DE GRUYTER

Mike Cullen, Melina A. Freitag, Stefan Kindermann, Robert Scheichl (Eds.)

LARGE SCALE INVERSE PROBLEMS

COMPUTATIONAL METHODS AND APPLICATIONS IN THE EARTH SCIENCES

RICAM
JOHANN·RADON·INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

ÖAW
Austrian Academy of Sciences

RADON SERIES ON COMPUTATIONAL AND APPLIED MATHEMATICS 13

DE
G

Large Scale Inverse Problems

Computational Methods and Applications in the Earth Sciences

Edited by
Mike Cullen
Melina A. Freitag
Stefan Kindermann
Robert Scheichl

DE GRUYTER

Mathematics Subject Classification 2010
Primary: 65F22, 47A52, 35R30, 47J06, 93E11, 62M20, 68U10, 94A08, 86A10, 86A05, 86A22;
Secondary: 93E10, 60G35, 62P12, 62C10, 62F15, 62M05, 90C06, 62G35, 62G08, 37N10, 65C40, 65C05, 65K10, 35A15

ISBN 978-3-11-028222-1
e-ISBN 978-3-11-028226-9
ISSN 1865-3707

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Typesetting: le-tex publishing services GmbH, Leipzig
Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Printed on acid-free paper
Printed in Germany

MIX
Papier aus verantwortungsvollen Quellen
FSC® C016439

www.degruyter.com

Mike Cullen, Melina A. Freitag, Stefan Kindermann, Robert Scheichl (Eds.)
Large Scale Inverse Problems

Radon Series on Computational and Applied Mathematics

Volume 13

Preface

This book contains five invited expository articles resulting from the workshop *"Large-Scale Inverse Problems and Applications in the Earth Sciences"* which took place from October 24th to October 28th, 2011, at the Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences at the Johannes Kepler University in Linz, Austria. This workshop was part of a special semester at the RICAM devoted to *"Multiscale Simulation and Analysis in Energy and the Environment"* which took place from October 3rd to December 16th, 2011. The special semester was designed around four workshops with the ambition to invoke interdisciplinary cooperation between engineers, hydrologists, meteorologists, and mathematicians.

The workshop on which this collection of articles is based was devoted more specifically to establishing ties between specialists engaged in research involving real-world applications, e.g. in meteorology, hydrology and geosciences, and experts in the theoretical background such as statisticians and mathematicians working on Bayesian inference, inverse problem and control theory.

The two central problems discussed at the workshop were the processing and handling of large scale data and models in earth sciences, and the efficient extraction of the relevant information from them. For instance, weather forecasting models involve hundreds of millions of degrees of freedom and the available data easily exceed millions of measurements per day. Since it is of no practical use to predict tomorrow's weather from today's data by a process that takes a couple of days, the need for efficient and fast methods to manage large amounts of data is obvious. The second crucial aspect is the extraction of information (in a broad sense) from these data. Since this information is often "hidden" or perhaps only accessible by indirect measurements, it takes special mathematical methods to distill and process it. A general mathematical methodology that is useful in this situation is that of inverse problems and regularization and, closely related, that of Bayesian inference. These two paths of information extraction can very roughly be distinguished by the fact that in the former, the information is usually considered a deterministic quantity, while in the latter, it is treated as a stochastic one.

A loose arrangement of the articles in this book follows this structuring of information extraction paradigms; all in view of large scale data and real-world applications:

- *Aspects of inverse problems, regularization and data assimilation.* The article by Freitag and Potthast provides a general theoretical framework for data assimilation, a special type of inverse problem and puts the theory of inverse problems in context, providing similarities and differences between general inverse problems and data assimilation problems. Lawless discusses state-of-the-art methodologies for data assimilation as a state estimation problem in current real-world applications, with particular emphasis on meteorology. In both cases, the need to treat spatial and temporal correlations effectively makes the application somewhat different from many other applications of inverse problems.

