NOVEL ALGORITHMS IN X-RAY COMPUTED TOMOGRAPHY IMAGING FROM UNDER-SAMPLED DATA

Zenith Purisha

Academic dissertation

To be presented for public examination,
with the permission of
the Faculty of Science of the University of Helsinki,
in Auditorium B123, Exactum, Kumpula Campus,
on 14th of May 2018 at 12 o'clock

Department of Mathematics and Statistics
Faculty of Science
University of Helsinki

HELSINKI 2018

ISBN 978-951-51-4221-4 (paperback)
ISBN 978-951-51-4222-1 (PDF)

https://ethesis.helsinki.fi/

Unigrafia Oy
HELSINKI 2018

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give a particular mention here.

First and foremost I want to thank my advisor Samuli Siltanen who gave me the golden opportunity to join his research group and to work on interesting research. Back in 2010 when I joined the summer school in University of Helsinki, his attractive presentation about X-ray tomography drew my attention and lead me to work on this topic for my thesis. I would like to express my deep gratitude to him for his patient guidance, enthusiastic encouragement and immense knowledge of this research work. Also, not to mention that he has taught me more than I could ever give him credit for here. He has shown me, by his example, what a good scientist and person should be.

I would like to express my very great appreciation to Elena Loli Piccolomini and Esther Klann, as well as my opponent Marta Betcke for their valuable time spent reading my thesis.

Furthermore, I am grateful to all of those with whom I have had the collaborations to work on this thesis: Tatiana Bubba, Heikki Haario, Aki Kallonen, Sakari S. Karhula, Juuso H. Ketola, Heikki Kröger, Marko Laine, Esa Niemi, Miika T. Nieminen, Juho Rimpeläinen, Simo Saarakkala. This fruitful collaboration has been immensely broadening my scientific knowledge and networking. Thanks to Tapio Helin, Martin Burger and Felix Lucka for the opportunity working on Bayesian overconfidence project, additionally to my studies.

My thesis will not work successfully without financial supports, therefore I would like to thank the Academy of Finland (Finnish Centre of Excellence in Inverse Problems Research 2012-2017, and grant nos. 250215, 268378, 284715 and 303786), Väisälä Foundation, TEKES / Business Finland "Heartbeat" project 6614/31/2016, FiDiPro project and NordForsk project AUQ-PDE (#74756).

I wish to acknowledge the help by Esa Niemi who patiently guide me in technical stuff especially in tomography computation since I joined the Inverse problems group. Thank you also for reading my thesis and giving constructive feedback including the grammar. My sincere thanks also goes to Inverse Problems group: Minh Mach, Rodrigo Bleyer, Lauri Ylinen, Hanne Kekkonen, Paola Elefante, Teemu Saksala, Markus Juvonen, Andreas Hauptmann, Alexander Meaney, Jonatan Lehtonen,

Matteo Santacesaria and all the rest for a big contribution in making a perfect research atmosphere. I also very much enjoyed the time we spent in taking X-ray measurements. My grateful thanks are also extended to Giovanni Canarecci and Marcelo Hartmann for being kind and nice officemates.

I am particularly grateful for the insightful comments by Janne Hakkarainen which improved a lot my thesis and the English in many parts. I thank him as well for discussions of Dynamic X-ray project additionally to my studies. Also thanks for many nice conversations.

Special thanks should be given to Maaria Rantala for interesting conversations about many things. I very much enjoyed the time we shared since I started my PhD study.

Supports and encouragements given by colleagues and friends from Universitas Gadjah Mada (UGM), especially Noorma, Nanang, and Niar were greatly appreciated. My special thanks are extended to friends from Indonesian students association (Indonesialaisten Opiskelijoiden Yhdistys, PPI Finlandia ry) who always bring joy, supports, discussions and encouragements. I always have the fighting spirit to finish my study here in Finland, far from our home country,...since I know, I am not alone. Special thanks also for Windi who has been supporting and helping me in many things before and during my stay in Helsinki.

I am deeply grateful for the presence of my friends outside of the PPI: Yayuk, Iis, Ayu, Fitri (and cute Khansa), Onel, Ai, Giorgia, Maretha. It always feels good to meet and talk to you about ridiculous to meaningful conversations (it always makes my brain recharged). Not to mention that, I very much enjoyed our time in cookings, eatings, travelings and also exploring Helsinki's cafes.

Nobody has been more important to me in the pursuit of this project than the members of the family. I would like to thank my big family and my mother in law for their supports. This thesis work is dedicated to my parents, my heroes, Suhadi Purwantara and Aisyah, I would like to offer my special thanks to them for their love, supports, guidance and wise counsel in whatever I pursue.

Terimakasih Pak, Bu, atas doanya yang tiada henti. Tanpa dukungan dari bapak ibu, Enit gak akan bisa sampai di titik sekarang ini.

Most importantly, this work is also dedicated to my husband, Isom Hilmi, who has been a constant source of support in a personal level and in an academic level. Thank you for your sympathetic ear and great patience at all times. You have been my pillar of strength through all my ups and downs. I am truly grateful for having you in my life.
Alhamdulillah...

Helsinki, April 2018

Zenith Purisha

The articles and the author's contribution are summarized below.

List of publications

I Purisha, Z. and Siltanen, S., 2016. Tomographic Inversion using NURBS and MCMC. In *Forging Connections between Computational Mathematics and Computational Geometry* (pp. 153-166). Springer, Cham.

