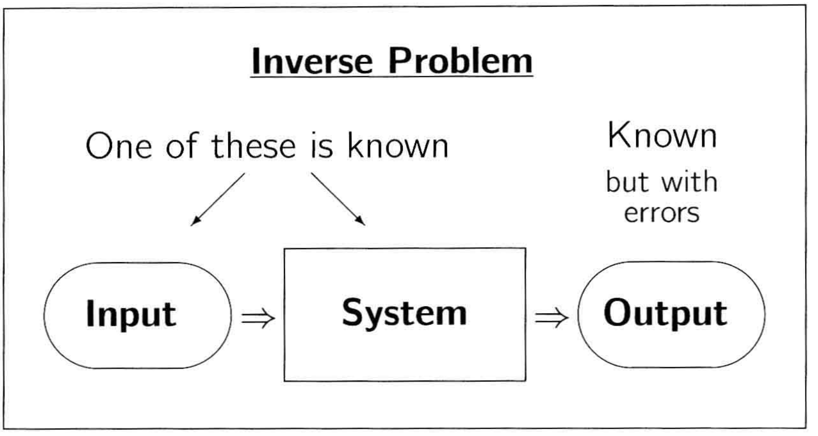

Figure 1.3. *The forward problem is to compute the output, given a system and the input to this system. The inverse problem is to compute either the input or the system, given the other two quantities. Note that in most situations we have imprecise (noisy) measurements of the output.*

Both are examples of a wide class of mathematical problems, referred to as *inverse problems*. These problems generally arise when we wish to compute information about internal or otherwise hidden data from outside (or otherwise accessible) measurements. See Figure 1.3 for a schematic illustration of an inverse problem.

Inverse problems, in turn, belong to the class of *ill-posed problems*. The term was coined in the early 20th century by Hadamard who worked on problems in mathematical physics, and he believed that ill-posed problems do not model real-world problems (he was wrong). Hadamard's definition says that a linear problem is well-posed if it satisfies the following three requirements:

- **Existence:** The problem must have a solution.
- **Uniqueness:** There must be only one solution to the problem.
- **Stability:** The solution must depend continuously on the data.

If the problem violates one or more of these requirements, it is said to be ill-posed.

The existence condition seems to be trivial—and yet we shall demonstrate that we can easily formulate problems that do not have a solution. Consider, for example, the overdetermined system

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} x = \begin{pmatrix} 1 \\ 2.2 \end{pmatrix}.$$

This problem does not have a solution; there is no x such that $x = 1$ and $2x = 2.2$. Violations of the existence criterion can often be fixed by a slight reformulation of the problem; for example, instead of the above system we can consider the associated least squares problem

$$\min_x \left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} x - \begin{pmatrix} 1 \\ 2.2 \end{pmatrix} \right\|_2^2 = \min_x \left((x-1)^2 + (2x-2.2)^2 \right),$$

which has the unique solution $x = 1.08$.

The uniqueness condition can be more critical; but again it can often be fixed by a reformulation of the problem—typically by adding additional requirements to the solution. If the requirements are carefully chosen, the solution becomes unique. For example, the underdetermined problem

$$x_1 + x_2 = 1 \qquad \text{(the world's simplest ill-posed problem)}$$

has infinitely many solutions; if we also require that the 2-norm of x, given by $\|x\|_2 = (x_1^2 + x_2^2)^{1/2}$, is minimum, then there is a unique solution $x_1 = x_2 = 1/2$.

The stability condition is much harder to "deal with" because a violation implies that arbitrarily small perturbations of data can produce arbitrarily large perturbations of the solution. At least, this is true for infinite-dimensional problems; for finite-dimensional problems the perturbation is always finite, but this is quite irrelevant if the perturbation of the solution is, say, of the order 10^{12}.

Again the key is to reformulate the problem such that the solution to the new problem is less sensitive to the perturbations. We say that we *stabilize* or *regularize* the problem, such that the solution becomes more stable and regular. As an example, consider the least squares problem $\min_x \|A\,x - b\|_2$ with coefficient matrix and right-hand side given by

$$A = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix}, \qquad b = A \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.01 \\ -0.03 \\ 0.02 \end{pmatrix} = \begin{pmatrix} 0.27 \\ 0.25 \\ 3.33 \end{pmatrix}.$$

Here, we can consider the vector $(0.01\,,\,-0.03\,,\,0.02)^T$ a perturbation of the exact right-hand side $(0.26\,,\,0.28\,,\,3.31)^T$. There is no vector x such that $A\,x = b$, and the least squares solution is given by

$$x_{\rm LS} = \begin{pmatrix} 7.01 \\ -8.40 \end{pmatrix} \qquad \Rightarrow \qquad \|A\,x_{\rm LS} - b\|_2 = 0.022\,.$$

Two other "solutions" with a small residual are

$$x' = \begin{pmatrix} 1.65 \\ 0 \end{pmatrix}, \qquad x'' = \begin{pmatrix} 0 \\ 2.58 \end{pmatrix} \qquad \Rightarrow$$

$$\|A\,x' - b\|_2 = 0.031, \qquad \|A\,x'' - b\|_2 = 0.036.$$

All three "solutions" $x_{\rm LS}$, x', and x'' have small residuals, yet they are far from the exact solution $(\,1\,,\,1\,)^T$!