- *Aspects of inverse problems and Bayesian inference.* The survey paper by Reich and Cotter gives an introduction to mathematical tools for data assimilation coming from Bayesian inference. In particular, ensemble filter techniques and Monte Carlo methods are discussed. In this case, the need to incorporate spatial and temporal correlations makes cost-effective implementation very challenging.
- *Aspects of inverse problems and regularization in imaging applications.* The article by Burger, Dirks and Müller is an overview of the process of acquiring, processing, and interpretation of data and the associated mathematical models in *imaging sciences.* While this article highlights the benefits of the nowadays very popular nonlinear (l_1-based) regularizations, the article by van den Doel, Ascher and Haber complements the picture by contrasting these benefits with the draw-backs of l_1-based approaches and by attempting to somewhat restore the "lost honor" of the more traditional and effective, linear l_2-type regularizations.

The review-type articles in this book contain basic material as well as many interesting aspects of inverse problems, regularization and data assimilation, with the provision of excellent and extensive references to the current literature. Hence, it should be of interest to both graduate students and researchers, and a valuable reference point for both practitioners and theoretical scientists.

We would like to thank the authors of these articles for their commendable contributions to this book. Without their time and commitment, the production of this book would not have been possible. We would also like to thank Nathan Smith (University of Bath) and Peter Jan van Leeuwen (University of Reading) who helped review the articles. Additionally, we would like to express our gratitude to the speakers and participants of the workshop, who contributed to a successful workshop in Linz.

Moreover, we would like to thank Prof. Heinz Engl, founder and former director of RICAM, and Prof. Ulrich Langer, former director of RICAM for their hospitality and for giving us the opportunity to organize this workshop at the RICAM. In addition, we would like to acknowledge the work of the administrative and computer support team at RICAM, Susanne Dujardin, Annette Weihs, Wolfgang Forsthuber and Florian Tischler, as well as the local scientific organizers Jörg Willems, Johannes Kraus and Erwin Karer. The special semester, the workshops and this book would not have been possible without their efforts.

More information on the special semester and the four workshops can be found at `http://www.ricam.oeaw.ac.at/specsem/specsem2011/`.

Exeter *Mike Cullen*
Bath *Melina A. Freitag*
Linz *Stefan Kindermann*
Bath *Robert Scheichl*

Contents

Melina A. Freitag and Roland W. E. Potthast

Synergy of inverse problems and data assimilation techniques

Abstract: This review article aims to provide a theoretical framework for data assimilation, a specific type of an inverse problem arising, for example, in numerical weather prediction, hydrology and geology.

We consider the general mathematical theory for inverse problems and regularization, before we treat Tikhonov regularization, as one of the most popular methods for solving inverse problems. We show that data assimilation techniques such as three-dimensional and four-dimensional variational data assimilation (3DVar and 4DVar) as well as the Kalman filter and Bayes' data assimilation are, in the linear case, a form of cycled Tikhonov regularization. We give an introduction to key data assimilation methods as currently used in practice, link them and show their similarities. We also give an overview of ensemble methods. Furthermore, we provide an error analysis for the data assimilation process in general, show research problems and give numerical examples for simple data assimilation problems. An extensive list of references is given for further reading.

Keywords: Inverse problems, ill-posedness, regularization theory, Tikhonov regularization, error analysis, 3DVar, 4DVar, Bayesian perspective, Kalman filter, Kalman smoother, ensemble methods, advection diffusion equation, Lorenz-95 system

2010 Mathematics Subject Classification: 65F22, 47A52, 35R30, 47J06, 93E11, 62M20

Melina A. Freitag: Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom, m.freitag@maths.bath.ac.uk
Roland W. E. Potthast: Department of Mathematics, University of Reading, Whiteknights, PO Box 220, RG6 6AX, U.K. and Research and Development, Deutscher Wetterdienst, Section FE 12, Frankfurter Strasse 135, 63067 Offenbach, Germany, r.w.e.potthast@reading.ac.uk

---

We would like to thank the Johann Radon Institute for Computational and Applied Mathematics (RICAM) for hosting us in Linz during the Special Semester on Multiscale Simulation & Analysis in Energy and the Environment. The second author would like to acknowledge EPSRC, UK, for funding research on the "error dynamics of data assimilation methods," and the German government's special research program on "Remote Sensing Techniques in Data Assimilation" at Deutscher Wetterdienst (DWD).

