DISSERTATIONS IN
FORESTRY AND
NATURAL SCIENCES

Kati Niinimäki

Computational Optimization Methods for Large-scale Inverse Problems

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences

UNIVERSITY OF
EASTERN FINLAND

KATI NIINIMÄKI

Computational optimization methods for large-scale inverse problems

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences
No 104

Academic Dissertation
To be presented by permission of the Faculty of Science and Forestry for public examination in the auditorium L22, Snellmania building at the University of Eastern Finland, Kuopio, on June, 12, 2013,
at 12 o'clock noon.

Department of Applied Physics

Kopijyvä
Kuopio, 2013
Editors: Prof. Pertti Pasanen, Prof. Pekka Kilpeläinen,
Prof. Kai Peiponen, Prof. Matti Vornanen

Distribution:
University of Eastern Finland Library / Sales of publications
P.O. Box 107, FI-80101 Joensuu, Finland
tel. +358-50-3058396
http://www.uef.fi/kirjasto

ISBN: 978-952-61-1098-1 (nid.)
ISBN: 978-952-61-1099-8 (PDF)
ISSNL: 1798-5668
ISSN: 1798-5668
ISSN: 1798-5676 (PDF)

Author's address:
University of Eastern Finland
Department of Applied Physics
P.O. Box 1627
70211 Kuopio
Finland
email: kati.niinimaki@uef.fi

Supervisors:
Associate Professor Ville Kolehmainen, PhD
University of Eastern Finland
Department of Applied Physics
Kuopio, Finland
email: ville.kolehmainen@uef.fi

Professor Samuli Siltanen, PhD
University of Helsinki
Department of Mathematics and Statistics
Helsinki,Finland
email: samuli.siltanen@helsinki.fi

Professor Jari Kaipio, PhD
University of Eastern Finland
Department of Applied Physics
Finland
email: jari@math.auckland.ac.nz

Reviewers:
Professor Martin Burger, PhD
Westfälische Wilhelms- Universität
Institute for Computational and Applied Mathematics
Münster, Germany
email: martin.burger@wwu.de

Docent Ozan Öktem, PhD
Royal Institute of Technology
Department of Mathematics
Stockholm, Sweden
email: ozan@kth.se

Opponent:
Dr. Carola-Bibiane Schönlieb
University of Cambridge
Department of Applied Mathematics and Theoretical Physics
Cambridge, England
email: C.B.Schoenlieb@damtp.cam.ac.uk

ABSTRACT

In this thesis, novel computational optimization methods for large-scale inverse problems are proposed. To overcome the difficulties related to the ill-posedness of practical inverse problems different approaches are considered when developing the computational methods proposed in this thesis. These approaches include Bayesian inversion, wavelet-based multiresolution analysis and constrained optimization methods. The performance of the proposed computational optimization methods are tested in four studies.

The first study proposes a wavelet-based Bayesian method that allows significant reduction of the number of unknowns in the inverse problem without deteriorating the image quality inside the region of interest. The inverse problem considered is the reconstruction of X-ray images from local tomography data.

The second study proposes a sparsity-promoting Bayesian inversion method that uses Besov B^1_{11} space norms with wavelet functions to describe the prior information about the quantity of interest. The computation of the maximum a posteriori (MAP) estimates with B^1_{11} prior is a non-differentiable optimization problem, therefore it is reformulated into a form of a quadratic program (QP) and a primal-dual interior-point (PD-IP) method is derived to solve the MAP estimates.

In the third study the PD-IP method is used to reconstruct discontinuous diffusion coefficients of a coupled parabolic system from limited observations based on a stability result for the inverse coefficient problem. In the fourth study PD-IP method is used to solve Bayesian MAP estimates from sparse angle X-ray tomography data. Besov B^1_{11} space prior with Haar wavelet basis is used to represent the distribution of the attenuation coefficients inside the image domain.

This thesis also proposes a novel method for selecting the Bayesian prior parameter. The method is based on a priori information about the sparsity of the unknown function. Based on the test

results the proposed parameter selection rule, as well as the proposed computational optimization methods, seem promising and adaptable for several practical applications.

AMS Classification: 62C10, 62F15, 42C40, 65T60, 30H25, 90C51, 90C25, 65R32, 62P10, 92C55, 65F22
Universal Decimal Classification: 517.956.27, 517.956.37, 517.956.47, 519.226, 519.6
INSPEC Thesaurus: optimisation; numerical analysis; Bayes methods; quadratic programming; convex programming; biomedical imaging; image reconstruction; tomography; diffusion
Yleinen suomalainen asiasanasto (YSA): Bayesialainen menetelmä; matemaattinen optimointi; laskentamenetelmät; kuvantaminen; tomografia; diffuusio

Acknowledgments

The research work of this thesis was mainly carried out at the Department of Applied Physics of the University of Eastern Finland on Kuopio campus. I want to thank all those who contributed to my research work toward this thesis. In particular, I want to thank the following people.

First, I want to thank my supervisors Associate Professor Ville Kolehmainen, PhD, and Professor Samuli Siltanen, PhD for their guidance during this work and for the productive co-operation. I am also grateful to my supervisor Professor Jari Kaipio, PhD for giving me the opportunity to work in his research group at the Department of Applied Physics. I want to thank all of my co-authors, especially Professor Matti Lassas, PhD for the fruitful collaboration.

I wish to thank the official pre-examiners Docent Ozan Öktem, PhD and Professor Martin Burger, PhD for the assessment of my thesis. In addition I wish to thank Professor Hamish Short, PhD for revising the linguistic of the thesis.

The work during the last two and a half year of this thesis was carried out in Marseille, France in a research laboratory, *Laboratoire d'Analyse, Topologie, Probabilités* (LATP), of the Aix-Marseille University. I am very grateful to Professor Patricia Gaitan,PhD, Professor Michel Cristofol, PhD and Professor Olivier Poisson, PhD for giving me the great opportunity to work with them and with the Applied Analysis research group and for the fruitful collaboration, for the inspiration and for their friendship. Also, I want to thank all of the staff and the members of the laboratory for the very warm welcome I received. Especially I wish to thank Professor Yves Dermenjian,PhD, Professor Florence Hubbert, PhD, Professor Franck Boyer, PhD and Professor Assia Bedabdallah, PhD, for their friendship and valuable advice. I wish to thank Professor Hary Rambelo for sharing his office with me at LATP.

I want to thank the faculty and the staff of the Department of

Applied Physics in Kuopio. Especially I wish to thank Joanna Sierpowska, PhD and Timo Lähivaara, PhD for their friendship and for the scientific and non-scientific discussions. I want to thank Mauri Puoskari, PhD for the computer related assistance during these years. Also, I thank all of the secretaries especially Soile Lempinen for the administrative support.

Finally, I thank all of my friends and family, especially my parents Anita and Sakari Niinimäki to whom I have inherited the persistence and the ambition which were valuable for me during this journey into the doctorate. My warmest thanks I dedicate to my husband Eero for his love and support.

This study was supported by the Finnish Funding Agency for Technology and Innovation (TEKES), the Academy of Finland, the Finnish Centre of Excellence in Inverse Problems Research 2006-2011, the Finnish Cultural Foundation North Savo Regional fund, the Emil Aaltonen Foundation, the Oskar Öflund Foundation, the Finnish Foundation for Technology Promotion, the Finnish-French Technology Society, the Embassy of France and by the CIMO an organization for international mobility and co-operation.

Kuopio May 15, 2013 *Kati Niinimäki*

ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
CBCT	Cone beam computed tomography
CM	Conditional mean
CS	Compressive sensing
CT	Computed tomography
EIT	Electrical impedance tomography
FBP	Filtered back projection
FS	Focal spot
IP	Interior-point
KKT	Karush Kuhn Tucker
LP	Linear programming
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
ML	Maximum likelihood
PC	Personal computer
PD	Primal-dual
PDE	Partial differential equations
PD-IP	Primal-dual interior-point
PET	Positron emission tomography
QP	Quadratic programming
ROI	Region of interest
TV	Total variation

NOMENCLATURE

$\mathbf{1}$	Vector of all ones
$a_{i,i}$	Reaction coefficients
A	Forward matrix
$\mathcal{A}, \mathcal{A}_{\mathcal{I}}, \mathcal{A}_{\mathcal{E}}$	Constraint matrices
$b, b_{\mathcal{I}}, b_{\mathcal{E}}$	Constraint vector
$B, B^{-1}, \widetilde{B}$	Wavelet transform matrices
B^s_{pq}	Besov space
$\mathbf{c}, c_1(x), c_2(x)$	Unknown diffusion coefficients
$\widetilde{\mathbf{c}}, \widetilde{c}_1(x), \widetilde{c}_2(x)$	Known diffusion coefficients
$c_{J_0,k}$	Approximation wavelet coefficients
C	Positive constant
d	QP vector
$D(\cdot)$	Dual objective function
$\mathcal{D}, \mathcal{D}_2$	Matrices related to time derivatives
E	Matrix related to spatial derivative
f	Unknown function or unknown vector
$F(z)$	Objective functional
$\mathcal{F}$	Mapping between spaces
g	Primal variable (vector)
G	Diagonal matrix with elements g_i
H	Matrix
h	Integer
h^+, h^-	Non-negative vectors
$\mathcal{H}$	Set of indices
i	Integer
I	Identity matrix
j	Wavelet integer related to the scale
J_0, J, J_{out}	Number of scaling levels
$\mathcal{J}$	Set of indices
k, k', k_1, k_2	Wavelet integer related to the spatial locations
ℓ	Integer related to the wavelet type
$L(\mathbb{R})$	Lebesgue space

$\mathcal{L}(\cdot)$	Lagrangian function
m	Measurement vector
M	Constant
$\mathcal{M}$	Space
$n, n_m, n_z, n_w, n_r, n_s$	Dimensions of vectors
N	Constant
N_t	Number of steps in time
N_x	Number of steps in space
$\mathcal{N}$	Gaussian density
$\mathcal{N}_i$	Four-point neighborhood for pixel i
p	Besov parameter
$P(\cdot)$	Primal objective function
P_j	Orthogonal projection operator
q	Besov parameter
Q	QP matrix
R	Model reduction matrix
r	Constant
$\mathbb{R}^n$	n dimensional space
s	Besov parameter
$S, \hat{S}$	Sparsity levels
$\mathcal{S}$	Subset of basis indices
t	Time variable
T	Time constant
$\mathcal{T}$	Projection operator
$\mathbb{T}, \mathbb{T}^2$	1D and 2D periodic spaces
$u_1, u_2, \mathbf{u}, \tilde{u}_1, \tilde{u}_2, \tilde{\mathbf{u}}$	Vectors
$u_{0,i}(\cdot)$	Initial conditions
v	Dual variable (vector)
$\mathcal{V}$	Diagonal matrix with elements v_i
V_j	Spaces
$w, \widetilde{w}$	Wavelet coefficient vector
$w_{j,k}, w_{j,k',\ell}$	Wavelet (detail) coefficients
$\mathcal{W}$	Diagonal matrix of weights
W_j	Spaces

x, x_1, x_2	Spatial variables
X	Space
y	Dual variable (vector)
z	Primal variable (vector)
$\mathbb{Z}$	Set of integers
α	Prior parameter
β	Smoothness parameter
γ_v	μ-complementarity
δ	Relative error
Δ	Search direction
ϵ	Noise vector
η	Positive constant
θ	Time constant
Θ	Domain
κ	QP constant
λ_{primal}	Primal step length multiplier
λ_{dual}	Dual step length multiplier
Λ	Lower bound for primal variable
μ	Central path parameter
ν	Integer
π	Probability distribution
ρ, ρ_g	Primal feasibilities
σ	Standard deviation
τ	Dual feasibility
$\phi(x), \phi_{j,k}(x)$	Scaling function (or father wavelet)
$\psi(x), \psi_{j,k}(x)$	Wavelet function (or mother wavelet)
$\omega, \Omega_{c_1}, \Omega_{c_2}, \Omega_{\text{ROI}}$	Subdomains
Ω	Domain
$\partial\Omega$	Boundary of Ω
Σ	Summation
$\partial_t, \partial_{tt}, \partial_x$	Partial derivatives

LIST OF PUBLICATIONS

This thesis consists of an overview and the following four original articles, which are referred to in the text by their Roman numerals **I-IV**:

I K. Niinimäki, S. Siltanen, and V. Kolehmainen. Bayesian multiresolution method for local tomography in dental x-ray imaging. *Physics in Medicine and Biology*, 52:6663-6678, 2007.

II V. Kolehmainen, M. Lassas, K. Niinimäki and S. Siltanen. Sparsity-promoting Bayesian inversion. *Inverse Problems*, 28:025005(28pp), 2012.

III M. Cristofol, P. Gaitan, K. Niinimäki and O. Poisson. Inverse problem for a coupled parabolic system with discontinuous conductivities: one-dimensional case. *Inverse Problems and Imaging* 7(1):159-182, 2013.

IV K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki and S. Siltanen. Sparse tomography. *SIAM Journal on Scientific Computing*, 2013, in press.

AUTHOR'S CONTRIBUTION

All publications in this thesis are results of joint work with the co-authors. The author of this thesis was the principal writer in **I** and **IV**. In **II** the writing task was equally divided among the authors. The author of this thesis wrote the numerical part of Publication **III** as well as computed the numerical results. The author of this thesis was responsible of developing all the computational optimization methods presented in Publications **I-IV**, carried out the numerical implementations and testings and computed the results on a Matlab® platform.

Contents

1 Introduction

Consider a linear model of the form

$$m = Af + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I) \tag{1.1}$$

where $m \in \mathbb{R}^{n_m}$ denotes measurements, $f \in \mathbb{R}^n$ denotes a physical quantity, $A \in \mathbb{R}^{n_m \times n}$ is a matrix describing the relationship between m and f and ϵ represents the measurement noise assumed to be additive Gaussian white noise with a standard deviation $\sigma > 0$. The inverse problem related to (1.1) is the following: recover f from indirect and noisy measurements m. Inverse problems are typically ill-posed i.e. the problem is non-unique and even small errors in the data m may cause arbitrarily large errors in the estimate of f.

Due to the ill-posed nature of inverse problems, *a priori* information needs to be incorporated into the model. In a statistical framework this can be done using Bayesian inversion methods [1–7]. In Bayesian inversion all quantities related to the problem are modelled as random variables. The randomness reflects the uncertainty concerning their true values. As a solution to the inverse problem a probability distribution of the unknown parameters is estimated when data m and all the available *a priori* information about f are given. This probability distribution is called the *posterior distribution*. Often the inverse problems are large-scale problems and the corresponding posterior distribution cannot be visualized directly, therefore, typically a single estimate of this distribution is presented as a solution. One of the most commonly used single estimates is the *maximum a posteriori* (MAP) estimates

$$f^{\mathrm{MAP}} = \arg\max_{f \in \mathbb{R}^n} \pi_{\mathrm{post}}(f|m) \propto \arg\max_{f \in \mathbb{R}^n} \pi(m|f)\pi_{\mathrm{pr}}(f),$$

where $\pi(m|f)$ is the likelihood function and $\pi_{\mathrm{pr}}(f)$ is the prior model. In Bayesian inversion the prior model $\pi_{\mathrm{pr}}(f)$ introduces into the model the information about the unknown f that is known prior to the measurements.

The *a priori* information can be quantitative or qualitative. In physical problems, for example, non-negativity is a general quantitative prior model. A type of qualitative prior information is, for example, that it is expected that f is sparse or that f has a sparse representation in some

basis. The sparsity means that $f \in \mathbb{R}^n$ has $n_s << n$ nonzero coefficients. For example, if f is assumed to contain smooth regions with relatively few sharp edges it often can be represented more sparsely on a wavelet basis [8–16]. Thus if the prior model is constructed such that this sparsity is respected, more accurate finite-dimensional approximations of f can be computed.

The sparse recovery of f is closely related to the concept of *compressive sensing* (CS). The field of CS grew out of the works of Candès, Romberg, Tao and Donoho [17–22] who showed that a signal, which has a sparse representation, can be recovered exactly from a small set of linear measurements. Since then the CS has been an active research topic.

In regularization the use of $\|\cdot\|_1$ norms instead of $\|\cdot\|_2$ norms is known to promote sparsity [17,22–25]. Thus in Bayesian frameworks, one might consider constructing the prior model such that it contains a norm of $\|\cdot\|_1$ type such as, for example, a total variation (TV) norm [25] or a wavelet-based Besov B^1_{11} norm [26,27] (see also **II** and **IV**).

The computation of the Bayesian MAP estimate with TV prior or with wavelet-based Besov prior is equivalent with the following optimization problem

$$f^{\text{MAP}} = \arg\min_{f\geq 0} \left(\frac{1}{2\sigma^2}\|Af - m\|_2^2 + \alpha\|f\| \right), \tag{1.2}$$

where α is the prior parameter and $\|f\| = \|f\|_{TV}$ for the TV prior, $\|f\| = \|f\|_{B^1_{11}}$ for the Besov B^1_{11} prior and the constraint ($f \geq 0$) is related to the non-negativity prior. In practical applications, when f^{MAP} is a large dimensional vector, the computation of the MAP estimates from (1.2) is a large-scale constrained optimization problem.

In a general form the constrained optimization problems can be stated as

$$\begin{array}{ll} \min_z & F(z) \\ \text{s.t} & \mathcal{A}_{\mathcal{I}}(z) \geq 0 \\ & \mathcal{A}_{\mathcal{E}}(z) = 0 \end{array}, \tag{1.3}$$

where $F(z)$ is the objective functional, $\mathcal{A}_{\mathcal{I}}$, $\mathcal{A}_{\mathcal{E}}$ denote the inequality and equality constraints, respectively and z is a vector containing the sought estimate of the unknown. There exist several methods that can be used to solve (1.3), these include linear programming (LP), such as the simplex method and the interior point (IP) method, quadratic programming (QP), such as the active set method, interior point method, gradient projection method and the augmented Lagrangian method, for example [28–31]. The choice, of which method to use, depends on the form of the functional

$F(z)$ and on the constraints. Often in problems that arise from practical applications, the matrices associated in the objective functional $F(z)$ and in the constraints $A_{\mathcal{I}}, A_{\mathcal{E}}$, are large in dimension but sparse. Hence it is preferable to select such a method that can handle large and sparse matrices effectively.

One approach that can exploit the sparsity of large matrices is the interior point (IP) method. Furthermore, it has been found that primal-dual (PD) IP methods have both good theoretical properties and also better performance in practice than other IP methods (see [29, 31–42], for example). Furthermore, [42] shows that primal-dual interior-point (PD-IP) methods are more efficient for minimizing a sum of norms or sum of absolute values than interior-point methods or other classical methods. Hence the PD-IP method is considered in this thesis for solving the large-scale constrained optimization problems encountered in Publications **II**-**IV**.

The aims and contents of this thesis

The aim of this thesis is to develop methods that can solve problems of the form (1.1) in large dimensional settings. Different approaches such as Bayesian inversion methods, wavelet-based multiresolution analysis and PD-IP methods were considered when developing the proposed computational optimization methods. The performance of the developed methods were tested in the four studies **I** - **IV**.

In Publication **I** Bayesian inversion methods were combined with wavelet-based multiresolution analysis. The considered problem was the problem of reconstructing X-ray images from dental local tomography data. Publication **I** proposed a method that allowed a remarkable reduction in the number of unknowns in the problem without deteriorating the image quality inside the region of interest (ROI). Test cases with both simulated and experimental data were considered. With the experimental data $\sim$ 90% of the unknown coefficients could be reduced leading to a similar reduction also in the computational time.