The reason for this behavior is that the matrix A is ill conditioned. When this is the case, it is well known from matrix computations that small perturbations of the right-hand side b can lead to large perturbations of the solution. It is also well known that a small residual does not imply that the perturbed solution is close to the exact solution.

Ill-conditioned problems are *effectively underdetermined.* For example, for the above problem we have

$$A \begin{pmatrix} -1.00 \\ 1.57 \end{pmatrix} = \begin{pmatrix} -0.0030 \\ 0.0027 \\ 0.0053 \end{pmatrix},$$

showing that the vector $(-1.00\,,\ 1.57)^T$ is "almost" a null vector for A. Hence we can add a large amount of this vector to the solution vector without changing the residual very much; the system behaves *almost* like an underdetermined system.

It turns out that we can modify the above problem such that the new solution is more stable, i.e., less sensitive to perturbations. For example, we can enforce an upper bound δ on the norm of the solution; i.e., we solve the modified problem:

$$\min_x \|A\,x - b\|_2 \qquad \text{subject to} \qquad \|x\|_2 \le \delta.$$

The solution x_δ depends in a unique but nonlinear way on δ; for example,

$$x_{0.1} = \begin{pmatrix} 0.08 \\ 0.05 \end{pmatrix}, \quad x_1 = \begin{pmatrix} 0.84 \\ 0.54 \end{pmatrix}, \quad x_{1.37} = \begin{pmatrix} 1.16 \\ 0.74 \end{pmatrix}, \quad x_{10} = \begin{pmatrix} 6.51 \\ -7.60 \end{pmatrix}.$$

The solution $x_{1.37}$ (for $\delta = 1.37$) is quite close to the exact solution. *By supplying the correct additional information we can compute a good approximate solution.* The main difficulty is how to choose the parameter δ when we have little knowledge about the exact solution.

Whenever we solve an inverse problem on a computer, we always face difficulties similar to the above, because the associated computational problem is ill conditioned. The purpose of this book is:

1. To discuss the inherent instability of inverse problems, in the form of first-kind Fredholm integral equations.

2. To explain why ill-conditioned systems of equations always arise when we discretize and solve these inverse problems.

3. To explain the fundamental "mechanisms" of this ill conditioning and how they reflect properties of the underlying problem.

4. To explain how we can modify the computational problem in order to stabilize the solution and make it less sensitive to errors.

5. To show how this can be done efficiently on a computer, using state-of-the-art methods from numerical analysis.

Regularization methods are at the heart of all this, and in the rest of this book we will develop these methods with a keen eye on the fundamental interplay between *insight* and *algorithms.*

Chapter 2

Meet the Fredholm Integral Equation of the First Kind

This book deals with one important class of linear inverse problems, namely, those that take the form of Fredholm integral equations of the first kind. These problems arise in many applications in science and technology, where they are used to describe the relationship between the source—the "hidden data"—and the measured data. Some examples are

- medical imaging (CT scanning, electro-cardiography, etc.),
- geophysical prospecting (search for oil, land-mines, etc.),
- image deblurring (astronomy, crime scene investigations, etc.),
- deconvolution of a measurement instrument's response.

If you want to work with linear inverse problems arising from first-kind Fredholm integral equations, you must make this integral equation your friend. In particular, you must understand the "psyche" of this beast and how it can play tricks on you if you are not careful. This chapter thus sets the stage for the remainder of the book by briefly surveying some important theoretical aspects and tools associated with first-kind Fredholm integral equations.

Readers unfamiliar with inner products, norms, etc. in function spaces may ask: How do I avoid reading this chapter? The answer is: Do not avoid it completely; read the first two sections. Readers more familiar with this kind of material are encouraged to read the first four sections, which provide important background material for the rest of the book.

2.1 A Model Problem from Geophysics

It is convenient to start with a simple model problem to illustrate our theory and algorithms. We will use a simplified problem from gravity surveying. An unknown mass distribution with density $f(t)$ is located at depth d below the surface, from 0 to 1 on the t axis shown in Figure 2.1. We assume there is no mass outside this source,