1 Introduction

Inverse problems appear in many applications and have received a great deal of attention from applied mathematicians, engineers and statisticians. They occur, for example, in geophysics, medical imaging (such as ultrasound, computerized tomography and electrical impedance tomography), computer vision, machine learning, statistical inference, geology, hydrology, atmospheric dynamics and many other important areas of physics and industrial mathematics.

This article aims to provide a theoretical framework for data assimilation, a specific inverse problem arising, for example, in numerical weather prediction (NWP) and hydrology [48, 57, 58, 70, 83]. A few introductory articles on data assimilation in the atmospheric and ocean sciences are available, mainly from the engineering and meteorological point of view, for example, [20, 44, 48, 51, 63, 66, 71]. However, a comprehensive mathematical analysis in light of the theory of the inverse problem is missing. This expository article aims to achieve this.

An inverse problem is a problem which is posed in a way that is inverse to most direct problems. The so-called direct problem we have in mind is that of determining the effect f from given causes and conditions φ when a definite physical or mathematical model H in form of a relation

$$H(\varphi) = f \tag{1.1}$$

is given. In general, the operator H is nonlinear and describes the governing equations that relate the model parameters to the observed data. Hence, in an inverse problem, we are looking for φ, that is, a special cause, state, parameter or condition of a mathematical model. The solution of an inverse problem can be described as the construction of φ from data f (see, for example, [22, 49]). We now consider the specific inverse problem arising in data assimilation which usually also contains a dynamic aspect.

Data assimilation is, loosely speaking, a method for combining observations of the state of a complex system with predictions from a computer model output of that same state where both the observations and the model output data contain errors and (in case of the observations) are often incomplete. The task in data assimilation (and hence the inverse problem) is seeking the best state estimate with the available information about the physical model and observations.

Let X be the state space. For the remainder of this article, we generally assume that X (and also Y) are Hilbert spaces unless otherwise stated. Let $\varphi \in X$, where φ is the state (of the atmosphere, for example), that is, a vector containing all state variables. Furthermore, let $\varphi_k \in X$ be the state at time t_k and $M_k : X \to X$ the (generally nonlinear) model operator at time t_k which describes the evolution of the states from time t_k to time t_{k+1}, that is, $\varphi_{k+1} = M_k(\varphi_k)$. For the moment, we consider a perfect model, that is, the true system dynamics are assumed to be known. We also use the notation

$$M_{k,\ell} = M_{k-1}M_{k-2}\cdots M_{\ell+1}M_{\ell}, \quad k > \ell \in \mathbb{N}_0, \tag{1.2}$$

to describe the evolution of the system dynamics from time t_ℓ to time t_k.

Let Y_k be the observation space at time t_k and $f_k \in Y_k$ be the observation vector, collecting all the observations at time t_k. Finally, let $H_k : X \to Y_k$ be the (generally nonlinear) observation operator at time t_k, mapping variables in the state space to variables in the observation space. The data assimilation problem can then be defined as follows.

Definition 1.1 (Data assimilation problem). Given observations $f_k \in Y_k$ at time t_k, determine the states $\varphi_k \in X$ from the operator equations

$$H_k(\varphi_k) = f_k, \quad k = 0, 1, 2, \ldots \tag{1.3}$$

subject to the model dynamics $M_k : X \to X$ given by $\varphi_{k+1} = M_k(\varphi_k)$, where $k = 0, 1, 2, \ldots$.