Publication **II** studied a discretization-invariant Bayesian inversion model. As prior models a Besov B^1_{11} prior and a total variation (TV) prior were considered. The method was tested numerically in the context of a 1D deconvolution task. The computation of Bayesian MAP estimates either with B^1_{11} prior or with TV prior is a non-differentiable optimization problem. In **II** this non-differentiable problem was reformulated into a QP form and a PD-IP method was derived in order to compute the MAP estimates. According to the results, MAP estimates that were

computed using the B^1_{11} prior with Haar wavelets were sparsity promoting, discretization-invariant and edge-preserving. Furthermore, a novel sparsity-based choice rule for selecting the prior parameter α was proposed. The proposed α-selection method seemed to perform robustly also under noisy conditions.

Publication **III** considered an inverse problem of recovering discontinuous diffusion coefficients related to a coupled parabolic system. A PD-IP method was used to recover these coefficients from limited observations. The results indicated that with the PD path following IP method accurate reconstructions of the discontinuous diffusion coefficients could be computed.

In Publication **IV** PD-IP method was used to compute Bayesian MAP estimates with Besov B^1_{11} prior. The considered problem was a tomographic reconstruction problem in 2D in a case where the projection data was sparsely sampled in the angular variable. The results demonstrated that the Bayesian MAP estimates performed robustly also when the number of projections were limited. Publication **IV** further developed the α-selection method of Publication **II** thus reducing the computational complexity of the method.

This thesis is organized as follows. Chapter 2 presents briefly 1D and 2D wavelets, Besov spaces and the related Besov norm. In Chapter 3 the Bayesian approach is introduced. Also a review on previous studies where TV and Besov priors have been used in Bayesian framework, is given. Chapter 4 describes briefly a general QP problem and introduces the PD path following IP method used in this thesis. Previous studies where PD-IP methods have been applied to practical inverse problems are also reviewed in Chapter 4. Chapter 5 reviews the studies and results of Publications **I** - **IV** and Chapter 6 summarizes and concludes the studies of this thesis.

2 Wavelets and Besov spaces

The interest in the use of wavelets and wavelet analysis has greatly increased since the early 1980's. Since then wavelets have been used in various applications including signal processing, image analysis, operator theory and many other applications. Many classes of functions can be represented by wavelets in a more compact way making wavelets an excellent tool for data compression and sparse recovery. Wavelets can also be considered for a time-frequency analysis of non-stationary signals since they are local both in time and in frequency(scale). Furthermore, wavelets can be created in such a way that they are suitable for multiresolution analysis. The multiresolution representation of a function yields a hierarchical framework for interpreting the information content at different resolutions.

This chapter gives a short introduction to wavelets by presenting the one- and two-dimensional (1D and 2D) wavelet functions and the corresponding wavelet expansions. Further, in this chapter Besov spaces and Besov norms are briefly introduced. The presentation in this chapter follows the standard references [8–16, 43].

2.1 WAVELET FUNCTIONS

In the following only orthonormal wavelets, which have a compact support and are suitable for multiresolution analysis, are considered.

2.1.1 One-dimensional case

Let $\psi_{j,k}(x)$ be a family of functions defined by dilatations and translations of a single function $\psi(x) \in L^2(\mathbb{R})$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),$$

$j, k \in \mathbb{Z}$. The function ψ is called the wavelet function (or the mother wavelet). For the wavelet function we have

$$\int \psi(x)\mathrm{d}x = 0, \quad \text{and} \quad \|\psi_{j,k}\|_{L^2} = 1.$$

Further, the closure of the linear span of $\{\psi_{j,k}, k \in \mathbb{Z}\}$ defines the spaces W_j [8,9,11,13,15,16]

$$W_j := \overline{span\{\psi_{j,k}, k \in \mathbb{Z}\}}.$$

Similarly let $\phi_{j,k}(x)$ denote the family of functions

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k),$$

where $j, k \in \mathbb{Z}$. The function ϕ is called the scaling function (or the father wavelet) and the closure of the linear spans of $\{\phi_{j,k}, k \in \mathbb{Z}\}$ defines the spaces V_j [8,9,11,13,15,16]

$$V_j := \overline{span\{\phi_{j,k}, k \in \mathbb{Z}\}}.$$

2.1.2 Two-dimensional case

Similarly to the one-dimensional case set

$$\begin{aligned} \phi_{j,k}(x) &= 2^j\phi(2^j x - k) \\ \psi_{j,k}(x) &= 2^j\psi(2^j x - k) \end{aligned}$$

and define the corresponding spaces

$$\begin{aligned} V_j &:= \overline{span\{\phi_{j,k}, k \in \mathbb{Z}\}} \\ W_j &:= \overline{span\{\psi_{j,k}, k \in \mathbb{Z}\}}, \end{aligned}$$

where $j \in \mathbb{Z}$. To obtain the 2D wavelet functions, the standard tensor product construction [8, 15] is used in this thesis. For the 2D location index let $k' = (k_1, k_2)$ denote independent translations in the coordinates $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$, respectively. Thus the 2D wavelets are defined as follows

$$\phi_{j,k'}(x_1, x_2) = \phi_{j,k_1}(x_1)\phi_{j,k_2}(x_2) \tag{2.1}$$

$$\psi^1_{j,k'}(x_1, x_2) = \phi_{j,k_1}(x_1)\psi_{j,k_2}(x_2) \tag{2.2}$$

$$\psi^2_{j,k'}(x_1, x_2) = \psi_{j,k_1}(x_1)\phi_{j,k_2}(x_2) \tag{2.3}$$

$$\psi^3_{j,k'}(x_1, x_2) = \psi_{j,k_1}(x_1)\psi_{j,k_2}(x_2) \tag{2.4}$$

2.2 MULTIRESOLUTION ANALYSIS

A multiresolution analysis (MRA) consists of a nested set of closed subspaces that satisfy

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \tag{2.5}$$

with

(i) $\overline{\left(\bigcup_{j\in\mathbb{Z}} V_j\right)} = L^2(\mathbb{R})$

(ii) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$

(iii) $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}, \quad j \in \mathbb{Z}$

(iv) $f(x) \in V_j \Leftrightarrow f(x - 2^{-j}k) \in V_j, \quad j \in \mathbb{Z}$

(v) $V_{j+1} = V_j \oplus W_j, \quad V_j \perp W_j$

The spaces $\{W_j\}$ satisfy similar conditions as the spaces $\{V_j\}$ hence if $f(x) \in W_j$ it follows that $f(2x) \in W_{j+1}$ and $f(x - 2^{-j}k) \in W_j$, for all $j \in \mathbb{Z}$.

Assume that there exists a function $\phi \in L^2(\mathbb{R})$ such that $\{\phi(x - k), \quad k \in \mathbb{Z}\}$ forms a basis of V_0. Then together with (iii) it follows that $\{\phi_{j,k}\}$ and $\{\psi_{j,k}\}$ form orthonormal bases of V_j and W_j, respectively. If the closed subspaces satisfy (2.5) and (i)-(v), then the orthonormal function ϕ generates an orthonormal MRA.

2.3 WAVELET EXPANSION

In the following let ψ and ϕ denote compactly supported scaling and wavelet functions, respectively, such that their support intersects the interval $[0, 1]$. Further, let $f : [0, 1] \to L^2(\mathbb{R})$ and let P_j denote an orthogonal projection operator onto V_j. By the property (v), the subspaces V_j and W_j are orthogonal and for $J_0 < j$,

$$V_j = V_{j-1} \oplus W_{j-1} = \cdots = V_{j-J_0} \oplus \left(W_{j-J_0} \oplus \cdots \oplus W_{j-1}\right).$$

Further, the property (i) ensures that $\lim_{j\to\infty} P_j f_j = f$, for all $f \in L^2(\mathbb{R})$, thus a function $f \in L^2(\mathbb{R})$ can be approximated (as closely as desired) by $P_j f_j \in V_j$

$$P_j f_j = \sum_{k=0}^{2^{J_0}-1} \langle f, \phi_{J_0,k}\rangle \phi_{J_0,k} + \sum_{j=J_0}^{\infty} \sum_{k=0}^{2^j-1} \langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

Denoting $c_{J_0,k} = \langle f, \phi_{J_0,k} \rangle$ and $w_{j,k} = \langle f, \psi_{j,k} \rangle$ the wavelet expansion of a 1D function f is written

$$f(x) = \sum_{k=0}^{2^{J_0}-1} c_{J_0,k}\phi_{J_0,k} + \sum_{j=J_0}^{\infty}\sum_{k=0}^{2^j-1} w_{j,k}\psi_{j,k}. \tag{2.6}$$

Similarly in 2D, $f : [0,1]^2 \to L^2(\mathbb{R}^2)$ can be expanded using the 2D scaling and wavelet functions

$$f(x_1,x_2) = \sum_{k_1=0}^{2^{J_0}-1}\sum_{k_2=0}^{2^{J_0}-1} c_{J_0,k'}\phi_{J_0,k'} + \sum_{j=0}^{\infty}\sum_{k_1=0}^{2^j-1}\sum_{k_2=0}^{2^j-1}\sum_{\ell=1}^{3} w_{j,k',\ell}\psi^{\ell}_{j,k'}, \tag{2.7}$$

where the index ℓ is used to denote the type of wavelet (see (2.1) - (2.4)).

2.4 BESOV SPACES AND BESOV NORM

Besov spaces $B^s_{pq}(\mathcal{M})$ are function spaces where $s \in \mathbb{R}$ is a smoothness parameter and $1 \leq p,q \leq \infty$ are integrability exponents. Here $\mathcal{M}$ denotes the appropriate space, for example, $\mathbb{R}, \mathbb{R}^2, \mathbb{T}$ or $\mathbb{T}^2$, where $\mathbb{T}$ and $\mathbb{T}^2$ denotes the 1D and 2D periodic spaces, respectively. The Besov space is equipped with a norm defined by

$$\|f\|_{B^s_{pq}} = \|f\|_{L^p(\Omega)} + |f|_{B^s_{pq}},$$

where $\Omega \subset \mathcal{M}$ and $|f|_{B^s_{pq}}$ is the Besov seminorm. Denoting the rth modulus of smoothness by

$$\omega_r(f,t)_p = \sup_{|h|\leq t} \|\Delta_h^{(r)} f\|_{L^p(\Omega)}, \quad t > 0, \quad r \in \mathbb{N},$$

where $\Delta_h^{(r)} = \Delta_h^{(r-1)} f(x+h) - \Delta_h^{(r-1)} f(x)$ is the rth forward difference, the Besov seminorm writes

$$|f|_{B^s_{pq}} = \left(\sum_{k=1}^{\infty} \left(2^{ks} \omega_r(f, 2^{-k}, \Omega)_p \right)^q \right)^{\frac{1}{q}},$$

with $s > 0$, $0 < p,q < \infty$ [11,44].

The Besov norm can also be expressed by means of wavelet coefficients. By choosing a sufficiently smooth wavelet basis, such that the

mother wavelet ψ and the scaling function ϕ are smooth enough, a function f belongs to $B^s_{pq}(\mathcal{M})$ if and only if [8,11,14–16,43]

$$\begin{aligned}\|f\|_{B^s_{pq}(\mathcal{M})} = & \left(\sum_{k=0}^{2^{J_0}-1} |c_{J_0,k}|^p\right)^{\frac{1}{p}} \\ & + \left(\sum_{j=J_0}^{\infty} 2^{jq\left(s+\frac{1}{2}-\frac{1}{p}\right)} \left(\sum_{k=0}^{2^j-1} |w_{j,k}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}\end{aligned} \tag{2.8}$$

is finite. Setting $p = q$ the 1D Besov norm of (2.8) is written

$$\|f\|_{B^s_{pp}} = \left(\sum_{k=0}^{2^{J_0}-1} |c_{J_0,k}|^p + \sum_{j=J_0}^{\infty} 2^{j\left(ps+\frac{p}{2}-1\right)} \sum_{k=0}^{2^j-1} |w_{j,k}|^p\right)^{\frac{1}{p}}. \tag{2.9}$$

The corresponding Besov norm in 2D is written

$$\begin{aligned}\|f\|_{B^s_{pp}} = & \left(\sum_{k_1=0}^{2^{J_0}-1} \sum_{k_2=0}^{2^{J_0}-1} |c_{J_0,k'}|^p \right. \\ & \left. + \sum_{j=J_0}^{\infty} 2^{jp\left(s+1-\frac{2}{p}\right)} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\ell=1}^{3} |w_{j,k',\ell}|^p\right)^{\frac{1}{p}}.\end{aligned} \tag{2.10}$$

3 Bayesian inversion

In the Bayesian approach to inverse problems all unknown variables are modelled as random variables. The randomness models the uncertainty on the variables true values and it can be expressed in terms of probability distributions of the quantities. In a Bayesian framework a complete solution to an inverse problem is a *posterior probability distribution* of the unknown quantity, given the measured data and the *a priori* information. As general references to Bayesian inversion see [1–7].

3.1 THE POSTERIOR DISTRIBUTION

Consider the problem of finding $f \in \mathbb{R}^n$ from the measurements $m \in \mathbb{R}^{n_m}$ when $f$and m are related by the model (1.1)

$$m = Af + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I).$$

The joint probability density of f and m can be written as

$$\pi(f, m) = \pi(f|m)\pi(m) = \pi(m|f)\pi(f) \tag{3.1}$$

where $\pi(f|m)$ is the posterior distribution of f given m, $\pi(m|f)$ is the likelihood function describing the measurements, $\pi(f)$ is the prior density representing the *a priori* information of the unknown and $\pi(m)$ is the marginal density of m.

A complete solution to the inverse problem, the posterior distribution, can be derived from (3.1), thus the posterior distribution writes

$$\pi_{\text{post}}(f|m) = \frac{\pi(m|f)\pi(f)}{\pi(m)}, \tag{3.2}$$

where the marginal density $\pi(m)$ is often considered as a normalizing constant. Equation (3.2) is known as the Bayes formula.

3.2 POINT ESTIMATES

Practical (real-world) inverse problems are often large-scale problems and the related posterior distribution is high dimensional and cannot be vi-

sualized directly. However, different point, spread and interval estimates can be computed in order to visualize the solution. The *maximum a posterior* (MAP) estimate

$$f^{\mathrm{MAP}} = \arg\max_{f \in \mathbb{R}^n} \pi_{\mathrm{post}}(f|m)$$

is one of the most common point estimates. The computation of the MAP estimate leads to a large-scale optimization problem.

Another commonly used point estimate is the *conditional means* (CM) estimate

$$f^{\mathrm{CM}} = \int f \pi(f|m) \mathrm{d}f.$$

The computation of the CM estimates leads to an integration problem in a high-dimensional space.

3.3 THE LIKELIHOOD FUNCTION

The construction of the likelihood function $\pi(m|f)$ is based on the forward model and on the information about the noise. An additive noise model ($m = Af + \epsilon$) with the assumption that the noise is Gaussian white noise, with a standard deviation $\sigma > 0$, is often a feasible choice. Further, assuming that the noise ϵ and the unknown f are mutually independent the likelihood function can be written as

$$\pi(m|f) = \pi_{\mathrm{noise}}(m - Af) = C \exp\left\{ -\frac{1}{2\sigma^2} \|Af - m\|_2^2 \right\}$$

where C is a normalizing constant.

3.4 PRIOR MODELS

In Bayesian inversion the prior model is the way to incorporate *a priori* information about f into the inversion process. The challenge of constructing a suitable prior model lies in the nature of the prior information.

The prior information is often qualitative, for example it can be expected that the solution is piecewise regular having only few irregularities (jumps). Then the prior model should be constructed such that sharp jumps in the reconstruction are allowed. Also it can be expected that the unknown function f is sparse or has a sparse representation in some transformation domain. For instance smooth functions with few local

irregularities can be represented sparsely on wavelet transformation domains.

This thesis considers two prior models that are adaptable to sparse representations, namely a wavelet-based Besov space prior model and a total variation (TV) prior model.

The Besov space prior can formally be written as

$$\pi_{\mathrm{pr}}(f) = \exp\left\{-\alpha \|f\|_{B^s_{pq}}\right\},$$

where $\alpha > 0$ is the prior parameter and $\|f\|_{B^s_{pq}}$ is the Besov space norm, which can be expressed by means of wavelet coefficients (see section 2.4).

The Besov priors in Bayesian frameworks have been studied in the literature. In [45] Besov space priors were used for wavelet denoising. Wavelet-based Besov priors have been applied to practical tomographic problems in [46,47]. In [26] Besov B^1_{11} space prior was studied in the context of discretization-invariant Bayesian inversion. In [26] it was shown that the use of B^1_{11} prior yields discretization-invariant Bayesian estimates in the sense that the estimates behave consistently at all resolutions and the same prior information can be used regardless of the discretization level. In [27] Besov prior and wavelets were applied to an inverse problem of finding diffusion coefficients related to an elliptic partial-differential equation. In deterministic regularization, in particular in Tikhonov regularization, Besov norm penalties have been extensively studied in the literature, see [48–53] for example.

The TV prior model can formally be written as

$$\pi_{\mathrm{pr}}(f) = \exp\left\{-\alpha \|f\|_{\mathrm{TV}}\right\}, \tag{3.3}$$

where $\|f\|_{\mathrm{TV}}$ is the TV norm of f.

The TV norm was originally introduced in 1992 by Rudin, Osher and Fatemi [25] as a regularizer for image restoration. Since then it has been used extensively. For example TV has been used for image restoration, denoising, deblurring, image inpainting, signal recovery and tomographic reconstruction problems both in deterministic and statistic frameworks. Thus the attempt to survey and to do justice to all of its contributors would be quite a considerable task and thus falls outside of the scope of this thesis. Therefore, here we only refer to works that are directly related to this thesis i.e. where TV prior has been used to recover Bayesian MAP estimates (see **I** and **II**).

Publication **I** considered a problem of reconstructing an image from

tomographic data. In [54–59] TV priors were also used when computing Bayesian MAP estimates from tomographic data. [54–57] considered X-ray tomography, [58] electrical impedance tomography (EIT) and [59] positron emission tomography (PET).

In Publication **II** TV priors were used to compute Bayesian MAP estimates in the context of a discretization-invariant Bayesian inversion. The considered problem was a deconvolution problem. The MAP estimates were recovered using a convex quadratic programming algorithm. For previous studies with deconvolution problems where TV priors have been used to compute Bayesian MAP estimates see [60–63], for example. In [60] MAP estimates were recovered under non-negativity constraint using a convex programming algorithm. [64] also considered TV priors for Bayesian inversion. The results of [64] showed that TV priors cannot be used for discretization-invariant Bayesian inversion.

3.5 COMPUTATION OF THE MAP ESTIMATE

The computation of the MAP estimates using the TV prior (3.3) is equivalent with the following minimization problem

$$f_{\mathrm{TV}}^{\mathrm{MAP}} = \arg\min_f \left(\frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha \sum_{i=1}^{n-1} |(Ef)_i| \right), \quad f \geq C$$

where α is the prior parameter, E denotes the discrete operator related to the first derivative of f and C is a constant (often $C = 0$).

The Besov space prior can be derived using the Besov norm of section 2.4, thus the computation of the MAP estimate using Besov space prior amounts to the following minimization problem

$$f_{B_{pq}^s}^{\mathrm{MAP}} = \arg\min_f \left(\frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha \|f\|_{B_{pq}^s}^p \right), \quad f \geq C$$

where a computationally efficient form of the Besov norm i.e. the power of p of the Besov norm, has been used.

Denoting the direct and inverse wavelet transforms ((2.6) and (2.7)) by

$$w = B^{-1}f \qquad \text{and} \qquad f = Bw,$$

respectively[1], and choosing the Besov parameters p, q and s such that $p = q = s = 1$, the computation of the MAP estimate (with B^1_{11} prior) is equivalent to

$$f^{\text{MAP}}_{B^s_{pq}} = \arg\min_f \left(\frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha \sum_{j=1}^{n} |\mathcal{W}B^{-1}f|_j \right), \quad f \geq C$$

where $\mathcal{W}$ is a diagonal matrix containing the powers of 2 weights of formulas (2.9) and (2.10). Note that in 2D $\mathcal{W} = I$ when $p = q = s = 1$ (see (2.10)).

[1] $w = B^{-1}f$ are the first n wavelet coefficients of f when the wavelets are ordered into a sequence and re-numbered by an index $j \in \mathbb{Z}$

4 Primal - dual interior - point method

In constrained optimization the publication of Karmarkar's study [32] of interior-point (IP) methods for linear programming in 1984 can be seen as a turning point in the development of modern IP methods. After that in 1990s primal-dual (PD) IP methods have proved themselves to be very efficient for several applications; they have good practical performance and they can be extended to wide classes of problems including linear programming (LP), quadratic programming (QP) and sequential QP problems. Most of the present PD-IP methods, including the one presented in this thesis, are based on the Mehrotra's predictor corrector method published in 1992 [33]. As general references for linear and quadratic programming see [28–31], for example.

This chapter is organized as follows. The section 4.1 present briefly a QP problem subject to equality and inequality constraints. The section 4.2 presents the primal-dual path following interior-point method that has been used in Publications **II-IV** and section 4.3 reviews previous studies on PD-IP methods related to this thesis. The considered constraints in Publications **II-IV** were linear equality and bounds-on-variable constraints, hence section 4.2 concentrates on presenting an optimization problem with quadratic objective function subject to linear equality and bounds-on-variable constraints. The extension to optimization problems with inequality constraint and/or nonlinear optimization problems can be derived using simple extensions to the approach presented in this chapter.

4.1 QUADRATIC PROGRAMMING

An optimization problem that has a quadratic objective function and linear constraints is called a quadratic program (QP). In a general form the QP problem can be stated as follows

$$\begin{array}{ll} \min_z & \left(\frac{1}{2} z^T Q z + z^T d + \kappa\right) \\ \text{s.t.} & \mathcal{A}_{\mathcal{E}} z = b_{\mathcal{E}} \\ & \mathcal{A}_{\mathcal{I}} z \geq b_{\mathcal{I}} \end{array} \tag{4.1}$$

where Q is a symmetric $n_z \times n_z$ matrix, z and d are vectors in $\mathbb{R}^{n_z}$ and the subscripts $\mathcal{E}$ and $\mathcal{I}$ denote the equality and inequality constraints, respectively. When the matrix Q is positive semidefinite the QP problem is convex, similarly when the matrix Q is indefinite the QP problem is nonconvex. In this thesis only the case of convex QP problems are considered. Methods for solving problems of the form (4.1) include active set methods, conjugate gradient methods, interior-point methods and augmented Lagrangian methods, for example [28–31].

4.2 PRIMAL-DUAL PATH FOLLOWING INTERIOR-POINT METHOD

A primal problem with a quadratic objective function and linear constraints can be presented as follows

$$\begin{aligned} \min_z \quad & P(z) \\ \text{s.t.} \quad & \mathcal{A}z = b \\ & z \geq \Lambda \end{aligned} \tag{4.2}$$

where $P(z) = \left(\frac{1}{2}z^T Q z + z^T d + \kappa\right)$ is the primal objective function.

To begin the derivation of the interior point method, slack variables $g_i, i = 1, \ldots, n_z$ are introduced to the inequality constraints thus rewriting the primal problem (4.2) as

$$\begin{aligned} \min_z \quad & P_\mu(z, g) \\ \text{s.t.} \quad & \mathcal{A}z = b \\ & z - \Lambda = g \end{aligned} \tag{4.3}$$

where the objective function $P_\mu(z, g) = \left(\frac{1}{2}z^T Q z + z^T d + \kappa - \mu \sum_{i=1}^{n_z} \log(g_i)\right)$ is the classical Fiacco-McCormick type logarithmic barrier function [65], with $\mu > 0$. The Lagrangian for this problems is

$$\begin{aligned} \mathcal{L}(z, g, y, v; \mu) = \quad & \tfrac{1}{2}z^T Q z + z^T d + \kappa - \mu \textstyle\sum_{i=1}^{n_z} \log(g_i) \\ & + y^T(b - \mathcal{A}z) + v^T(\Lambda - z + g) \end{aligned}, \tag{4.4}$$

where y and v are the Lagrangian multipliers for the constraints $\mathcal{A}z = b$ and $x - \Lambda = g$, respectively. The Lagrangian multipliers (y and v) are also the dual variables of the associated Lagrangian dual problem (or simply

the dual problem)

$$\begin{aligned} \max_z \quad & D(z,y,v) \\ \text{s.t.} \quad & \mathcal{A}^T y + v - Qz = b, \\ & v \geq 0 \\ & y \quad \text{free} \end{aligned}$$

where $D(z,y,v) = \left(\kappa + b^T y + \Lambda^T v - \frac{1}{2} z^T Q z\right)$ is the dual objective function. The dual variable v is complementary to the non-negative primal variable g, which implies that also v is non-negative. The dual variable y is associated with the primal equality constraint making y a free variable.

In Lagrangian duality the basic idea is to take the constraints of the problem into account by augmenting the objective function by a weighted sum of the constraints. The duality has an important role in the PD-IP methods since the dual objective function $D(z,y,v)$ provides a lower bound for the primal objective function $P(z)$. Further, at the optimal point (z^*, g^*, y^*, v^*) the solution of the primal objective function $P(z^*)$ is equal to the solution of the dual objective function $D(z^*, y^*, v^*)$, i.e. $P(z^*) \equiv D(z^*, y^*, v^*)$.

4.2.1 The KKT-conditions and the central path

The optimality conditions for the constrained problem can be derived by stating the conditions for a minimum of (4.4)

$$\nabla_z \mathcal{L}(z,g,y,v;\mu) = Qz + d - \mathcal{A}^T y - v = 0 \tag{4.5}$$
$$\nabla_g \mathcal{L}(z,g,y,v;\mu) = -\mu G^{-1}\mathbf{1} + v = 0 \tag{4.6}$$
$$\nabla_y \mathcal{L}(z,g,y,v;\mu) = b - \mathcal{A}z = 0 \tag{4.7}$$
$$\nabla_v \mathcal{L}(z,g,y,v;\mu) = \Lambda - z + g = 0 \tag{4.8}$$

where $\mathbf{1}$ is a vector of all ones and G denotes diagonal matrix with elements g_i. Multiplying (4.6) with G the conditions (4.5) - (4.8) yield the first-order necessary conditions, often referred to as the *Karush-Kuhn-Tucker*

(KKT) conditions, for the primal-dual system, namely the conditions

$$\begin{aligned} \mathcal{A}^T y + v - Qz &= d && (4.9)\\ \mathcal{A}z &= b && (4.10)\\ z - g &= \Lambda && (4.11)\\ G\mathcal{V}\mathbf{1} &= \mu\mathbf{1} && (4.12) \end{aligned}$$

where $\mathcal{V}$ denotes diagonal matrix with elements v_i. Equation (4.9) is called the dual feasibility, (4.10) and (4.11) the primal feasibilities and (4.12) is the μ-complementarity. When the matrix Q in (4.9) is positive semidefinite these KKT-conditions are both necessary and sufficient optimality conditions for the QP problem (4.2) [28, 30].

The parameter μ parameterizes the central path, which is followed in the PD path following IP method. The central path is defined by the trajectory $\mathcal{P}$: $\mathcal{P}\{(z_\mu, g_\mu, y_\mu, v_\mu)|\mu > 0\}$. For each $\mu > 0$ the associated central path defines a point in the primal-dual space simultaneously satisfying the conditions (4.9) - (4.12). As $\mu \to 0$ the trajectory $\mathcal{P}$ converges to the optimal solution of both the primal and dual problems. At the optimal point (z^*, g^*, y^*, v^*), $\mu \equiv 0$ and $P(z^*) = D(z^*, y^*, v^*)$ and furthermore, the barrier objective function $P_\mu(z^*, g^*)$ at the optimal point (x^*,g^*) is equivalent with the original primal objective function $P(z^*)$ hence the QP problem (4.2) can be solved by finding a solution to the system (4.9) - (4.12). Writing the conditions (4.9) - (4.12) in a form of a mapping $\mathcal{F} : \mathbb{R}^{3n_z+n_y} \to \mathbb{R}^{3n_z+n_y}$

$$\mathcal{F}(z, g, y, v; \mu) = \begin{bmatrix} Qz - \mathcal{A}^T y - v + d \\ \mathcal{A}z - b \\ z - \Lambda - g \\ G\mathcal{V}\mathbf{1} - \mu\mathbf{1} \end{bmatrix} = 0,$$

applying Newton's method and assuming (for the moment) that μ is fixed, one obtains the following linear system, which yields the search directions

$$\begin{bmatrix} -Q & 0 & \mathcal{A}^T & I \\ I & 0 & 0 & 0 \\ I & -I & 0 & 0 \\ 0 & G^{-1}\mathcal{V} & 0 & I \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta g \\ \Delta y \\ \Delta v \end{bmatrix} = \begin{bmatrix} \tau \\ \rho \\ \rho_g \\ \gamma_v \end{bmatrix},$$

where

$$\begin{aligned} \tau &= Qz + d - \mathcal{A}^T y - v + r && (4.13)\\ \rho &= \mathcal{A}z - b && (4.14)\\ \rho_g &= z - \Lambda - g && (4.15)\\ \gamma_v &= v - \mu G^{-1}\mathbf{1} + G^{-1}\Delta G \Delta \mathcal{V} && (4.16) \end{aligned}$$

The variables Δg and Δv can be eliminated without producing any off-diagonal entries in the remaining system hence giving a more compact system i.e. the reduced KKT system

$$\begin{bmatrix} -(Q + G^{-1}\mathcal{V}) & \mathcal{A}^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta y \end{bmatrix} = \begin{bmatrix} \tau - G^{-1}\mathcal{V}\rho_g - \gamma_v \\ \rho \end{bmatrix} \quad (4.17)$$

with Δg and Δv defined as

$$\begin{aligned} \Delta g &= \mathcal{V}^{-1} G(\gamma_v - \Delta v)\\ \Delta v &= \gamma_v + G^{-1}\mathcal{V}(\rho_g - \Delta z). \end{aligned}$$

4.2.2 The predictor-corrector approach

The algorithm used in this thesis, to solve the QP problem, is based on Mehrotra's predictor-corrector method [33]. Following the Mehrotra approach the method presented in this thesis proceeds iteratively starting from an initial point (z^0, g^0, y^0, v^0) through a sequence of points determined by

$$\begin{aligned} z^{i+1} &= z^i + \lambda_{\text{primal}} \Delta z && (4.18)\\ g^{i+1} &= g^i + \lambda_{\text{primal}} \Delta g && (4.19)\\ y^{i+1} &= y^i + \lambda_{\text{dual}} \Delta y && (4.20)\\ v^{i+1} &= v^i + \lambda_{\text{dual}} \Delta v && (4.21) \end{aligned}$$

where λ_{primal} is a step length multiplier for the primal variables z and g and λ_{dual} is a step length multiplier for the dual variables y and v. The search directions $(\Delta z, \Delta g, \Delta y, \Delta v)$ are determined using the predictor corrector method.

The computation of $(\Delta z, \Delta g, \Delta y, \Delta v)$ in (4.18) - (4.21) can be summarized as follows:

1) Set $\mu = 0, \Delta G = 0$ and $\Delta \mathcal{V} = 0$ and solve (4.17) in order to compute

the predictor steps $(\Delta z^{\text{pre}}, \Delta g^{\text{pre}}, \Delta y^{\text{pre}}, \Delta v^{\text{pre}})$

2) Calculate a value for μ. In this thesis μ is computed using a similar method first proposed by Vanderbei and Shanno in 1999 [35], thus μ is computed by the equation

$$\mu = 0.1 \times \min\left(0.05\frac{1-\xi}{\xi}, 2\right)^3 \frac{v^T g}{n_z}, \tag{4.22}$$

where $\xi = \frac{\min_i v_i g_i}{v^T g / n_z}$

3) Compute γ_v (4.16) with $(\Delta g^{\text{pre}}, \Delta v^{\text{pre}})$ of step 1) using μ of step 2) (equation (4.22))

4) Solve (4.17) for $(\Delta z, \Delta g, \Delta y, \Delta v)$.

4.2.3 Selection of the step length

The PD-IP method of this thesis uses different step length multipliers λ_{primal} and λ_{dual} for the primal (z, g) and dual (y, v) variables, respectively. The unequal step length multipliers for primal and dual variables are obtained as follows. First the maximal feasible step lengths that enforce the positivity of the primal variable g and the dual variable v are computed using a simple ratio test

$$\begin{aligned}
\lambda_{\text{primal}}^{\max} &= \min\left(\min_{\Delta g_i < 0}\left(-\frac{g_i}{\Delta g_i}\right), 1\right), \quad i = 1, \ldots, n_z \\
\lambda_{\text{dual}}^{\max} &= \min\left(\min_{\Delta v_i < 0}\left(-\frac{v_i}{\Delta v_i}\right), 1\right), \quad i = 1, \ldots, n_z,
\end{aligned}$$

in order to keep $g^{(i+1)}$ and $v^{(i+1)}$ positive. Then the dual step length multiplier is obtained by setting: $\lambda_{\text{dual}} = 0.95 \times \lambda_{\text{dual}}^{\max}$. For the primal step length multiplier, a backtracking line search is used on interval $[0, 0.95 \times \lambda_{\text{primal}}^{\max}]$, such that also the primal constraints, $\mathcal{A}z = b, x - \Lambda = g$, are taken into account thus ensuring a progress also towards primal feasibility.

4.3 LITERATURE REVIEW

Since the works by Karmarkar in 1984 [32] and Mehrotra in 1992 [33] PD-IP methods have been an active research topic. Both theoretical and practical performances of PD-IP methods have been extensively studied within various application areas. Thus referring to all of them would be

quite a considerable task and thus falls beyond the scope of this thesis. The references given in the following sections refer to the works that are most closely related to the topic of this thesis.

The references with respect to the publicly available convex QP algorithms are provided section 4.3.1. Section 4.3.2, on the other hand, reviews a few of the studies where PD-IP methods have been applied to large-scale inverse problems arising from practical applications.

4.3.1 Algorithms

For standard convex OP problems (4.1) there exist general purpose solvers such as Matlab®'s quadprog [66], Scilab®'s qpsolver [67], MOSEK [68], CVX [69], CPLEX [70] and LOQO [71, 72], for example. In quadprog the user may choose between three methods: the trust-region method, interior-point method or active-set method. In CVX the user can choose between two modes: the semidefinite programming mode or the geometric programming mode. MOSEK, CPLEX and LOQO use the state-of-the-art interior-point method. Further, CPLEX and LOQO use the Mehrothra's [33] predictor corrector approach.

However, often most of these standard solvers can only solve small and medium size problems and fail to solve large-scale problems. For example, the user's guide of CVX suggests an image size of 50×50 for image reconstruction problems. Therefore, in application areas where the problems arise from real world and are large-scale problems standard general purpose solvers may fail to work and customized algorithms are needed[1].

4.3.2 Applications

In [36] Johnson and Sofer derived a specialized PD-IP method for 3D emission tomography reconstruction problem. Within the Bayesian framework MAP estimates with positivity constrained were recovered using a PD-IP method. In [74] the authors used PD-IP method to compute maximum likelihood (ML) estimates from 3D positron emission tomography (PET) data.

In [40] Borsic *et al* derived a PD-IP framework to compute electrical impedance tomography (EIT) reconstructions. As a regularization for the

[1] As a solver for large-scale nonlinear optimization problems see [73], for example

EIT problem they used TV regularization. In [41] Borsic and Adler derived a PD-IP method that used $\|\cdot\|_1$ norm either in the data fitting term, or in the regularization term or on both. Reconstructions from EIT data were computed with a PD-IP method. In [75] an EIT image reconstruction problem was also considered and a PD-IP method was used to solve problems with $\|\cdot\|_1$ and mixed norm formulations.

PD-IP methods have also been used for PDE constrained problems. In [76–83] PD-IP methods were used to solve PDE constrained optimal control problems. In [84,85] PD-IP methods were applied to optimal control problems with parabolic constraints.

In [38] Kim *et al* used PD-IP method for sparse signal recovery problem of magnetic resonance imaging and Zhang *et al* for bioluminescene tomography in [86]. PD-IP methods have also been considered for deblurring problems [37], for image restoration problems [39] and for compressed sensing [87].

5 Review of the Publications

This chapter reviews the studies and results of Publications **I** - **IV**. In Publication **I** the reconstruction problem arose from dental local tomography, Publication **II** handled a deconvolution problem, in Publication **III** the task was to recover diffusion coefficients related to a coupled parabolic system and in Publication **IV** the reconstruction problem originated from sparse angle X-ray tomography.

5.1 PUBLICATION I: RECONSTRUCTING X-RAY IMAGES FROM LOCAL TOMOGRAPHY DATA

The motivation of this research stemmed from dental X-ray imaging, in particular from dental Cone Beam Computed Tomography (CBCT) imaging. Typically in dental CBCT imaging a C-arm with an X-ray source and a digital detector rotates around the patient's head collecting 2D projection images. The CBCT imaging is an example of local tomography. In local tomography the aim is to reconstruct the distribution of the attenuation coefficients inside the region of interest (ROI) from a set of truncated projection images [88–92]. For an illustration of the geometry, global vs. local tomography, see figure 5.1. Compared to the traditional CT scanners the dental CBCT devices provide higher resolution and lower costs. However, the dental CBCT produces sparse projection data as opposed to the global and dense CT data since the projections are truncated and possibly coarsely sampled in the angular variable. The truncation of the projection images is typically enforced by limited detector size or by the intention to minimize the radiation dose to critical organs. The image reconstruction from such data is an ill-posed inverse problem.

The attenuation coefficients in X-ray tomography have often been represented by piecewise constant voxel bases, see [93–96] for example. However, the use of voxel-based models in dental CBCT imaging may become computationally problematic due to the high resolution requirement; the dentist needs to see detail of size ~ 0.1 mm inside the ROI. If, however, the whole volume Ω, that contains a cross-section of the patient's head, is covered with voxels of size $0.1 \times 0.1 \times 0.1$ mm^3, the number of unknowns in the inverse problem becomes impractically large. One option could be to cover only the ROI with small enough voxel and not to model the other

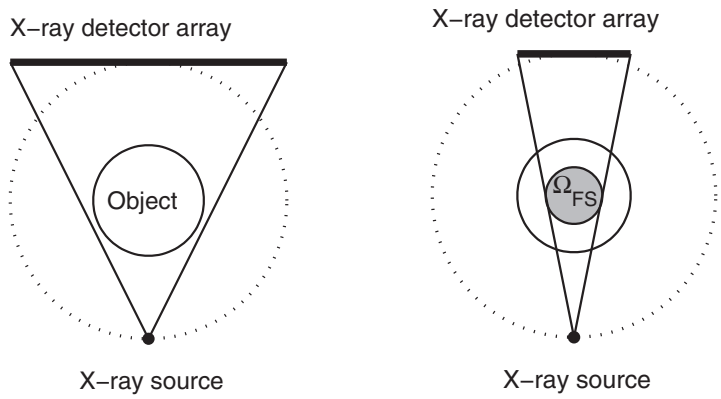

*Figure 5.1: Global (left) and local (right) tomography. The source and detector panel rotate around the object along the dotted circle. Ω_{FS} represents the sub domain that is present in all of the projection images in the local tomography case (*FS = *Focal spot). Usually the region of interest is chosen to be Ω_{FS}.*

tissues at all; this, however, will typically lead to severe reconstruction artefacts. Publication **I** proposed a wavelet-based Bayesian multiresolution method for local tomography where a Besov space prior was used for the dental structures.