In numerical weather prediction, the operator M_k involves the solution of a time-dependent nonlinear partial differential equation. Usually, the observation operator H_k is dynamic, that is, it changes at every time step. However, for simplicity, we often let $H_k := H$. Both the operator H_k and the data f_k contain errors. Also, in practice, the dynamical model M_k involves errors, that is, M_k does not represent the true system dynamics because of model errors. For a detailed account on errors occurring in the data assimilation problem, we refer to Section 4. Moreover, the model dynamics represented by the nonlinear operators M_k are usually chaotic. In the context of data assimilation, additional information might be given through known prior information (background information) about the state variable denoted by $\varphi_k^{(b)} \in X$.

The operator equation (1.3) (see also (1.1)) is usually ill-posed, that is, at least one of the following well-posedness conditions according to Hadamard [33] is not satisfied.

Definition 1.2 (Well-Posedness [49, 82]). Let X, Y be normed spaces and $H : X \to Y$ be a nonlinear mapping. Then, the operator equation $H(\varphi) = f$ from (1.1) is called well-posed if the following holds:

- Existence: For every $f \in Y$, there exists at least one $\varphi \in X$ such that $H(\varphi) = f$, that is, the operator H is surjective.
- Uniqueness: The solution φ from $H(\varphi) = f$ is unique, that is, the operator H is injective.
- Stability: The solution φ depends continuously on the data f, that is, it is stable with respect to perturbations in f.

Equation (1.1) is ill-posed if it is not well-posed.

Note that for a general nonlinear operator H, both the existence and uniqueness of the operator equation need not be satisfied. If the existence condition in Definition 1.2 is not satisfied, then it is possible that $f \in \mathcal{R}(H)$. However, for a perturbed right-hand side f^δ, we have $f^\delta \notin \mathcal{R}(H)$, where $\mathcal{R}(H) = \{f \in Y, f = H(\varphi), \varphi \in X\}$ is the range of H. Existence of a generalized solution can sometimes (for instance, in the finite-dimensional case) be ensured by solving the minimization problem

$$\min \|f - H(\varphi)\|_Y^2 , \tag{1.4}$$

which is equivalent to (1.1) if $f \in \mathcal{R}(H)$. The norm $\|\cdot\|_Y$ is a generic norm in Y. The second condition in Definition 1.2 implies that an inverse operator $H^{-1} : \mathcal{R}(H) \subseteq Y \to X$ with $H^{-1}(f) = \varphi$ exists. If the uniqueness condition is not satisfied, then it is possible to ensure uniqueness by looking for special solutions, for example, solutions that are closest to a reference element $\varphi^* \in X$, or, solutions with a minimum norm. Hence, at least in the linear case, uniqueness can be ensured if

$$\|f - H(\varphi_{uni})\|_Y = \min_{\varphi \in X} \|f - H(\varphi)\|_Y , \tag{1.5}$$

where $\|\varphi_{uni} - \varphi^*\|_X = \min\{\|\varphi - \varphi^*\|_X, \varphi \in X, \varphi \text{ is a minimizer in (1.5)}\}$. The third condition in Definition 1.2 implies that the inverse operator $H^{-1} : \mathcal{R}(H) \subseteq Y \to X$ is continuous. Usually, this problem is the most severe one as small perturbations in the right-hand side $f \in Y$ lead to large errors in the solution $\varphi \in X$ and the problem needs to be regularized. We will look at this aspect in Section 2.