5.1.1 The wavelet-based Bayesian multiresolution

The basic idea in the wavelet-based multiresolution method is to reduce the number of unknowns in the inverse problem by using a high resolution only inside the ROI and a coarser resolution elsewhere in the reconstruction volume. This can be done by using wavelets to represent the tissue structures. Thanks to the locality, scaling and shifting properties of wavelets [8, 10, 11, 13–16, 43] this kind of multiresolution representation can be achieved.

Let $f = Bw, w \in \mathbb{R}^{n_w}$ denote the full wavelet expansion of an image f with all the available scales J used over all of the image domain and let $\mathcal{H} = \{1, \ldots, n_w\}$ denote the associated indexing of the basis in the wavelet expansion. Further, let $\mathcal{S} \subset \mathcal{H}$ denote the subset of basis indices that contains all the scales up to J in Ω_{ROI} and the scale up to J_{out} in $\Omega \backslash \Omega_{\text{ROI}}$, such that $J_{\text{out}} \leq J$. Now the multiresolution representation of the image f can (formally) be written as

$$f = \tilde{B}\tilde{w}, \quad \tilde{w} = Rw \in \mathbb{R}^{n_r}, \quad \tilde{B} = BR^T, \quad n_r \leq n_w \tag{5.1}$$

where $R \in \mathbb{R}^{n_r \times n_w}$ is a model reduction matrix with elements

$$R_{i,h} = \begin{cases} 1 & \mathcal{S}_i = h \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{S}_i$ denotes the ith element of $\mathcal{S}$. Note that $f \in \mathbb{R}^n$ and $w \in \mathbb{R}^{n_w}$ with $n_w \geq n$ depending on the chosen wavelet basis and on the boundary conditions used.

Using equation (5.1) the reduced model for the projection measurements in the multiresolution approach becomes

$$m = A\tilde{B}\tilde{w} + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I). \tag{5.2}$$

Following section 3.5 the computation of the MAP estimate related to the reduced model (5.2) is equivalent with the following minimization problem

$$\tilde{w}_{\text{MAP}} = \arg\min_{\tilde{w}} \left(\frac{1}{2\sigma^2} \|A\tilde{B}\tilde{w} - m\|_2^2 + \alpha \|\tilde{B}\tilde{w}\|_{B^s_{pq}}^p \right). \tag{5.3}$$

The minimization of (5.3) in **I** was realized by using the nonlinear conjugate gradient Polak-Ribière method [97].

5.1.2 Materials and methods

Two different test cases of local tomography, corresponding to a dental CBCT situation in 2D, were considered:

(i) simulated data consisting of 187 projections from a total opening angle of 187° with 1° projection interval

(ii) experimental data consisting of 23 projections from sparsely distributed directions with total opening angle of 187 ° with 8.5° projection intervals

In the inverse problem the size of the pixel grid used over the whole image domain Ω was 256×256 for the simulated case and 498×498 for the experimental case.

The following choices were used for both test cases. Daubechies 6 wavelets were used in the multiresolution model (5.3) and a half-point symmetric padding was used for the treatment of the boundaries. The prior parameter α was set to $\alpha = 30$. The Besov parameters p, q and s were set to $p = q = 1.5$ and $s = 0.5$. In [46] the authors used the same

choices for p, q and s and demonstrated that such values are well suited for dental imaging. Furthermore, this parameter selection enables the use of gradient-based minimization methods in the computations of the MAP estimates since the functional in (5.3) becomes differentiable.

In Publication **I** the maximum available numbers of scaling levels J in Ω_{ROI} were used thus giving $J = 5$ for the simulated case and $J = 6$ for the experimental case. To evaluate the performance of the multiresolution method the number of scaling levels J_{out} was gradually decreased and the effect of this decrease on the image quality in Ω_{ROI} was studied. With the simulated data five reconstructions were computed having $J_{\text{out}} = 5, 4, 3, 2, 1$ in $\Omega \backslash \Omega_{\text{ROI}}$, respectively. J in Ω_{ROI} was set to $J = 5$ for all of the reconstructions. Corresponding to the experimental data six reconstructions were computed with $J_{\text{out}} = 6, 5, 4, 3, 2, 1$ in $\Omega \backslash \Omega_{\text{ROI}}$, respectively and $J = 6$ in Ω_{ROI} for all of the six reconstructions.

As a reference for the wavelet-based multiresolution method, pixel-based MAP estimates with total variation (TV) prior were computed. TV priors were originally introduced for modelling blocky objects [23–25] and they have been found to perform well for dental structures, see [54, 57], for example.

In order to use the gradient-based optimization method when computing the MAP estimates with TV prior, an approximate TV was used in Publication **I**. The computation of the MAP estimates with the approximate TV prior is equivalent to the following minimization problem

$$f_{\text{TV}}^{\text{MAP}} = \arg\min_f \left(\frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha \sum_{i=1}^{n} \sum_{\nu \in \mathcal{N}_i} \sqrt{(\beta + (f_i - f_\nu))} \right), \quad (5.4)$$

where $\mathcal{N}_i$ denotes a four-point neighborhood for pixel i and β is a smoothing parameter. In Publication **I** the parameter α and β in (5.4) were set to $\alpha = 0.1$ and $\beta = 10^{-4}$, respectively.

The idea of the multiresolution method for reducing the number of unknowns by using higher resolution only inside the ROI was extended analogously to the pixel domain in **I**. This was done by considering Ω_{ROI} as the whole image domain and neglecting the contribution of tissues in $\Omega \backslash \Omega_{\text{ROI}}$ to the projection measurements. Denoting by $\mathcal{J}$ the set of indexes in Ω_{ROI} the pixel-based ROI-only measurement model (i.e. $\Omega = \Omega_{\text{ROI}}$) was obtained by

$$m \approx A_{\text{ROI}} f_{\text{ROI}} + \epsilon \quad A_{\text{ROI}} = A(:, \mathcal{J}), \quad f_{\text{ROI}} = f(\mathcal{J}). \quad (5.5)$$

5.1.3 Results

Simulated local tomography data

The original jawbone phantom that was used to simulate the projection data is shown on the left in figure 5.2. The size of the original phantom was 300×300 pixels. The white circle in figure 5.2 denotes the Ω_{FS}, which was chosen as the ROI ($\Omega_{\text{FS}} = \Omega_{\text{ROI}}$). For each of 187 projections 300 line integrals were chosen leading to a projection data vector $m \in \mathbb{R}^{56100}$. Additive Gaussian white noise with a standard deviation of 1 % of the maximum of the computed projections was added to the data.

Figure 5.2 shows reconstructions with the multiresolution model. In the middle of figure 5.2 is a reconstruction using $J_{\text{out}} = J = 5$, the resulting number of unknown wavelet coefficients was 76 990. On the right in figure 5.2 is a reconstruction using $J_{\text{out}} = 1$ in $\Omega \backslash \Omega_{\text{ROI}}$ and $J = 5$ in Ω_{ROI}, the resulting number of unknown wavelet coefficients was 14 070. The corresponding computational times[1] for the MAP estimates were 3 minutes 18 seconds for the full wavelet reconstruction (i.e. $J = J_{\text{out}} = 5$) and 1 minutes 40 seconds for the multiresolution reconstruction with $J_{\text{out}} = 1$.

Figure 5.3 presents the ROI details of the original phantom (top left) and MAP reconstruction with the multiresolution method using $J_{\text{out}} = 1$ and $J = 5$ in ROI (top right).

In order to study quantitatively how the image quality is affected when J_{out} was decreased, the following relative reconstruction errors were computed

$$\delta_\Theta = \frac{\|f_{\text{true}} - f\|_{L^2(\Theta)}}{\|f_{\text{true}}\|_{L^2(\Theta)}} \times 100\% \tag{5.6}$$

either in the whole image domain ($\Theta = \Omega$) or in the ROI ($\Theta = \Omega_{\text{ROI}}$). Table 5.2 shows the relative errors both with respect to the original phantom and with respect to the full wavelet basis reconstruction (i.e. $J_{\text{out}} = J = 5$).

The bottom row in figure 5.3 presents the reconstructions obtained using the TV prior as the prior distribution. On the left of the bottom row is a MAP estimate with TV prior when the whole image domain Ω is covered with pixels (numbers of unknown pixels was 65 536) and on the right is a TV reconstruction obtained by using the TV ROI-only model of equation (5.5), the corresponding number of unknown pixels was 7484.

[1] All computations of Publication **I** were done with a desktop PC (model: Intel® Pentium® 4CPU 3.2 GHz with 4 GB RAM using Matlab® version 7.2(R2006a)

Table 5.1: The effect of the number (J_{out}) of scaling levels outside the ROI on the relative L^2-reconstruction errors (5.6) in the local tomography reconstructions with the multiresolution model. The number of scaling levels inside the ROI was $J = 5$ in all cases. n_w denotes the number of unknown wavelet coefficients in the inverse problem. δ^{W} indicates to the reconstruction error when the full wavelet basis reconstruction with $J_{\text{out}} = 5$ is used as the reference.

J_{out}	1	2	3	4	5
n_w	14070	15120	18800	31470	76990
δ_Ω (%)	36.6	36.6	36.4	36.4	36.3
$\delta_{\Omega^{\text{W}}}$ (%)	4.1	3.9	2.8	1.6	0
$\delta_{\Omega_{\text{ROI}}}$ (%)	24.0	23.9	23.8	23.7	23.7
$\delta_{\Omega^W_{ROI}}$ (%)	0.35	0.28	0.14	0.04	0

The relative reconstruction errors in Ω_{ROI} for the MAP estimates with TV prior were $\delta_{\Omega_{\text{ROI}}} = 23,7\%$ for the full pixel model and $\delta_{\Omega_{\text{ROI}}} = 123,7\%$ for the ROI-only model.

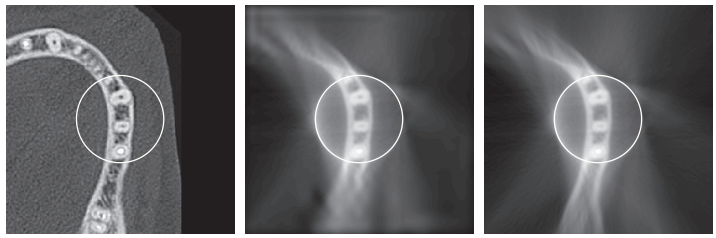

Figure 5.2: The MAP *estimates with the multiresolution model from the local tomography data of the jawbone phantom (original phantom on the left). The* ROI *is marked with a white circle. In all of the reconstructions, the number of scaling levels inside the* ROI *was $J = 5$. The number of the scaling levels outside the* ROI *was $J_{\text{out}} = 1$ (middle), and $J_{\text{out}} = 5$ (right).*

In publications [98,99] the multiresolution method was further tested by using only the approximation coefficients $c_{J_0,k}$ in $\Omega \backslash \Omega_{\text{ROI}}$ and by adding extra fillings into the jawbone phantom such that the fillings were just cut outside of the ROI. Typically such fillings cause severe artefacts in the reconstructions. The results are presented in figures 5.4 and 5.5.

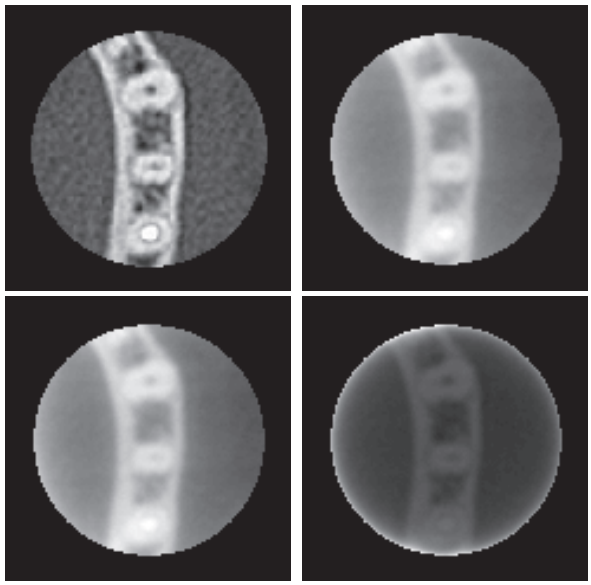

Figure 5.3: The ROI *details of the* MAP *estimates. On the top row; the original phantom (on the left), the* MAP *estimate with the multiresolution model using* $J_{\text{out}} = 1$ *scaling levels outside the* ROI *while the number of scaling levels inside the* ROI *was* $J = 5$ *(on the right). On the bottom row; the* MAP *estimate with the total variation prior (on the left) and the* MAP *estimate with the* TV *prior when the matrix* A_{ROI} *is used as the forward model (i.e., the whole image domain* $\Omega = \Omega_{\text{ROI}}$*) (on the right). (Note that each image has its own gray scale.)*

Experimental local tomography data

The projection radiographs of a jawbone specimen were acquired using a commercial intraoral X-ray detector Sigma® and a dental X-ray source Focus®. The projection images were collected using a similar CBCT geometry that is used in the commercial dental CBCT scanners. The measurement geometry of the experimental setup is illustrated in figure 5.6. One row of the projection images was used for the 2D reconstructions. The number of data was $m \in \mathbb{R}^{15272}$.

Figure 5.7 shows the multiresolution reconstructions computed from the experimental data. The ROI is marked by a white circle. On the left of figure 5.7 is a MAP reconstruction with $J_{\text{out}} = J = 6$ (number of unknown wavelet coefficients was 269 310) and on the right is a reconstruction with $J_{\text{out}} = 1$ in $\Omega \backslash \Omega_{\text{ROI}}$ and $J = 6$ in Ω_{ROI} (the number of unknown wavelet

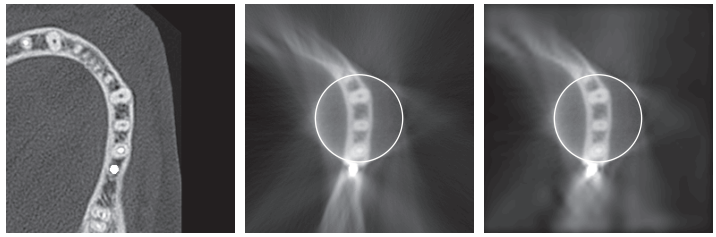

Figure 5.4: The MAP *estimates with the multiresolution model from the local tomography data of the jawbone phantom with fillings. The* ROI *is marked with a white circle. In all of the reconstructions, the number of scaling levels inside the* ROI *was* $J = 5$*. The number of the scaling levels outside the* ROI *was* $J_{\text{out}} = c_{J_0,k}$ *(in the middle), and* $J_{\text{out}} = 5$ *(on the right).*

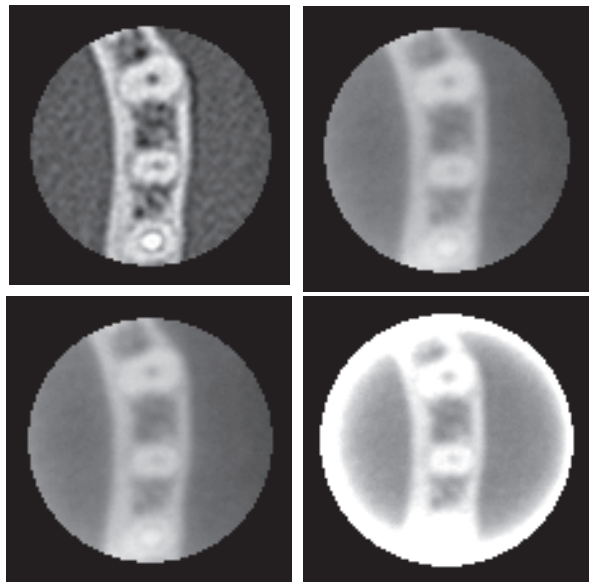

Figure 5.5: The ROI *details of the* MAP *estimates of figure 5.4. On the top row; the jawbone phantom (on the left), a* MAP *estimate with the multiresolution model using* $J_{\text{out}} = c_{J_0,k}$ *outside the* ROI *while the number of scaling levels inside the* ROI*was* $J = 5$ *(on the right). On the bottom row; a* MAP *estimate with the* TV *prior (on the left) and a* MAP *estimate with the* TV *prior when the matrix* A_{ROI} *is used as forward model (i.e., the whole image domain* $\Omega = \Omega_{\text{ROI}}$*) (on the right).*

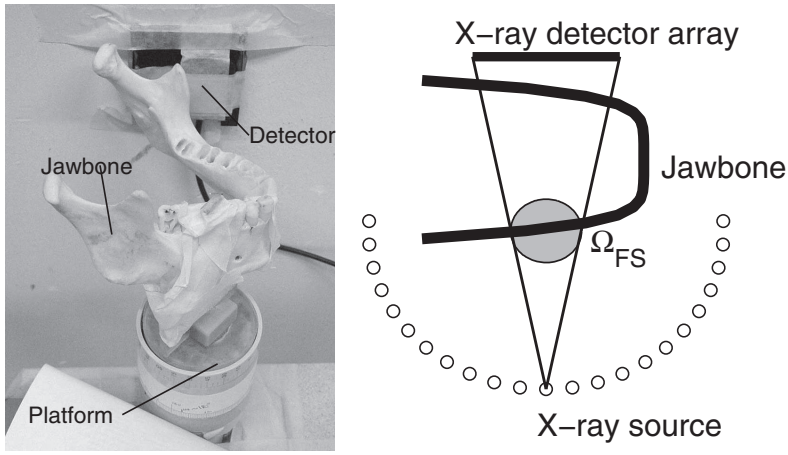

Figure 5.6: Left; a photograph of the measurement setup. Right; an illustration of the measurement geometry for the experiment with the jaw bone specimen.

coefficients was 29130). The corresponding computation times were 14 minutes 54 seconds and 2 minutes 14 seconds, respectively.

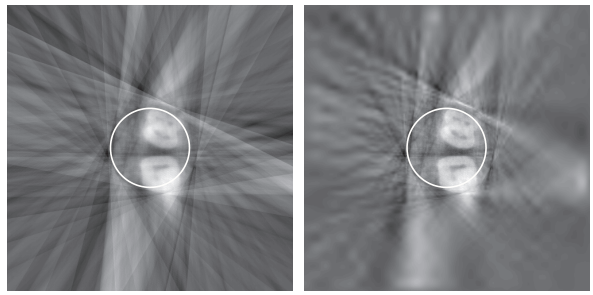

Figure 5.7: MAP *estimates with the multiresolution model from the measured local tomography data of the jaw bone specimen. In all of the reconstructions, the number of scaling levels inside the* ROI *was* $J = 6$*. The number of the scaling levels outside the* ROI *was* $J_{\text{out}} = 6$ *(left), and* $J_{\text{out}} = 1$ *(right). The* ROI *is marked with a white circle.*

Figure 5.8 shows the ROI details of the MAP estimates. On the left in figure 5.8 is a MAP estimate with $J_{\text{out}} = 1$ in $\Omega \backslash \Omega_{\text{ROI}}$ and $J = 6$ in Ω_{ROI}. In the middle is a MAP estimate with the full pixel-based TV model (5.4) (the number of the unknown pixels was 248 004) and on the right is a MAP estimate with TV prior using the ROI-only model (5.5) (the number of the unknown pixels was 19 044). The computation times were

12 minutes 36 seconds for the full pixel-based TV model and 2 minutes 14 seconds for the multiresolution model.