From the above discussion, it follows that the operator equation (1.3) is well-posed if the operator H_k is bijective and has a well-defined inverse operator H_k^{-1} which is continuous. A least squares solution can be found by solving the minimization problem

$$\min_{\varphi_k \in X} \|f_k - H_k(\varphi_k)\|_Y^2 , \quad k = 0, 1, 2, \dots . \tag{1.6}$$

We can solve (1.6) at every time step k, which is a sequential data assimilation problem. If we include the nonlinear model dynamics constraint $M_k : X \to X$ given by $\varphi_{k+1} = M_k(\varphi_k)$, over the time steps t_k, $k = 0, \dots, K$, and take the sum of the least squares problem in every time step, the minimization problem becomes

$$\min_{\varphi_k \in X} \sum_{k=0}^{K} \|f_k - H_k(\varphi_k)\|_Y^2 = \min_{\varphi_0 \in X} \sum_{k=0}^{K} \|f_k - H_k M_{k,0}(\varphi_0)\|_Y^2 , \tag{1.7}$$

where $M_{k,0}$ denotes the evolution of the model operator from time t_0 to time t_k, that is, $M_{k,0} = M_{k-1} M_{k-2} \cdots M_0$, using the system dynamics (1.2), and $M_{k,k} = I$. Both the sequential data assimilation system (1.6) and the data assimilation system (1.7) can be written in the form

$$\min_{\varphi \in X} \left\| \overline{f} - \overline{H}(\varphi) \right\|_Y^2 , \tag{1.8}$$

with an appropriate operator $\overline{H}$. Problem (1.8) is equivalent to $\overline{H}(\varphi) = \overline{f}$ (cf. (1.1)) if $\overline{f} \in \mathcal{R}(\overline{H})$. For the sequential assimilation system (1.6), we have $\overline{H} := H_k$, $\overline{f} := f_k$ and $\varphi := \varphi_k$ at every step $k = 0, 1, \ldots$. For the system (1.7), we have $\varphi := \varphi_0$,

$$\overline{H} := \begin{bmatrix} H_0 \\ H_1 M_{1,0} \\ H_2 M_{2,0} \\ \vdots \\ H_K M_{K,0} \end{bmatrix} \quad \text{and} \quad \overline{f} := \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_K \end{bmatrix}.$$

In general, $\overline{H}$ is a *nonlinear* operator since both the model dynamics M_k and the observation operators H_k are nonlinear. If the equation $\overline{H}(\varphi) = \overline{f}$ is well-posed, then $\overline{H}$ has a well-defined continuous inverse operator $\overline{H}^{-1}$ and $\mathcal{R}(\overline{H}) = Y$.

Now, if $\overline{H}$ is a *linear* operator in Banach spaces, then well-posedness follows from the first two conditions in Definition 1.2, which are equivalent to $\mathcal{R}(\overline{H}) = Y$ and $\mathcal{N}(\overline{H}) = \{0\}$ where $\mathcal{N}(\overline{H})$ is the null space of $\overline{H}$. Moreover, if $\overline{H}$ is a *linear* operator on a finite-dimensional Hilbert space (in particular, if $\mathcal{R}(\overline{H})$ is of finite dimension), then the stability condition in Definition 1.2 holds automatically and well-posedness follows from either one of the first two conditions in 1.2. (The last condition in Definition 1.2 follows from the compactness of the unit ball in finite dimensions [49].) For linear $\overline{H}$, the uniqueness condition $\mathcal{N}(\overline{H}) = \{0\}$ is clearly satisfied if the *observability matrix* $\overline{H}$ has full row rank. In this case, the system is observable, that is, it is possible to determine the behavior of the entire system from the systems output, see [47, 73].

The remaining question is the stability of the (injective) operator equation $\overline{H}(\varphi) = \overline{f}$ (or $H\varphi = H(\varphi) = f$, a notation which we are going to use from now on) for a compact linear operator $H : X \to Y$ in infinite dimensions. As a compact linear operator is always ill-posed in an infinite-dimensional space (as $\mathcal{R}(H)$ is not closed), we need some form of regularization.