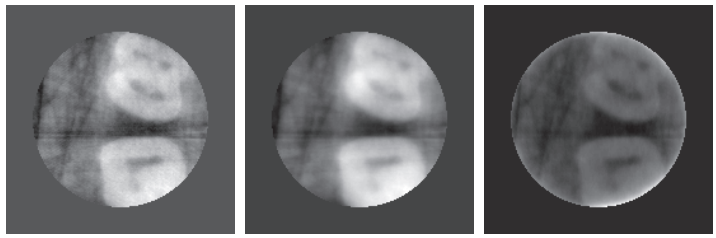

Figure 5.8: The ROI *details of the* MAP *estimates with the multiresolution model and with the total variation prior computed from the measured local tomography data of the jaw bone specimen. The reconstructions from left to right: a* MAP *estimate with the multiresolution method with* $J_{\text{out}} = 1$ *and* $J = 6$*, a* MAP *estimate with the* TV *prior and a* MAP *estimate with the* TV *prior using the* ROI*-only model, respectively. (Note that each image has its own gray scale.)*

5.2 PUBLICATION II: RECOVERING A SIGNAL FROM 1D CONVOLUTION DATA

Consider a linear model

$$m = Af + \epsilon$$

where $m \in \mathbb{R}^{n_m}$ is the measurement data vector, A is a linear operator modelling the measurements, f is a continuous function and $\epsilon \in \mathbb{R}^{n_m}$ is the measurement noise. In practice f is the unknown physical quantity that one wants to estimate. The computational solutions of this inverse problem are however, based on discrete models for approximating the continuous function $f \in X$. Hence, instead of studying the linear model $m = Af + \epsilon$ in a continuous setting, we study the finite computational model

$$m = Af_{(n)} + \epsilon, \tag{5.7}$$

where $f_{(n)} = \mathcal{T}_n f \in X$ and $\mathcal{T}_n : X \to X$ is a linear projection operator with n-dimensional range $\dim(\mathcal{T}_n(X)) = n < \infty$. Using Bayesian inversion methods and following Chapter 3 the Bayesian MAP estimate related to the inverse problem (5.7) is defined by

$$\pi_{\text{post}}(f_{(n)}^{\text{MAP}} \,|\, m) = \max\{\pi_{\text{post}}(f_{(n)} \,|\, m) \,:\, f_{(n)} \in \mathcal{T}_n(X)\}. \tag{5.8}$$

In the Bayesian framework it would be useful in practise if the Bayesian MAP and CM estimates were discretization-invariant since that would ensure that the MAP and CM estimates behave consistently at different resolutions. Furthermore, the same Bayesian prior model could be used regardless of the discretization level. Thus one of the objectives of Publication **II** was to derive a discretization-invariant Bayesian inversion model such that it would also be sparsity-promoting.

In regularization the use of $\|\cdot\|_1$ norms instead of $\|\cdot\|_2$ norms is known to promote sparsity [17, 23–25, 100]. In particular the TV norm is widely used in computational inversion. In the Bayesian framework, a sparsity-promoting approach can be obtained by constructing a prior model that contains a norm of $\|\cdot\|_1$ type. Publication **II** considered two such prior models; the TV prior $\pi_{\text{pr}}(f) = \exp(-\alpha \sum_{i=1}^{n} |(Ef)_i|)$, where E

is a periodic difference matrix defined by

$$E = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & & \cdots & 0 & -1 & 1 \\ 1 & & \cdots & 0 & 0 & -1 \end{bmatrix}$$

and the Besov B^1_{11} space prior $\pi_{\text{pr}}(f) = \exp(-\alpha \sum_{i=1}^n |(\mathcal{W}B^{-1}f)_i|)$, where $\mathcal{W}$ is a diagonal matrix containing the powers of 2 weights (see section 2.4).

The computation of the Bayesian MAP estimates with TV and B^1_{11} priors amounts to the following optimization problem

$$f^{\text{MAP}} = \arg\min_{f \in \mathbb{R}^n} \left\{ \frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha \sum_{i=1}^n |(Hf)_i| \right\} \tag{5.9}$$

where α is the prior parameter, $H = E$ for total variation prior and $H = \mathcal{W}B^{-1}$ for Besov B^1_{11} space prior. (Note that the vector $f^{\text{MAP}} \in \mathbb{R}^n$ then contains the values of the function $f^{\text{MAP}}_{(n)}(x)$, defined in (5.8), at the grid points.)

This optimization problem is non-differentiable and gradient-based optimization methods are not applicable. In Publication **II** the optimization problem of (5.9) was reformulated into a QP form and PD-IP method of Chapter 4 was used to compute the MAP estimates.

5.2.1 PD-IP method for sparsity-promoting Bayesian inversion

The derivation of the PD-IP method starts by reformulating (5.9) into a QP form. Thus denoting $Hf = h^+ - h^-$, where $h^+, h^- \geq 0$ the problem (5.9) becomes

$$\min_f \left\{ \frac{1}{2\sigma^2} f^T A^T A f - \frac{1}{\sigma^2} f^T A^T m + \alpha \mathbf{1}^T h^+ + \alpha \mathbf{1}^T h^- + \frac{1}{2\sigma^2} m^T m \right\}, \tag{5.10}$$

where **1** is a vector of all ones. Now denoting $z = \begin{bmatrix} f \\ h^+ \\ h^- \end{bmatrix}$, $Q = \begin{bmatrix} \frac{1}{\sigma^2}A^TA & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $d = \begin{bmatrix} -\frac{1}{2\sigma^2}A^Tm \\ \alpha\mathbf{1} \\ \alpha\mathbf{1} \end{bmatrix}$ and $\kappa = \frac{1}{2\sigma^2}m^Tm$, the minimization of (5.10) becomes

$$\begin{aligned} \min_z \quad & \left(\frac{1}{2}z^TQz + d^Tz + \kappa\right) \\ \text{s.t.} \quad & \mathcal{A}z = 0 \\ & z \geq \begin{bmatrix} -N \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

where $\mathcal{A} = \begin{bmatrix} \mathcal{W}B^{-1} & -I & I \end{bmatrix}$ for the B^1_{11} prior and $\mathcal{A} = \begin{bmatrix} E & -I & I \end{bmatrix}$ for the TV prior, I denoting the identity matrix. Based on theory (see [26] and **II**) f was unconstrained, i.e. $-\infty \leq f \leq \infty$, however, in practical computations we cannot set $f \geq -\infty$ therefore, we set $f \geq -N$ with $N >> 0$ so that the constraint on f was practically ineffective.

5.2.2 Materials and methods

The data used in the numerical computations in Publication **II** was acquired by taking A of equation (1.1) to correspond to a periodic convolution operator and by adding Gaussian white noise with a standard deviation $\sigma > 0$ s.t.

$$m = (Af)(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I).$$

$(Af)(x)$ was evaluated numerically using a finer grid than was used in any computations of the MAP estimates thus avoiding the inverse crime[1].

As a reference method MAP and CM estimates were computed using TV prior. The discretization-invariance of the MAP and CM estimates with TV prior were studied in [64]. In [64] the authors showed that the use of TV priors in Bayesian inversion fails of produce discretization-invariant MAP and CM estimates. In Publication **II** new numerical examples illus-

[1] The term *inverse crime* refers to obtaining unrealistically optimistic results. In practice this means that the discretization level is the same both in the numerical simulations and in inversion.

trating this discretization dilemma were provided.

With both the TV and B_{11}^1 models the MAP estimates were computed using the PD-IP method and the CM estimates were recovered using a Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. For more details to the computation of the CM estimates with the MCMC method see **II**.

A novel sparsity-based method for the selection of the prior parameter α was proposed in Publication **II**. The idea of the sparsity-based choice rule is that we want to promote estimates that have similar levels of sparsity (S) in a given wavelet basis as the *a priori* estimated sparsity level $(\hat{S})$ of the unknown. The sparsity-based choice rule can be summarized as follows:

1) take a finely spaced collection of parameters α such that
$$0 < \alpha^{(1)} < \cdots < \alpha^{(M)} < \infty$$
with M sufficiently large

2) compute the corresponding MAP estimates
$$f^{\mathrm{MAP}}(\alpha^{(1)}), \ldots, f^{\mathrm{MAP}}(\alpha^{(M)})$$
and calculate the numbers of wavelet coefficients S that are nonzero
$$(S(\alpha^{(1)}), \ldots, S(\alpha^{(M)}))$$
(in practice the absolute values of the wavelet coefficients are taken to be above a small and positive threshold level)

3) Select $\alpha = \alpha^i$ s.t. $S(\alpha) \approx \hat{S}$, using the data $(\alpha^i, S(\alpha^{(i)})), i = 1, \ldots, M$. $\hat{S}$ denotes the *a priori* estimated sparsity level of f in the wavelet domain

In Publication **II** the $\alpha^{(i)}$'s ranged in the interval $\left[\ 10^{-5}, 10^{8}\ \right]$, $M = 500$ and the small but positive threshold level for the wavelet coefficients was set to 10^{-6}.

5.2.3 Results

A piecewise linear target function (shown in figure 5.9) was used to simulate the measurement data such that the measurement vector consisted of 63 elements. Figure 5.9 shows the realization of the measurements m (on the right). For the treatment of the boundaries a periodic boundary condition was used in **II**.

In Publication **II** the convergence of estimates with the B_{11}^1 prior were studied numerically. To that end numerical results were computed using eight different discretization levels, namely $n = [64, 128, \ldots, 8192]$. The

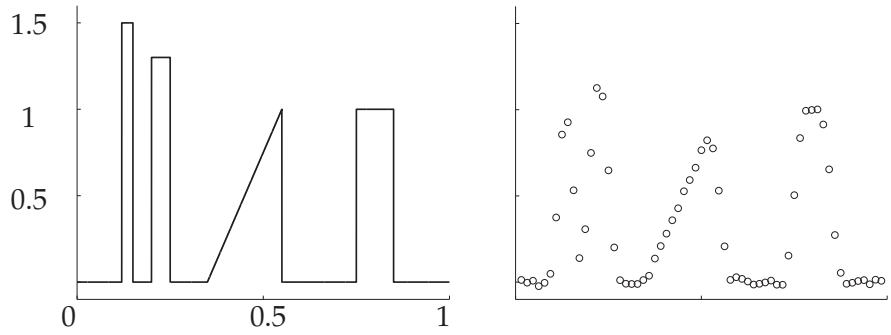

Figure 5.9: Left: the piecewise linear target function f. Right: the simulated measurement vector $m \in \mathbb{R}^{63}$ corrupted by white noise with a relative amplitude of 1%. The coordinate axis limits are the same in both of the plots.

theoretical existence for the limit $\lim_{n\to\infty} f_{(n)}^{\mathrm{MAP}}$ with B_{11}^1 prior was proven in **II** (for more details see **II**) and the proofs of the existence of the limit $\lim_{n\to\infty} f_{(n)}^{\mathrm{CM}}$ were given in [26].

Based on theory the Daubechies 7 wavelets would have provided a valid basis for the B_{11}^1 space but the numerical results showed that Bayesian estimates using B_{11}^1 prior with Daubechies 7 wavelets failed to preserve edges in the reconstructions, whereas the use of Haar wavelets yielded convergent discretization-invariant estimates that were sparse and edge-preserving, hence Haar wavelets were used in the computations.

The value for $\hat{S}$ was determined by computing the numbers of nonzero wavelet coefficients of the target function (presented on the left in figure 5.9) in Haar wavelet basis at the level $n = 256$, thus representing an optimal case where the number of nonzero wavelet coefficient is exactly known. The resulting $\hat{S}$ was $\hat{S} = 82$.

The proposed α-choice rule was tested using three data sets with different standard deviations for the noise; namely $\sigma = 5\%, \sigma = 1\%$ and $\sigma = 0.1\%$. The results with $\sigma = 1\%$ are presented in figure 5.10, this data was used in further computations. According to the proposed sparsity-based choice rule with $\hat{S} = 82$ the prior parameter α, determined at the resolution $n = 256$, resulted $\alpha = 5.39$, as illustrated in figure 5.10. With $\alpha = 5.39$ the number of nonzero wavelet coefficients in the corresponding MAP estimate was $S(5.39) = 44$ $(n = 256)$. The selected $\alpha = 5.39$ was

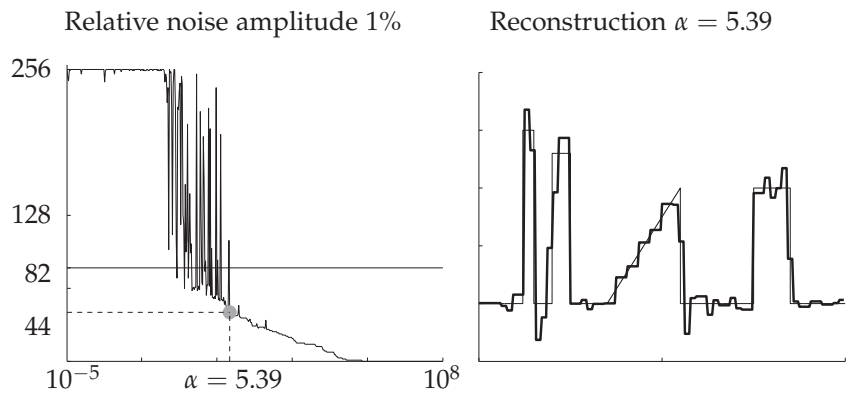

Figure 5.10: The Sparsity-based choice of the regularization parameter. Left: plots of numbers of nonzero wavelet coefficients in the MAP *estimates computed with various values of parameter α. Results for $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = 1\%$ are presented. Right: the original function (thin line) and the* MAP *estimate (thick line) using the selected α.*

then used to compute the MAP and CM estimates with B^1_{11} prior at all discretization levels ($64 \leq n \leq 8192$). The results are presented in figure 5.11.

In order to study the convergence of the MAP and CM estimates quantitatively, relative reconstruction errors were computed with respect to the estimates computed at the discretization level $n = 8192$. See table 5.2 for the results. From the table 5.2 and from the superposition images of figure 5.11 it can be seen that the Bayesian MAP and CM estimates with B^1_{11} prior are discretization-invariant.

Also the sparsity of the estimates was studied by counting the percentages of scaled wavelet coefficients $\alpha 2^{j/2} w_{j,k}$ that had $|\alpha 2^{j/2} w_{j,k}| < r$ with the prior probability of 10^{-6}. This criterion gave $r = 10^{-6}$ as the tolerance for the numerical zero when measuring sparsity. The results are presented in table 5.2. As can be seen from table 5.2, selecting the prior parameter α according to the proposed choice rule leads to MAP estimates that are quite sparse and have roughly the same number of non-zero wavelet coefficients for all n.

In [64] it was shown that Bayesian estimates with TV prior are not discretization-invariant, this was also illustrated in Publication **II** by computing the MAP estimates with TV prior and by selecting the prior parameter α in two ways: (i) $\alpha = 89 \ \forall \ n$ and (ii) $\alpha = 3.93\sqrt{n}$. The superpositions of MAP and CM estimates with α's as stated in (i) and (ii) are

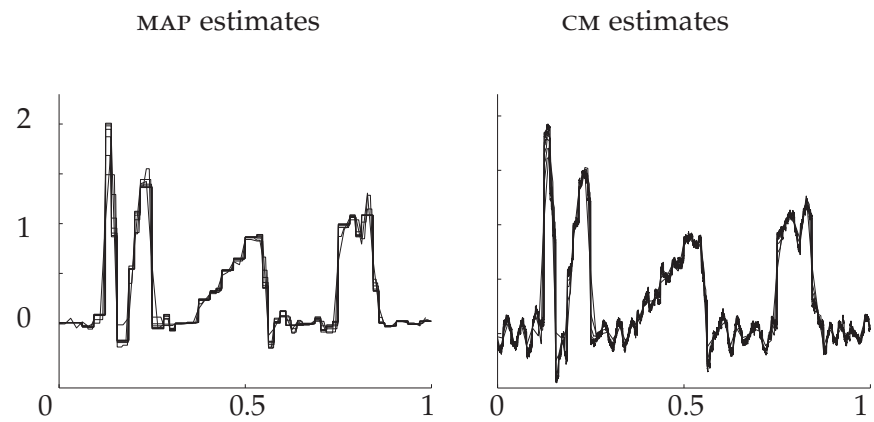

Figure 5.11: Superposition of the CM *and* MAP *estimates. Note the quite close agreement of the estimates over a significant range of discretization levels (*$64 \leq n \leq 8192$*). This is due to the convergence properties of the discretization-invariant Besov space prior.*

Table 5.2: The relative errors ($\delta_{\mathrm{MAP/CM}}$*) and the sparsities (*$S_{\mathrm{MAP/CM}}$*) of* MAP *and* CM *estimates with Besov space* $B^1_{11}(\mathbb{T})$ *prior using Haar wavelet basis. The relative errors are computed with respect to the estimate at resolution* $n = 8192$ *and the numbers* S_{MAP} *and* S_{CM} *present the numbers of nonzero scaled wavelet coefficients* $\alpha 2^{j/2} w_{j,k}$ *in the estimates computed using* 10^{-6} *as a tolerance value.*

n	64	128	256	512	1024	2048	4096	8192
δ_{MAP}	0.28	0.31	0.23	0.16	0.10	0.07	0.03	0
δ_{CM}	0.25	0.25	0.20	0.15	0.11	0.09	0.07	0
S_{CM}	64	128	256	512	1024	2048	4095	8191
S_{MAP}	47	41	44	40	42	41	41	42

presented in figure 5.12. From the figure 5.12 it can be seen how using a constant parameter $\alpha \in \mathbb{R}$ leads to convergent MAP estimates and divergent CM estimates. On the other hand, a choice of the form $\alpha = \alpha_0 \sqrt{n}$ is the only possibility to achieve convergent CM estimates. However, such a choice leads to MAP estimates converging to zero and to CM estimates

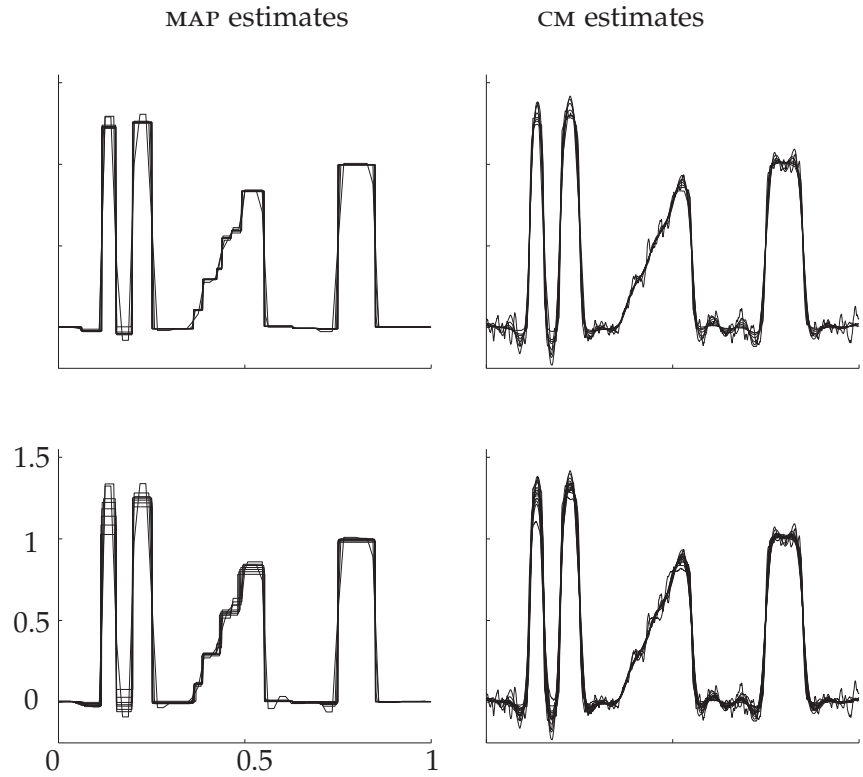

Figure 5.12: Superposition of the MAP *(left column) and* CM *(right column) estimates with the* TV *prior. Top row: Superpositions of the reconstructions using* $\alpha = 89\ \forall n$. *Bottom row: Superpositions of the reconstructions using* $\alpha = 3.93\sqrt{n}$. *The coordinate axes limits are the same in all of the plots.*

that are smooth in the high resolution limit, thus failing to preserve edges.

5.3 PUBLICATION III: RECOVERING DIFFUSION COEFFICIENTS OF A 1D COUPLED PARABOLIC SYSTEM

Consider the following linear parabolic system

$$\begin{cases} \partial_t u_1 - \partial_x(c_1 \partial_x) u_1 &= a_{11}u_1 + a_{12}u_2 \text{ in } \Omega \times (0,T), \\ \partial_t u_2 - \partial_x(c_2 \partial_x) u_2 &= a_{21}u_1 + a_{22}u_2 \text{ in } \Omega \times (0,T), \\ u_i(t,x) &= 0 \text{ on } \partial\Omega \times (0,T),\ i = 1,2, \\ u_i(0,x) &= u_{0,i}(x) \text{ in } \Omega,\ i = 1,2, \end{cases} \tag{5.11}$$

where $\Omega =]0,1[$, $T > 0$, $c_1, c_2 \geq \eta > 0$ are the diffusion coefficients and $a_{i,h}$ are the reaction coefficients. The diffusion coefficients $c_i(x), i = 1,2$ are restricted to be time-independent and to belong to a set of positive piecewise smooth functions in Ω and the coupling coefficients (a_{12} and a_{21}) are assumed to satisfy the conditions

$$|a_{21}| \geq \eta > 0 \text{ and } |a_{12}| \geq \eta > 0 \text{ in } \omega,$$

where ω is a non empty arbitrarily fixed open interval $\omega \subset\subset \Omega$.

The inverse problem is the following. Determine the discontinuous diffusion coefficients, $c_1(x)$ and $c_2(x)$, related to the system (5.11), from the observation data $(U_2(\mathbf{u_0};\mathbf{c})|_{(t_0,T)\times\omega}, \mathbf{U}(\mathbf{u_0};\mathbf{c})|_{t=\theta})$, where $\mathbf{c} = (c_1, c_2)$, $\mathbf{u_0} = (u_{0,1}, u_{0,2})$, $t_0 \in (0,T)$, $\theta \in (t_0,T)$ and $\mathbf{U}(\mathbf{u_0};\mathbf{c}) = (U_1(\mathbf{u_0};\mathbf{c}), U_2(\mathbf{u_0};\mathbf{c}))$.

In the theoretical part of Publication **III** Carleman estimates for the inverse problem related to the system (5.11) with observations of only one component (i.e. $U_2(\mathbf{u_0};\mathbf{c})|_{(t_0,T)\times\omega}$) were given. These Carleman estimates were then used to provide the main stability results for recovering $c_1(x)$ and $c_2(x)$ from the observation data $(U_2(\mathbf{u_0};\mathbf{c})|_{(t_0,T)\times\omega}, \mathbf{U}(\mathbf{u_0};\mathbf{c})|_{t=\theta})$. For more details about the theoretical results see **III**.

In the numerical part of Publication **III** the discontinuous diffusion coefficients were recovered using similar PD-IP method as presented in Chapter 4.

5.3.1 PD-IP method for recovering $c_1(x)$ and $c_2(x)$

According to the main theoretical results of Publication **III** (see Theorem 1.1, Corollary 1 and section 2 in **III**) the discontinuous diffusion coefficients $c_1(x)$ and $c_2(x)$ could be identified simultaneously by solving the following constrained optimization problem

$$\min_{u_2} \frac{1}{2} \left\{ \|\partial_t(u_2 - \tilde{u}_2)\|^2_{L^2((t_0,T)\times\omega)} + \|\partial_{tt}(u_2 - \tilde{u}_2)\|^2_{L^2((t_0,T)\times\omega)} \right\} \tag{5.12}$$

subject to

$$\begin{cases} \partial_t u_1 - \partial_x(c_1 \partial_x u_1) = a_{11}u_1 + a_{12}u_2 & \text{in} \quad \Omega \times (0,T) \\ \partial_t u_2 - \partial_x(c_2 \partial_x u_2) = a_{21}u_1 + a_{22}u_2 & \text{in} \quad \Omega \times (0,T) \\ u_i(t,x) = 0 & \text{on} \quad \partial\Omega \times (0,T) \\ u_i(0,x) = u_{0,i}(x) & \text{in} \quad \Omega \\ \mathbf{u}(\theta,x) = \tilde{\mathbf{u}}(\theta,x) & \text{in} \quad \Omega \end{cases} \tag{5.13}$$

with $c_i^{\min} \leq c_i \leq c_i^{\max}, i = 1,2,\ \theta \in [t_0, T]$ and $\tilde{\mathbf{u}} = (\tilde{u}_1, \tilde{u}_2)$, where $\tilde{u}_2$ denotes the measurement (i.e. $U_2(\mathbf{u_0}; \mathbf{c})|_{(t_0,T)\times\omega}$).

Note that the functional in (5.12) depends only on u_2 whereas the constraints (5.13) depend on both $\mathbf{u} = (u_1, u_2)$ and $\mathbf{c} = (c_1, c_2)$ and the aim is to recover $\mathbf{c} = (c_1, c_2)$. This implies that (5.13) becomes non-linear in view of the numerical computations. Thus for the numerical computations a linearized version of (5.13) was considered having both $\mathbf{u} = (u_1, u_2)$ and $\mathbf{c} = (c_1, c_2)$ as unknowns in the problem. The linearized version of the constraints is written

$$\begin{cases} \partial_t u_1 - \partial_x(\tilde{c}_1\partial_x)u_1 - a_{11}u_1 - a_{12}u_2 - \partial_x(c_1\partial_x)\tilde{u}_1 = -\partial_x(\tilde{c}_1\partial_x)\tilde{u}_1 & \text{in } \Omega_t, \\ \partial_t u_2 - \partial_x(\tilde{c}_2\partial_x)u_2 - a_{21}u_1 - a_{22}u_2 - \partial_x(c_2\partial_x)\tilde{u}_2 = -\partial_x(\tilde{c}_2\partial_x)\tilde{u}_2 & \text{in } \Omega_t, \\ u_i(t,x) = 0 \quad \text{on } \partial\Omega \times (0,T), i = 1,2, \\ u_i(0,x) = u_{0,i}(x) \quad \text{in } \Omega \\ \mathbf{u}(\theta,x) = \tilde{\mathbf{u}}(\theta,x) \quad \text{in } \Omega \end{cases} \tag{5.14}$$

where $\Omega_t = \Omega \times (0,T)$. Denoting the unknowns (u_1, u_2, c_1, c_2) by a vector

$$z = \begin{bmatrix} u_1 \\ u_2 \\ c_1 \\ c_2 \end{bmatrix},$$

the constraints (5.14) are written in a matrix form such that $\mathcal{A}z = b$ is equivalent to (5.14) with b denoting the right-hand side of (5.14).

Now denoting $\mathcal{D} = \partial_t$ and $\mathcal{D}_2 = \partial_{tt}$ the functional in (5.12) can be rewritten as

$$\min_z \quad \left\{ \tfrac{1}{2}u_2^T(\mathcal{D}^T\mathcal{D} + \mathcal{D}_2^T\mathcal{D}_2)u_2 - u_2^T(\mathcal{D}^T\mathcal{D} + \mathcal{D}_2^T\mathcal{D}_2)\tilde{u}_2 + \tfrac{1}{2}\tilde{u}_2^T(\mathcal{D}^T\mathcal{D} + \mathcal{D}_2^T\mathcal{D}_2)\tilde{u}_2 \right\}.$$

Further, denoting

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathcal{D}^T\mathcal{D} + \mathcal{D}_2^T\mathcal{D}_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ -(\mathcal{D}^T\mathcal{D} + \mathcal{D}_2^T\mathcal{D}_2) \\ 0 \\ 0 \end{bmatrix}$$

and $\kappa = \frac{1}{2}\tilde{u}_2^T(\mathcal{D}^T\mathcal{D} + \mathcal{D}_2^T\mathcal{D}_2)\tilde{u}_2$ the optimization problem related to equa-

tions (5.12) and (5.13) becomes

$$\begin{array}{ll} \min_z & \left(\frac{1}{2}z^T Q z + d^T z + \kappa\right) \\ \text{s.t} & \mathcal{A}z = b \\ \begin{bmatrix} -N \\ -N \\ c_1^{\min} \\ c_2^{\min} \end{bmatrix} & \leq z \leq \begin{bmatrix} N \\ N \\ c_1^{\max} \\ c_2^{\max} \end{bmatrix} \end{array}$$

where the upper and lower bounds for the primal variable z are related to the following inequalities; $0 < c_i^{\min} \leq c_i \leq c_i^{\max} < C$, $C > 0$ and $-N \leq u_i \leq N$, $i = 1,2$ with some $N >> 0$ so that the constraints on u_i's are practically ineffective.

5.3.2 Results

The observation data used in the numerical computations was simulated by setting $T = 1$ and adding Gaussian white noise with a standard deviation of 1 % of the maximum of the noiseless measurements to the data. A more dense grid was used for the generation of the data than was used in any of the computations.

The problem (5.14) was discretized by using the explicit Euler's method for the approximation of the time derivatives and finite difference (FD) method for the approximation of the spatial derivatives. In the following N_x and N_t denote the numbers of steps in space and in time, respectively.

Two different realizations of the discontinuous diffusion coefficients were considered having different locations of ω's (ω is a subdomain of Ω where the observation data $U_2(\mathbf{u_0}; \mathbf{c})|_{(t_0,T)\times\omega}$ is measured). The piecewise smooth targets ($\widetilde{c}_1(x)$ and $\widetilde{c}_2(x)$), as well as the subdomains (ω's) are presented in figure 5.13.

For the numerical computations of recovering $c_1(x)$ and $c_2(x)$, N_x and N_t were set to $N_x = 128$ and $N_t = 80$, respectively. According to the theory (see **III** and [101]) the $\theta \in [t_0, T]$ was selected such that the requirement $|\partial_x \tilde{\mathbf{u}}(\theta, x)| \geq \delta > 0$ in $\Omega \backslash \omega$ was satisfied.

In order to study the results quantitatively relative errors

$$\delta_{\mathrm{c_i}} = \frac{\|\widetilde{c}_i(x) - c_i(x)\|_{L^2}}{\|\widetilde{c}_i(x)\|_{L^2}}, \qquad i = 1,2$$

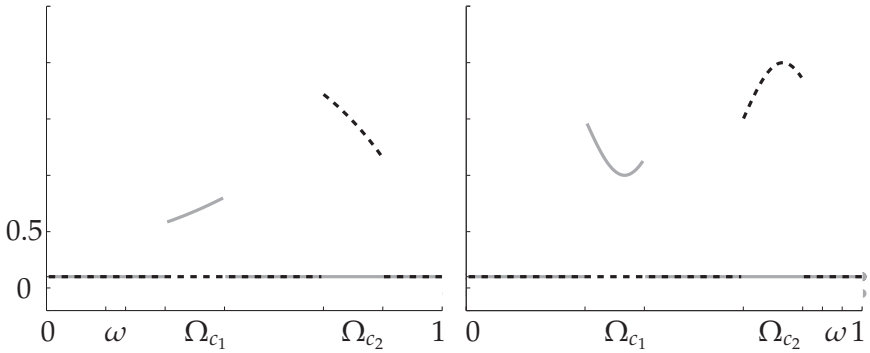

Figure 5.13: . The target coefficient $\widetilde{c}_1(x)$ and $\widetilde{c}_2(x)$. $\widetilde{c}_1(x)$ is denoted by a solid gray line and $\widetilde{c}_2(x)$ by a dashed back line. The coordinate axis limits are the same in both of the plots.

were computed with respect to the true targets.

Figure 5.14 shows the results for the first test case. The relative errors were $\delta_{c_1} = 0.1541$ and $\delta_{c_2} = 0.1127$ for $c_1(x)$ and $c_2(x)$, respectively. The corresponding results for the second test case are presented in figure 5.15. The corresponding relative errors were $\delta_{c_1} = 0.1443$ and $\delta_{c_2} = 0.0802$.

5.4 PUBLICATION IV: RECONSTRUCTION OF 2D X-RAY IMAGES FROM SPARSE PROJECTION DATA

In medical X-ray tomography it is necessary to keep the radiation dose to patients as low as possible but at the same time accurate reconstructions are needed for the diagnostic and surgical purposes. The radiation dose can be reduced, for example, by truncating the projection images (which leads to a local tomography problem) or by reducing the angular sampling rate of the projections. However, both of these procedures lead to problems that are ill-posed.

In Publication **IV** the study concentrated on reconstructing X-ray images from tomographic projection data collected from a full view-angle but from only few directions. The issue of ill-posedness was tackled by using Bayesian inversion methods. As the prior model Besov B^1_{11} space prior with Haar wavelet basis was considered since it allowed to compute Bayesian MAP estimates that preserved edges, were discretization-invariant and had sparse representations on the wavelet domain, see [26]

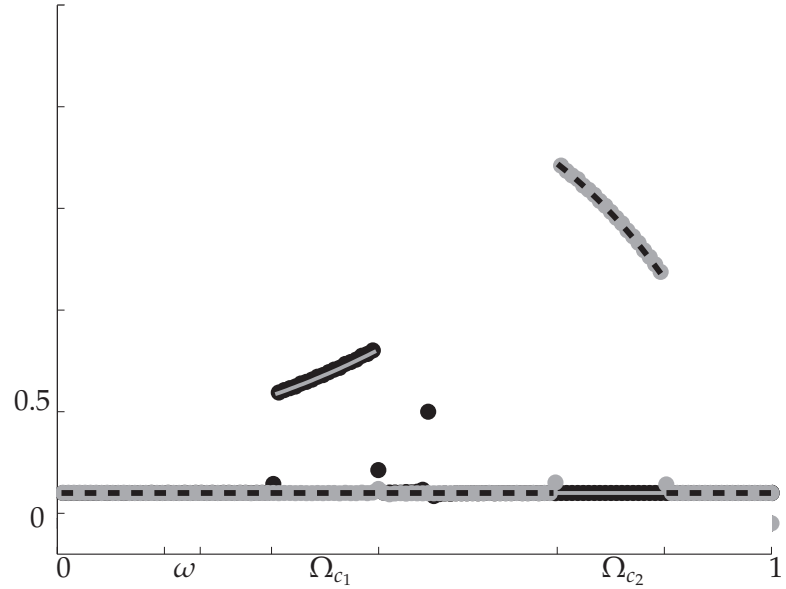

Figure 5.14: Reconstructions of $c_1(x)$ and $c_2(x)$. The results for $c_1(x)$ are denoted by black dots and the results for $c_2(x)$ are denoted by gray dots. $\widetilde{c}_1(x)$ denoted by a solid gray line and $\widetilde{c}_2(x)$ by a dashed black line. The relative error for $c_1(x)$ and $c_2(x)$ are $\delta_{c_1} = 0.1541$ in Ω and $\delta_{c_2} = 0.1127$ in Ω, respectively.

and **II**.

5.4.1 PD path following IP method for 2D image reconstruction problem with sparse angle data

The computation of the Bayesian MAP estimate in 2D using B^1_{11} prior amounts to the following optimization problem

$$f^{\mathrm{MAP}} = \arg\min_{f \geq 0} \left\{ \frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha \sum_{\nu=1}^{n} |(B^{-1}f)_\nu| \right\}, \tag{5.15}$$

where A is a matrix that implements the transformation from the pixel values to the projection data, $f \in \mathbb{R}^n$ is a vector that represent the X-ray attenuation function, $m \in \mathbb{R}^{n_m}$ is the projection data vector, α is the prior parameter and B^{-1} is a matrix that denotes the direct wavelet transform, i.e. $w = B^{-1}f$ when the wavelets are ordered to a sequence and re-numbered by an index $\nu \in \mathbb{Z}_+$.

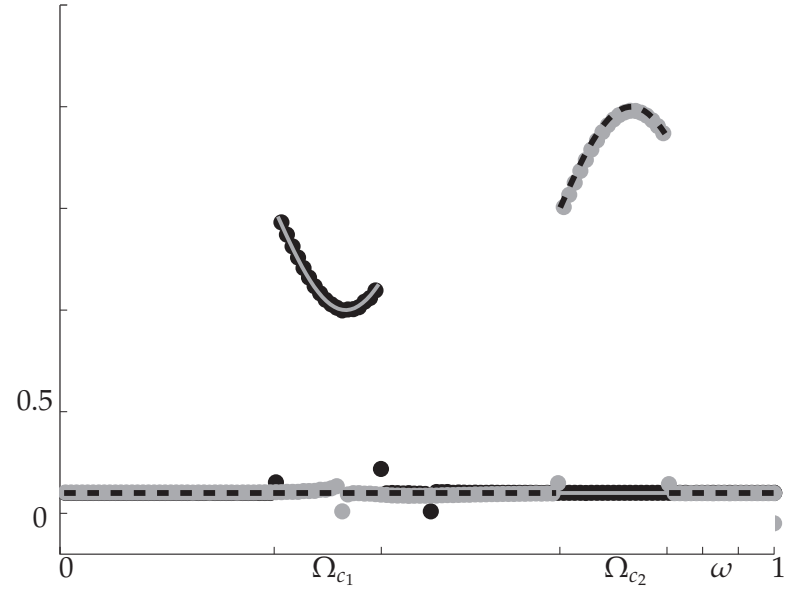

Figure 5.15: . Reconstructions of $c_1(x)$ and $c_2(x)$. The results for $c_1(x)$ are denoted by black dots and the results for $c_2(x)$ are denoted by gray dots. $\widetilde{c}_1(x)$ is denoted by a solid gray line and $\widetilde{c}_2(x)$ by a dashed back line. The relative errors for $c_1(x)$ and $c_2(x)$ are $\delta_{c_1} = 0.1443$ in Ω and $\delta_{c_2} = 0.0802$ in Ω, respectively.

Denoting $B^{-1}f = h^+ - h^-, h^+, h^- \geq 0$ the minimization of (5.15) can be similarly reformulated into a QP form as presented in section 5.2. Thus the MAP estimates can be computed by solving the following constrained optimization problem

$$\begin{aligned} \min_z \quad & \left(\frac{1}{2}z^T Q z + d^T z + \kappa\right) \\ \text{s.t.} \quad & \mathcal{A}z = 0 \\ & z \geq 0, \end{aligned}$$

where $z = \begin{bmatrix} f \\ h^+ \\ h^- \end{bmatrix}, Q = \begin{bmatrix} \frac{1}{\sigma^2}A^TA & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \end{bmatrix}, d = \begin{bmatrix} -\frac{1}{2\sigma^2}A^Tm \\ \alpha\mathbf{1} \\ \alpha\mathbf{1} \end{bmatrix}, \kappa = \frac{1}{2\sigma^2}m^Tm$ and $\mathcal{A} = \begin{bmatrix} B^{-1} & -I & I \end{bmatrix}$.

5.4.2 Materials and methods

Two different test cases of sparse angle tomography in 2D were considered; (i) simulated data and (ii) experimental data.

The simulated data was computed using Shepp-Logan phantom. A Gaussian white noise with a standard deviation of 1% of the maximum of the noiseless data was added to the projection data. The data was simulated such that the inverse crime was avoided (see section 5.2 for the explanation of inverse crime). Five different data sets were considered consisting of 148, 74, 37, 19 and 13 projections with uniform angular sampling from angular spanning of 180°.

The experimental data was measured using a custom-build μCT device supplied by Phoenix-Xray Systems + Service GmbH (Germany). A walnut was used as a target object. A set of 90 projection images were acquired over a 180° rotation with uniform angular steps of 2° between the projections. For this study only the middle cross-section of the walnut was selected. From the measured data four data sets were picked consisting of 90, 45, 30 and 15 uniformly sampled projections from a total opening angle of 180 degrees.