Note that the discretization of an infinite-dimensional unstable ill-posed problem naturally leads to a finite-dimensional problem which is well-posed, that is, according to Definition 1.2. However, the discrete problem will be ill-conditioned, that is, an error in the input data will still lead to large errors in the solution. Hence, some form of regularization is also needed for finite-dimensional problems arising from infinite-dimensional ill-posed operators.

In the following, we consider compact linear operators H for which a singular value decomposition exists (see, for example, [49]).

Lemma 1.3 (Singular system of compact linear operators). *Let $H : X \to Y$ be a compact linear operator. Then, there exist sets of indices $J = \{1, \ldots, m\}$ for dim$(R(H)) = m$ and $J = \mathbb{N}$ for dim$(R(H)) = \infty$, orthonormal systems $\{u_j\}_{j\in J}$ in X and $\{v_j\}_{j\in J}$*

in Y and a sequence $\{\sigma_j\}_{j\in J}$ of positive real numbers with the following properties:

$$\{\sigma_j\}_{j\in J} \text{ is non-increasing} \quad \text{and} \quad \lim_{j\to\infty} \sigma_j = 0 \text{ for } J = \mathbb{N}\,, \tag{1.9}$$

$$Hu_j = \sigma_j v_j,\ (j \in J) \quad \text{and} \quad H^* v_j = \sigma_j u_j\,, \quad (j \in J)\,. \tag{1.10}$$

For all $\varphi \in X$, there exists an element $\varphi_0 \in \mathcal{N}(H)$ with

$$\varphi = \varphi_0 + \sum_{j\in J} \left\langle \varphi, u_j \right\rangle_X u_j \quad \text{and} \quad H\varphi = \sum_{j\in J} \sigma_j \left\langle \varphi, u_j \right\rangle_X v_j\,. \tag{1.11}$$

Furthermore,

$$H^* f = \sum_{j\in J} \sigma_j \left\langle f, v_j \right\rangle_Y u_j \tag{1.12}$$

holds for all $f \in Y$. The countable set of triples $\{\sigma_j, u_j, v_j\}_{j\in J}$ is called a singular system, $\{\sigma_j\}_{j\in J}$ are called singular values, $\{u_j\}_{j\in J}$ are right singular vectors and form an orthonormal basis for $\mathcal{N}(H)^\perp$ and $\{v_j\}_{j\in J}$ are left singular vectors and form an orthonormal basis for $\overline{\mathcal{R}(H)}$.

In the following, we mostly consider compact linear operators, although the concept of ill-posedness can be extended to nonlinear operators [23, 40, 49, 82] by considering linearizations of the nonlinear problem using, for example, the Fréchet derivative of the nonlinear operator. One can show that for compact nonlinear operators, the Fréchet derivative is compact as well, leading to the concept of locally ill-posed problems for nonlinear operator equations. For solving nonlinear problems computationally, usually some form of linearization is required. Hence, most of our results for linear problems can be extended to the case of iterative solutions to nonlinear problems (where a linear problem needs to be solved at each iteration).

2 Regularization theory

Problems of the form $H\varphi = f$ with a compact operator H are ill-posed in infinite dimensions since the inverse of H is not uniformly bounded. However, in order to solve $H\varphi = f$ (or, for $f \notin \mathcal{R}(H)$, its equivalent minimization problem $\min \|H\varphi - f\|^2$), regularization is needed.

Let $H : X \to Y$ and denote its adjoint operator by $H^* : Y \to X$. Furthermore, let φ be the unique solution to the least squares minimization problem $\min \|H\varphi - f\|^2$. Then, the solution to the minimization problem is equivalent to the solution of the normal equations

$$H^* H\varphi = H^* f\,. \tag{1.13}$$

Clearly, if $H : X \to Y$ is compact, then $H^* H$ is compact and the normal equations (1.13) remain ill-posed. However, if we replace (1.13) by

$$(\alpha I + H^* H)\,\varphi_\alpha = \alpha\varphi_\alpha + H^* H\varphi_\alpha = H^* f \tag{1.14}$$