As a reference for the Bayesian method Filtered Back Projection (FBP) reconstructions were computed. The FBP reconstructions were computed using the Matlab®'s iradon function with a Hamming window. In addition reconstruction using Tikhonov regularization were computed from the simulated data. The X-ray image reconstruction using classical Tikhonov regularization is equivalent to the following optimization problem

$$f_{\text{Tik}} = \min_{f \in \mathbb{R}^n} \left(\frac{1}{2\sigma^2} \|Af - m\|_2^2 + \alpha_{\text{M}} \|f\|_2^2 \right), \quad f \geq 0. \tag{5.16}$$

The regularization parameter α_{M} in equation (5.16) was selected using Morozov's discrepancy principle. The minimization problem of (5.16) was solved using similar PD-IP method as presented in section 4.2.

For the selection of the prior parameter α for the Bayesian MAP estimates Publication **IV** proposed a method called *S-curve* method. The S-curve method is based on the α-choice rule previously proposed in the Publication **II**. The S-curve method of **IV** can be summarized as follows

1) Take a small collection of α's ranging on the interval $[0, \infty]$ s.t
$$0 < \alpha^{(1)} < \cdots < \alpha^{(i)} < \cdots < \alpha^{(M)} < \infty$$

2) Compute the corresponding estimates
$$f^{\text{MAP}}(\alpha^{(1)}), \ldots, f^{\text{MAP}}(\alpha^{(i)}), \ldots, f^{\text{MAP}}(\alpha^{(M)})$$

and calculate the numbers of the elements in the set (#{·})

$$S(\alpha) = \#\{\nu : 1 \leq \nu \leq n \quad | \quad |(B^{-1}f(\alpha))_\nu| > \eta\}$$

3) Fit a smooth interpolating curve to the data $\{\alpha^{(i)}, S(\alpha^{(i)})\}, i \in [1, M]$ to obtain the S-curve

4) Use the fitted curve to select α s.t.

$$S(\alpha) = \hat{S},$$

where $\hat{S}$ is the *a priori* estimated sparsity level of the unknown f.

In Publication **IV** M and η were set to $M = 20$ and $\eta = 10^{-6}$, respectively. Due to the fact that MAP estimates with Besov B^1_{11} prior are discretization-invariant (discussed in Publication **II**) the selection of the prior parameter could be done on a coarser grid than was used in the computations of the final MAP estimates.

The *a priori* estimates for the sparsity of f, i.e. $\hat{S}$, in the two test cases were acquired as follows. For the simulated data $\hat{S}$ was computed using Shepp-Logan phantom at resolution 128×128 with $\eta = 10^{-6}$, corresponding to the situation where $\hat{S}$ is exactly known. In order to estimate $\hat{S}$ in the experimental data case, three walnuts were split carefully in half and digital photographs were taken of the halfs. The digital photographs were then resized into size 128×128 and values $\hat{S}_1, \hat{S}_2, \hat{S}_3$ were computed with a threshold level of $\eta = 10^{-6}$. The average of $\hat{S}_1, \hat{S}_2$ and $\hat{S}_3$ was then used as the $\hat{S}$. This corresponded to a situation where $\hat{S}$ cannot be known exactly and needs to be estimated. By computing the $\hat{S}_1, \hat{S}_2$ and $\hat{S}_3$ of the photographs we thus simulated the use of high-resolution CT atlases for the estimation of an approximation of $\hat{S}$.

The performance to the estimates (MAP, FBP and Tikhonov) were studied as the projection directions were decreased.

5.4.3 Results

The selection of α's for each data set (five for the simulated data and four for the experimental) was done on a 128×128 grid and the actual reconstructions were computed on a 256×256 grid. Figure 5.16 shows the sparsity-based choice of α for one of the simulated data sets. The data set presented in figure 5.16 contains 148 uniformly distributed projection angles. Similarly figure 5.18 shows the α choice rule for one of the experimental data sets (90 uniformly spaced projection angles from a total opening angle of 180°).

In real life when sparse angle data has been measured the prior parameter α needs to be selected using that data. Therefore in **IV** the α's

Table 5.3: The relative errors of the MAP ,FBP *and Tikhonov reconstructions. The relative errors are computed with respect to the original Shepp-Logan phantom at resolution* 256×256.

	148	74	37	19	13
δ_{MAP}	0.10	0.12	0.12	0.13	0.17
δ_{FBP}	0.14	0.16	0.27	0.51	0.78
δ_{Tik}	0.10	0.12	0.13	0.23	0.29

were computed separately for all of the data sets. The chosen α's for the MAP estimates with respect to each data sets are denoted in figure 5.19 for the simulated case and in figure 5.21 for the experimental case. The corresponding MAP, FBP and Tikhonov reconstructions in the simulated case with data sets consisting of 148, 37 and 13 projection angles, respectively, are presented in figure 5.19. The MAP and FBP reconstructions computed from the experimental data are presented in figure 5.21.

In order to study the MAP,FBP and Tikhonov reconstruction quantitatively relative reconstruction errors were computed with respect to the original Shepp-Logan phantom at resolution 256×256. The results are presented in table 5.3.

In **IV** also the L-curve method was applied to a simulated data set consisting of 37 projections. The MAP estimates were computed with 400 values of α ranging on interval $[10^{-4}, 10^{7}]$. The resulting reconstruction with α selected using the L-curve method is presented on the left in figure 5.17. Reconstruction from the same data set with α selected using the S-curve method and the Morozov discrepancy principle are also presented in figure 5.17. The results for the L-curve method are presented here only for one sparse angle data set but the results with other data sets were similar.

In **IV** we also studied what would happen if $\hat{S}$, that is used to determine α, is estimated from X-ray images of healthy subjects and the data at hand is measured from a subject with anomalies (such as tumor). The results are presented in figure 5.20.

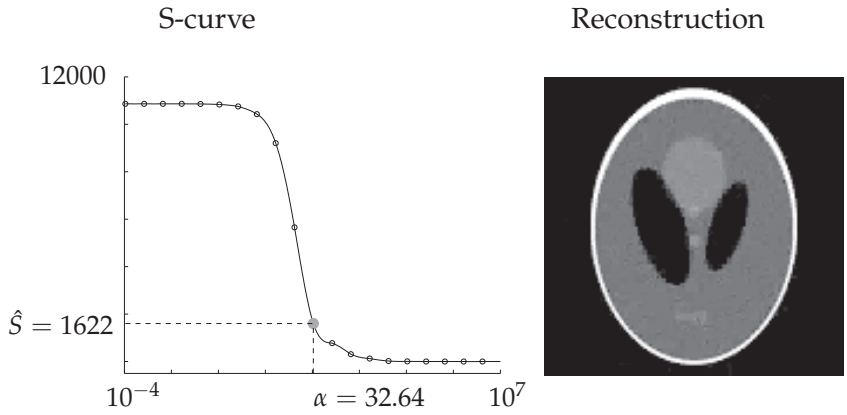

Figure 5.16: Sparsity-based choice of the prior parameter α. Left: The numbers of nonzero wavelet coefficients in Bayesian MAP *estimates computed with 20 values of α ranging in the interval* $[10^{-4}, 10^{7}]$ *(denoted by ∘) and a plot of the interpolation curve used to determine the value of α. Right: The reconstruction (*$n = 128 \times 128$*) using the selected α.*

Figure 5.17: . L-curve method, S-curve method and Morozov discrepancy principle applied to data set with 37 projections. Reconstructions with selected alphas from left to right are: the MAP *estimate with α selected using the L-curve method* $\alpha_{\mathrm{L}} = 33022$*, the* MAP *estimate with α selected using the S-curve method* $\alpha = 15.24$ *and the Tikhonov reconstruction with the α selected using the Morozov discrepancy principle* $\alpha_{\mathrm{M}} = 6.68$*, respectively. The reconstruction resolution is* 128×128*.*

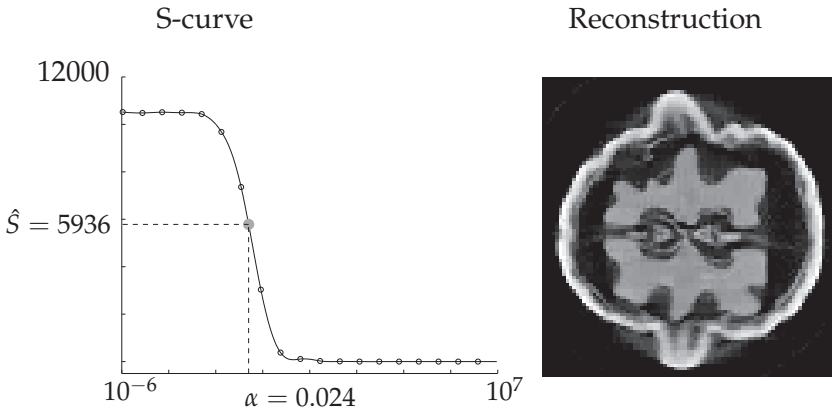

Figure 5.18: Sparsity based choice of α. Left: The numbers of nonzero wavelet coefficients in Bayesian MAP *estimates computed with 20 values of α ranging in the interval* $[10^{-6}, 10^{7}]$ *(denoted by ∘) and plot of the interpolation curve used to determine the value of α. Right: The* MAP *estimate (resolution* 128×128*) using the selected value of α.*

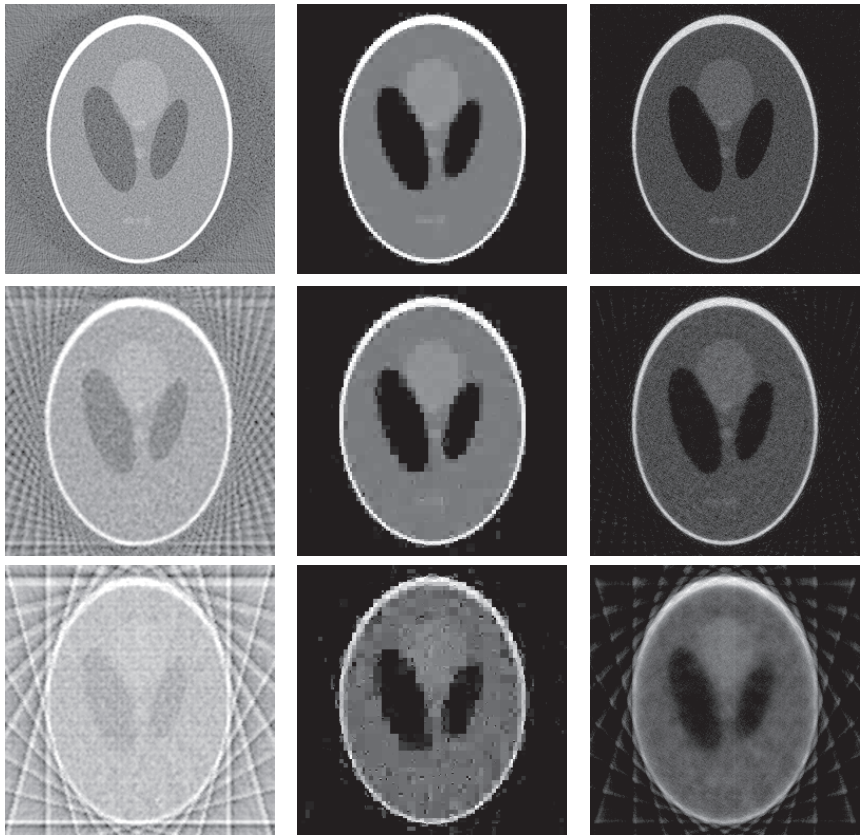

Figure 5.19: Sparse-angle tomography reconstructions of the Shepp-Logan phantom. Reconstruction of the FBP *(left column),* MAP *estimates with B_{11}^1 prior (middle column) and Tikhonov regularization (right column). The number of projections and the values of the parameters α are. Top row: 148 projections, $\alpha = 32.64$, and $\alpha_{\mathrm{M}} = 0.0586$. Middle row: 37 projections, $\alpha = 15.24$ and $\alpha_{\mathrm{M}} = 0.0092$. Bottom row: 13 projections, $\alpha = 5.18$ and $\alpha_{\mathrm{M}} = 0.0176$. α denotes the prior parameter selected using the S-curve method and α_{M} denotes the regularization parameters selected using the Morozov discrepancy principle. The reconstruction resolution is 256×256.*

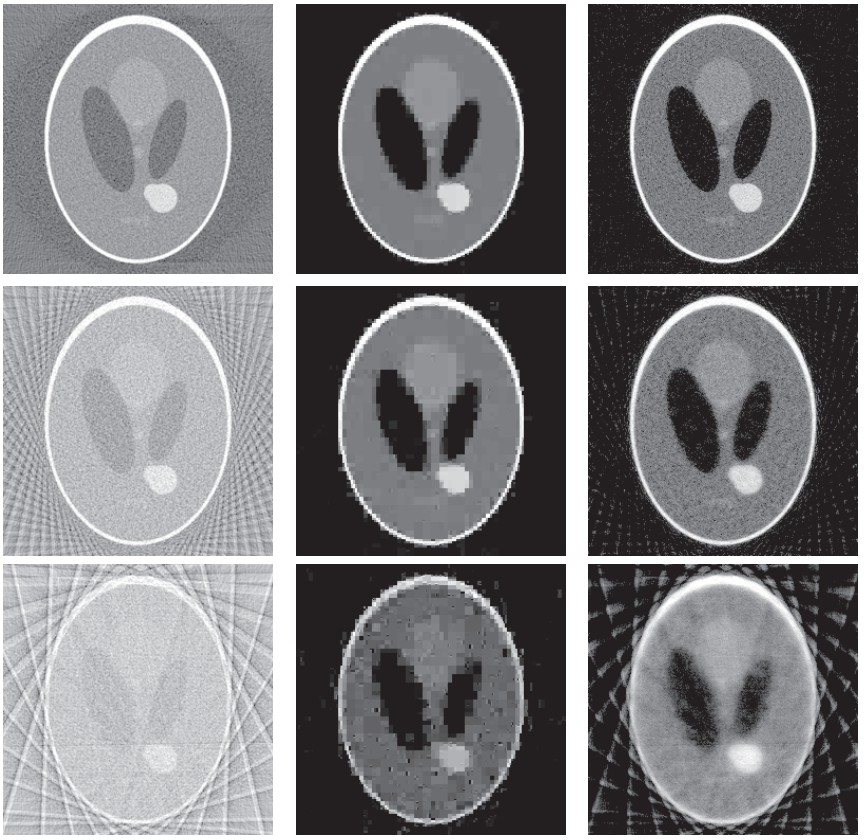

Figure 5.20: Sparse-angle tomography reconstructions of the Shepp-Logan phantom consisting of a tumor. Reconstruction of the FBP *(left column),* MAP *estimates with* B^1_{11} *prior (middle column) and the Tikhonov regularization (right column). The number of projections and the values of the parameters* α *are. Top row: 148 projections,* $\alpha = 32.64$*, and* $\alpha_{\mathrm{M}} = 0.0586$*. Middle row: 37 projections,* $\alpha = 15.24$ *and* $\alpha_{\mathrm{M}} = 0.0092$*. Bottom row: 13 projections,* $\alpha = 5.18$ *and* $\alpha_{\mathrm{M}} = 0.0176$*.* α *denotes the prior parameter selected using the S-curve method and* α_{M} *denotes the regularization parameters selected using the Morozov's discrepancy principle. The reconstruction resolution is* 256×256*.*

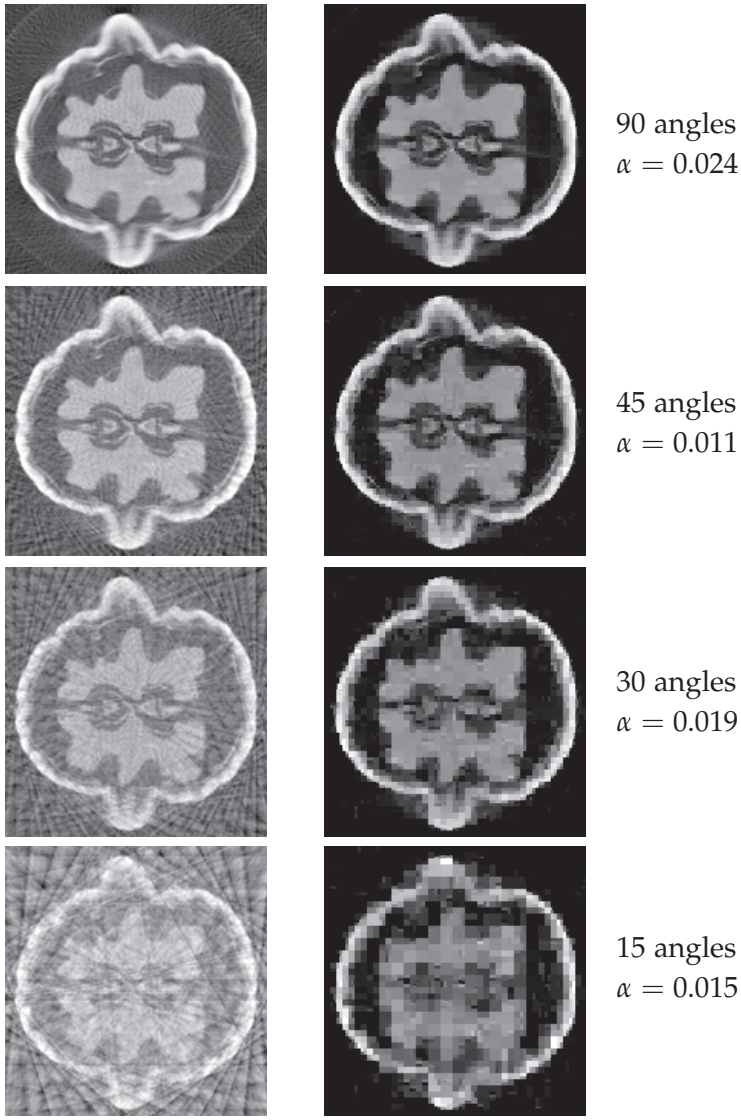

Figure 5.21: Sparse angle tomography reconstructions of the walnut data. Reconstruction using FBP *(left column) and* MAP *estimates with* B_{11}^1 *prior (right column). The numbers of equally distributed projections angles and the values of* α *selected using the S-curve method are denoted on the right. The reconstructions are computed at resolution* 256×256.

6 Summary and Conclusions

In several real world problems, a reconstruction of the unknown quantity needs to be recovered from indirect measurements. The majority of such problems are also large-scale and ill-posed. To overcome the difficulties related to such inverse problems, several approaches including Bayesian methods, wavelet-based methods and constrained optimization methods, were considered while developing the computational methods proposed in this thesis. The performances of the proposed computational optimization methods were evaluated in the four studies.

Publication **I** considered the problem of reconstructing images from local X-ray tomography data. The objective in **I** was to develop a method that enables reduction of the number of unknowns in the inverse problem without degrading the image quality inside the region of interest (ROI). In **I** a Bayesian wavelet-based multiresolution method was proposed where high resolution was used inside the ROI and much coarser resolution elsewhere in the image domain. The intuition behind the wavelet-based multiresolution method was the fact that the fine scale details of the tissue structures outside the ROI cannot be reconstructed well (even in a qualitative sense) from the local tomography data [89, 92], since they are present only in a few of the projection measurements. The results of Publication **I** indicated that by using a coarse wavelet scale in $\Omega \backslash \Omega_{\text{ROI}}$ the contribution of the tissue structures outside Ω_{ROI} could be taken satisfactorily into account. Furthermore, the proposed method yielded similar image quality in Ω_{ROI} as the state-of-the-art pixel-based method while significantly reducing the degrees of freedom of the inverse problem.

Publication **II** considered discretization-invariant Bayesian inversion with Besov B^1_{11} prior, in particular estimates that preserved edges and were sparsity-promoting. The computation of the MAP estimates with B^1_{11} prior is a non-differentiable optimization problem thus it was reformulated into a QP form and PD-IP method was derived in order to recover the MAP estimates. Furthermore, Publication **II** proposed a novel method for selecting the prior parameter α. The proposed choice rule was based on a priori information about the sparsity of the unknown function. The results demonstrated that the use of Haar wavelets with Besov B^1_{11} space prior enabled the recovery of discretization-invariant MAP estimates that preserved sharp edges and had approximately the same level of sparsity

as the a priori estimated sparsity level of the unknown quantity. Furthermore, the proposed parameter selection rule seemed to work robustly also in noisy conditions.

In Publication **III** the object was to recover discontinuous diffusion coefficients from limited observation data based on a stability result for the inverse coefficient problem. The diffusion coefficients were related to a coupled parabolic system of reaction-diffusion type. According to the main theoretical result of Publication **III**, the diffusion coefficients could be recovered by minimizing a quadratic function subject to constraints. The constraints included a linearized version of the original coupled parabolic system as well as bounds-on-variable constraints. Hence the developed PD-IP method was expanded to non-stationary problems in Publication **III**. The performance of the method was studied using two test cases. The results indicated that the PD-IP method allowed to recover accurate reconstructions of the discontinuous diffusion coefficients.

Publication **IV** studied the performance of a wavelet-based sparsity-promoting Bayesian method in the context of sparse angle tomography. Bayesian MAP estimates with Besov B^1_{11} prior and Haar wavelet functions were computed from both simulated and experimental data. The developed PD path following IP method was used to recover the 2D reconstructions. As a reference method reconstruction with FBP and Tikhonov regularization were computed. The results suggested that the MAP estimates with B^1_{11} prior performed robustly also when the number of projections was decreased. Further, Publication **IV** proposed modifications to the sparsity-based α-choice rule presented in **II**. The proposed modifications made the method computationally less complex since now only a few reconstructions are needed to be computed in order to select a value for α. The proposed method was called the S-curve method. The proposed S-curve method was compared to the L-curve method. The results indicated that the proposed S-curve method performs well with sparse angle tomography data while the L-curve method failed to produce useful results. In addition the performance of the S-curve method was evaluated with the following test. Supposing that the a priori sparsity level is estimated from a healthy subject, what happens if this estimate is then used to compute reconstruction from tumor data, will the tumor be detectable? According to the results it seemed that since wavelets are mostly needed near the edges of the image, also the edges of the tumor are visible as well as the other structures.

In many practical applications reconstructions with high resolution need to be computed from indirect, incomplete and noisy measurements.

The high resolution requirement means that the number of degrees of freedom in the inverse problem is large, making the problem computationally demanding. Furthermore, often several restrictions or constraints related to the physics have to be taken into account when estimates of the sought-for quantity are computed. For example, if the solution is required to be non-negative the computational method used has to be such that non-negativity of the solution can be enforced. This, however, can be difficult to do in large-dimensional setting using classical computational methods. A quadratic programming approach provides a natural platform to enforce different constraints. Furthermore, PD-IP methods have been found to perform well also with large-scale problems. The results of this thesis imply that feasible estimates can be computed using tailored PD-IP methods in the contexts of different large-scale constrained problems. Also, the use of Besov space prior with wavelet functions within the Bayesian framework seems promising, enabling for example recovery of sparsity-promoting and edge-preserving Bayesian MAP estimates. Further, within local tomography problems it is possible to reduce the number of unknowns without a significant loss of image quality inside the ROI. Such a method could be considered, for example, when developing the next generation's low-dose X-ray tomography devices. In conclusion the methods proposed in this thesis seem promising and suitable for several areas of application.

References

[1] K. Hanson, "Bayesian and related methods in image reconstruction from incomplete data," *Image Recovery: Theory and Application* 79–125 (1987).

[2] M. S. Lehtinen, *On statistical inversion theory. Theory and applications of Inverse Problems* (Longman, London, 1988).

[3] K. Mosegaard and A. Tarantola, "Monte Carlo sampling of solutions to inverse problems," *Journal of Geophysical Research B* **14,** 12431–12447 (1995).

[4] G. K. Nicholls and C. Fox, "Prior modelling and posterior sampling in impedance imaging," in *Bayesian Inference for Inverse Problems*, Vol. 3459, SPIE proceedings series (Society of Photo-Optical Instrumentation Engineers, Bellingham WA, 1998), pp. 116–127.

[5] L. Tenorio, "Statistical regularization of inverse problems," *SIAM Review* **43,** 347–366 (2001).

[6] S. N. Evans and P. B. Stark, "Inverse Problems as statistics," *Inverse Problems* **18,** R55–R97 (2002).

[7] J. P. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Vol. 160 of *Applied Mathematical Sciences* (Springer Verlag, 2004).

[8] Y. Meyer, *Wavelets and operators*, Vol. 1, (Cambridge University Press, Cambridge, UK, 1995).

[9] C. Chui, *Wavelets: A Mathematical Tool for Signal Analysis* (SIAM, Philadelphia, 1997).

[10] E. Aboufadel and S. Schlicker, *Discovering wavelets* (John Wiley & Sons, 1999).

[11] B. Vidakovic, *Statistical modeling by wavelets* (John Wiley, 1999).

[12] S. Jaffard, Y. Meyer, and R. D. Ryan, *Wavelets: Tools for Science & Technology* (SIAM, 2001).

[13] F. Keinert, *Wavelets and Multiwavelets* (Chapman & Hall, 2004).

[14] S. Jaffard, "Beyond Besov spaces part 1: Distributions of wavelet coefficients," *Journal of Fourier Analysis and Applications* **10,** 221–246 (2004).

[15] I. Daubechies, *Ten Lectures on Wavelets (Ninth printing, 2006)*, Vol. 61 of *CBMS-NSF Regional conference series in applied mathematics* (SIAM, 2006).

[16] H. Triebel, *Function Spaces and Wavelets on Domains*, Vol. 7 of *Tracts in Mathematics* (European Mathematical Society, 2008).

[17] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory* **52,** 489–509 (2006).

[18] E. Candès, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics* **59,** 1207–1223 (2006).

[19] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics* **59,** 1207–1223 (2006).

[20] E. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics* **6,** 227–254 (2006).

[21] E. Candès and T. Tao, "Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," **52,** 2406–2425 (2006).

[22] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory* **52,** 1289–1306 (2006).

[23] D. C. Dobson and F. Santosa, "An image enhancement technique for electrical impedance tomography," *Inverse Problems* **10,** 317–334 (1994).

[24] D. C. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM Journal on Applied Mathematics* **56,** 1181–1198 (1996).

[25] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms.," *Physica D* **60,** 259–268 (1992).

[26] M. Lassas, E. Saksman, and S. Siltanen, "Discretization-invariant Bayesian inversion and Besov space priors," *Inverse Problems and Imaging* **3,** 87–122 (2009).

[27] M. Dashti, S. Harris, and A. Stuard, "Besov priors for Bayesian inverse problems," *Inverse Problems and Imaging* **6,** 183–200 (2012).

[28] J. Nocedal and S. Wright, *Numerical Optimization*, 2 ed. (Springer Verlag, New York, 2006).

[29] S. Wright, *Primal-Dual Interior-Point Methods*, Vol. 54, (SIAM, Philadelphia PA, 1997).

[30] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2009).

[31] R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, Vol. 114 of *International Series in Operations Research & Management Science*, 3 ed. (Springer Verlag, New York, 2008).

[32] N. Karmarkar, "A new polinimial.time algorithm for linear programming," *Combinacatorica* **4,** 373–395 (1984).

[33] S. Mehrotra, "On the implementation of a primal-dual interior point method," *SIAM Journal on Optimization* **2,** 575–601 (1992).

[34] I. J. Lustig, M. R.E., and S. D. F., "Computational experience with primal-dual interior point method for linear programming," *Linear Algebra and its Applications* **152,** 191–222 (1991).

[35] R. J. Vanderbei and D. F. Shanno, "An interior-point algorithm for nonconvex nonlinear programming," *Computational Optimization and Applications* **13,** 231–252 (1999).

[36] C. A. Johnson and A. Sofer, "A primal-dual method for large-scale image reconstruction in emission tomography," *SIAM Journal on Optimization* **11,** 691–715 (2000/01).

[37] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, "Efficient minimization methods of mixed *l*2-*l*1 and *l*1-*l*1 norms for image restoration," *SIAM Journal on Scientific Computing* **27,** 1881–1902 (2006).

[38] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An Interior-Point Method for Large-Scale l1-Regularized Least Squares," *IEEE Journal of Selected Topics in Quantum Electronics* **1,** 606 –617 (2007).

[39] M. Nikolova, M. K. Ng, S. Zhang, and W.-K. Ching, "Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization," *SIAM Journal on Imaging Sciences* **1,** 2–25 (2008).

[40] A. Borsic, B. Graham, A. Adler, and W. Lionheart, "In vivo impedance imaging with total variation regularization," *IEEE Transactions on Medical Imaging* **29,** 44–54 (2010).

[41] A. Borsic and A. Adler, "A primal dual -interior point framework for using the L1-norm or the L2 norm on the data and regularization terms of inverse problems," (2012), submitted to Inverse Problems.

[42] K. D. Andersen, E. Christiansen, A. R. Conn, and M. L. Overton, "An efficient primal-dual interior-point method for minimizing a sum of Euclidean norms," *SIAM Journal on Scientific Computing* **22,** 243–262 (electronic) (2000).

[43] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelets, approximation, and statistical applications*, Vol. 129 of *Lecture Notes in Statistics* (Springer-Verlag, New York, 1998).

[44] R. A. DeVore, G. C. Kyriazis, and P. Wang, "Multiscale Characterizations of Besov Spaces on Bounded Domains," *Journal of Approximation Theory* **93,** 273–292 (1998).

[45] D. Leporini and J. Pesquet, "Bayesian wavelet denoising: Besov priors and non-Gaussian noises," *Signal processing* **81,** 55–67 (2001).

[46] M. Rantala, S. Vänskä, S. Järvenpää, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen, "Wavelet-based reconstruction for limited-angle X-Ray tomography," *IEEE Transactions on Medical Imaging* **25,** 210–217 (2006).

[47] S. Vänskä, M. Lassas, and S. Siltanen, "Statistical X-Ray tomography using empirical Besov priors," *International Journal of Tomography & Statistics* **11,** 3–32 (2009).

[48] A. Chambolle, R. A. DeVore, N. yong Lee, and B. J. Lucier, "Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal through Wavelet Shrinkage," *IEEE Transactions on Image Processing* **7,** 319–335 (1998).

[49] 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,** 1413–1457 (2004).

[50] H. Choi and R. Baraniuk, "Multiple wavelet basis image denoising using Besov ball projections," *IEEE, Signal Processing Letters* **11,** 717–720 (2004).

[51] R. Ramlau and G. Teschke, "A Tikhonov-based projection iteration for nonlinear Ill-posed problems with sparsity constraints," *Numerische Mathematik* **104,** 177–203 (2006), 10.1007/s00211-006-0016-3.

[52] R. Ramlau, "Regularization properties of Tikhonov regularization with sparsity constraints," *Electronic Transactions on Numerical Analysis* **30,** 54–74 (2008).

[53] S. Loustau, "Penalized empirical risk minimization oer Besov spaces," *Electronic Journal of Statistics* **3,** 824–850 (2009).

[54] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. P. 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,** 1465–1490 (2003).

[55] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo, "Statistical inversion for medical X-ray tomography with few radiographs: I. General theory," *Physics in medicine and biology* **48,** 1437–1463 (2003).

[56] V. Kolehmainen, A. Vanne, S. Siltanen, S. Järvenpää, J. Kaipio, M. Lassas, and M. Kalke, "Bayesian inversion method for 3D dental X-Ray imaging," *Elektrotechnik und Informationstechnik* **124,** 248–253 (2007).

[57] 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," *EEE Transactions on Medical Imaging* **25,** 218–228 (2006).

[58] J. P. Kaipio, V. Kolehmainen, E. Somersal, and M. Vauhkonen, "Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography," *Inverse Problems* **16,** 1487–1522 (2000).

[59] J. Bardsley, D. Calvetti, and E. Somersalo, "Hierarchical regularization for edge-preserving reconstruction of PET images," *Inverse Problems* **26** (2010).

[60] J. M. Bardsley, "An efficient computational method for total variation-penalized poisson likelihood estimation," *Inverse Problems and Imaging* **2,** 167–185 (2008).

[61] B. Amizic, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Sparse bayesian blind image deconvolution with parameter estimation," in *European Signal Processing Conference* (2010), pp. 626–630.

[62] S. D. Babacan, J. Wang, R. Molina, and A. K. Katsaggelos, "Bayesian Blind Deconvolution From Differently Exposed Image Pairs," *IEEE Transactions on Image Processing* **19,** 2874–2888 (2010).

[63] T. R. Greeshma and P. M. Ameeramol, "Bayesian MAP Model for Edge Preserving Image Restoration: A Survey," in *International Journal of Computer Applications* (2011).

[64] M. Lassas and S. Siltanen, "Can one use total variation prior for edge-preserving Bayesian inversion?," *Inverse Problems* **20,** 1537–1564 (2004).

[65] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (John Wiley and Sons, New York, NY, 1968).

[66] "MATLAB, the language of technical computing," http://www.mathworks.fr/index.html (2012).

[67] "Scilab Software for Numerical Computation," http://www.scilab.org/ (2012).

[68] "The MOSEK Optimization Tools," www.mosek.fi (2012).

[69] "CVX: Matlab Software for Disciplined Convex Programming," http://cvxr.com/cvx (2012).

[70] "CPLEX," http://www.aimms.com/cplex-solver-for-linear-programming2?gclid=CID4obHZ0rICFUEMfAodSnoA9w (2012).

[71] "LOQO," http://www.princeton.edu/ rvdb/loqo/LOQO.html (2012).

[72] R. J. Vanderbei, "LOQO User's Manual - Version 4.05," (2006).

[73] "KNITRO," http://www.ziena.com/knitro.htm (**!!YEAR!!**).

[74] C. Johnson, J. Seidel, and A. Sofer, "Interior-point methodology for 3-D PET reconstruction," *IEEE Trans. Med. Imaging* **19,** 271 –285 (2000).

[75] Y. Mamatjan, A. Borsic, D. Grusoy, and A. Adler, "Electrical Impedance Tomography Image Reconstruction with ℓ_1 data and image norms," (2012), submitted to IEEE Transactions on Biomedical Engineering.

[76] F. Alvarez, J. Bolte, J. Bonnans, and F. Silva, "Asymptotic expansions for interior penalty solutions of control constrained linear-quadratic problems," *Mathematical Programming* **135,** 473–507 (2012).

[77] M. Hinze and A. Schiela, "Discretization of interior point methods for state constrained elliptic optimal control problems: optimal error estimates and parameter adjustment," *Computational Optimization and Applications* **48,** 581–600 (2010).

[78] A. Schiela, "Barrier methods for optimal control problems with state constraints," *SIAM Journal on Optimization* **20,** 1002–1031 (2009).

[79] A. Schiela and M. Weiser, "Superlinear convergence of the control reduced interior point method for PDE constrained optimization," *Computational Optimization and Applications* **39,** 369–393 (2008).

[80] A. Schiela and A. Günter, "An interior point algorithm with inexact step computation in function space for state constrained optimal control," *Numerische Mathematik* **199,** 373–407 (2011).

[81] M. Ulbrich and U. S., "Primal-dual interior point methods for PDE-constrained optimization," *Mathematical Programming* **117,** 435–485 (2009).

[82] M. Weiser, T. Gänzler, and A. Schiela, "A control reduced primal interior point method for PDE constrained optimizatio," *Computational Optimization and Applications* **41,** 127–145 (2008).

[83] W. Wollner, "A posteriori error estimates for a finite element discretization of interior point methods for an elliptic optimization problem with state constraints," *Computational Optimization and Applications* **47,** 133–159 (2010).

[84] I. Neitzel, U. Prüfert, and T. Slawig, "Strategies for time-dependent PDE control using an integrated modeling and simulation environment. Part one: problems without inequality constraints," *Numerical Algorithms* **50,** 241–269 (2009).

[85] U. Prüfert and F. Tröltzsch, "An interior point method for a parabolic optimal control problem with regularized pointwise state

constraints," *ZAMM - Journal of Applied Mathematics and Mechanics* **87,** 564–589 (2007).

[86] Q. Zhang, H. Zhao, D. Chen, X. Qu, X. Chen, X. He, W. Li, Z. Hu, J. Liu, J. Liang, and J. Tian, "Source sparsity based primal-dual interior-point method for three-dimensional bioluminescence tomography," *Optics Communications* **284,** 5871–5876 (2011).

[87] K. Fountoulakis, J. Gondzio, and P. Zhlobich, "Matrix-free Interior Point Method for Compressed Sensing Problems," (2012).

[88] K. T. Smith and K. F., "Mathematical foundations of computed tomography," *Applied Optics* **24,** 3950–3957 (1985).

[89] F. Natterer, *The mathematics of computerized tomography*, Vol. 32 of *SIAM Classics in Applied Mathematics* (SIAM, 2001).

[90] A. Faridani, E. L. Ritman, and K. T. Smith, "Local tomography," *SIAM Journal on Applied Mathematics* **52,** 459–484 (1992), Examples became disorganized while the article was in press. A reorganization is in: Examples of local tomography, SIAM Journal on Applied Mathematics, vol. 52, num. 4, p. 1193–1198.

[91] A. Faridani, D. Finch, E. Ritman, and K. Smith, "Local tomography II," *SIAM Journal on Applied Mathematics* **57,** 1095–1127 (1997).

[92] P. Maaß, "The interior Radon transform," *SIAM Journal on Applied Mathematics* **52,** 710–724 (1992).

[93] K. M. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *Journal of the Optical Society of America* **73,** 1501–1509 (1983).

[94] K. Sauer, S. J. James, and K. Klifa, "Bayesian estimation of 3-D objects from few radiographs," *IEEE Transactions on Nuclear Science* **41,** 1780–1790 (1994).

[95] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing* **2,** 296–310 (1993).

[96] J. Zheng, S. Saquib, K. Sauer, and C. Bouman, "Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence," *IEEE Transactions on Image Processing* **9,** 1745–1759 (2000).

[97] E. Polak and G. Ribière, "Note sur la convergence de méthodes de directions conjuguées.," *Revue Française d'Informatique et de Recherche Opérationnelle* **16,** 35–43 (1969).

[98] K. Niinimäki, V. Kolehmainen, and S. Siltanen, "Bayesian multiresolution method for local tomography," in *Frontiers in Optics*, Laser Science XXV/Adaptive Optics: Methods, Analysis and Applications/Advances in Optical Materials/Computational Optical Sensing and Imaging/Femtosecond Laser Microfabrication/Signal Recovery and Synthesis (2009).

[99] K. Niinimäki, V. Kolehmainen, and S. Siltanen, "Bayesian multiresolution method for local X-ray tomography in delta radiology," in *Computational Imaging VII*, Vol. 7245 (2009).

[100] D. A. Lorenz, "Convergence rates and source conditions for Tikhonov regularization with sparsity constraints," *Journal of Inverse and Ill-posed Problems* **16,** 463–478 (2008).

[101] O. Poisson, "Uniqueness and Hölder stability of discontinuous diffusion coefficients in three related inverse problems for the heat equation," *Inverse Problems* **24** (2008).

Kati Niinimäki

Computational Optimization Methods for Large-scale Inverse Problems

Large-scale inverse problems, where accurate reconstructions needs to be computed from indirect and noisy measurements, are encountered in several application areas. Due to the ill-posed nature of inverse problems a priori information needs to be incorporated into the inversion model. This thesis proposes computational methods that are adaptable to solve such large-scale inverse problems. In particular considered are constrained optimization methods, such as primal-dual interior-point methods.

UNIVERSITY OF
EASTERN FINLAND

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences

ISBN 978-952-61-1098-1