II Purisha, Z. and Siltanen, S., 2014. Tomographic Reconstruction of Homogeneous 2D Geometric Models with Unknown Attenuation. In *IFIP Conference on System Modeling and Optimization* (pp. 247-256). Springer, Berlin, Heidelberg.

III Haario, H., Kallonen, A., Laine, M., Niemi, E., Purisha, Z. and Siltanen, S., 2017. Shape recovery for sparse-data tomography. In *Mathematical Methods in the Applied Sciences*.

IV Purisha, Z., Rimpeläinen, J., Bubba, T. and Siltanen, S., 2017. Controlled wavelet domain sparsity for x-ray tomography. In *Measurement Science and Technology*.

V Purisha, Z., Karhula, S.S., Ketola, J.H., Rimpeläinen, J., Nieminen, M.T., Saarakkala, S., Heikki Kröger and Siltanen, S., 2018. An automatic regularization method: An application for 3D X-ray micro-CT reconstruction using sparse data. *preprint arXiv*: 1708.02067.

Author's contributions

I The author has crucial part in developing the algorithm and the numerical studies. She had a major part of writing.

II The development of the algorithm is mainly due to the author. She conducted the computation. She had a major part of writing.

III The development of the algorithm is mainly due to the author. She conducted the computation. The analysis of the results and the writing were done in equal parts with S.Siltanen.

IV The author and J.Rimpeläinen had an equivalent role in the development of the algorithm. The author conducted the computation. The analysis of the results and the writing were done in equal parts with the other authors.

V The development of the algorithm is mainly due to the author. She conducted the computation. The analysis of the results were done in equal parts with S.S. Karhula and J.H. Ketola. The author had a major part of writing.

Contents

1 Introduction

In many applications in real-world problems, we need to interpret indirect measurements of a target of interest. As an example, in airport security, baggage screening is required to show what is inside each bag. The measurements are collected by penetrating X-rays through the baggage. In mathematics, a problem how to model the measurement process is called the direct or forward problem. The opposite of forward problem is the so-called inverse problem, in which we collect all available observations from indirect measurements, quite often incomplete, and use the data to obtain the target of interest. In this example, the collected X-ray measurement data are used to reconstruct the inside of the baggage. There are many other examples of inverse problems, for instance, image denoising to remove noise from photographic images, remote sensing to monitor land use using reflected energy (often from sunlight), imaging the subsurface of the Earth using seismic waves produced by earthquakes.

Forward problems are usually *well-posed*, meaning that they are numerically stable and can be solved reliably. However, typically inverse problems are *ill-posed*, which means the problem either has *no solution* or has *many solutions*, or the solution procedure is *unstable* (having an arbitrarily small error in the measurement data may cause large errors in the solutions). Most difficulties in solving *ill-posed* problems come from the instability of the solution. In practice, the measured data given by the measurement device are always corrupted by noise. Mathematically, the inverse problem is difficult to solve due to the sensitivity to the measurement noise.

In this overview, we focus on *X-ray tomography*, a subfield of inverse problems in which an object is penetrated by X-rays along multiple views and the measurements are the integration along the ray paths. The aim of X-ray tomography is to reveal the inner structures of the object using a non-invasive procedure from the X-ray measured data. X-ray tomography, also known as *computerized tomography* (CT), was introduced by Allan Cormack and Godfrey Hounsfield (Nobel laureate) in the 1970s [10, 11]. CT has been widely implemented for visualizing the internal structure of products in industrial applications [1, 15, 5] and of patients in medical imaging [32, 12, 49]. The inverse problem of the X-ray attenuation reconstruction (the target structure) is well understood using standard methods in which dense measured X-ray data from a full angle of view is required to obtain a satisfying reconstruction image. As an example, Figure 1.1 shows the human trabecular bone reconstructions

from full angle using Feldmann, David and Kress (FDK) method [3]. Unfortunately, in many applications, sufficient information is not available. In reverse engineering, for example, X-ray tomography is used to duplicate and manufacture an existing object. However, to convert the data into a discrete model representation can be very costly computationally and time-consuming [28]. Therefore, by using only limited or sparse data of tomographic images would be one way to cope the problem.

Another example is in biomedical imaging in which the full data are often difficult to collect due to:

- the needs to reduce the radiation dose to the patient or *in vivo* samples since the high dose can modify the properties of living tissues,
- the long acquisition time to obtain high-resolution data of *in vitro* samples, or
- restrictions in the geometry of the measurement setting, for example in mammographic imaging [42, 44, 55, 56] and electron imaging [19].

In the situation in which only limited projection data are available, it is well known that the standard CT method does not yield satisfactory reconstructions [30]. *A priori* information such as enforcing regularity in the model is required in the reconstruction algorithm in order to compensate for the incomplete datasets. As an example, a Bayesian framework is a statistical approach to solve the inverse problems with statistical inference scheme. In this approach, the prior knowledge is formulated as a prior distribution, and the measurement model is constructed as a likelihood. Bayes' formula combines the prior distribution and the likelihood together as a posterior distribution [29]. In the Bayesian framework, the posterior density is the solution of the inverse problem. Popular statistical point estimates are:

- *maximum a posteriori* estimate (MAP): the highest mode of the posterior and
- *conditional mean* estimate (CM): the mean or the expected value of the posterior.

Finding the MAP estimate requires a solution of an optimization problem. In order to obtain CM estimate, a more powerful method is needed, for instance, Markov chain Monte Carlo (MCMC).

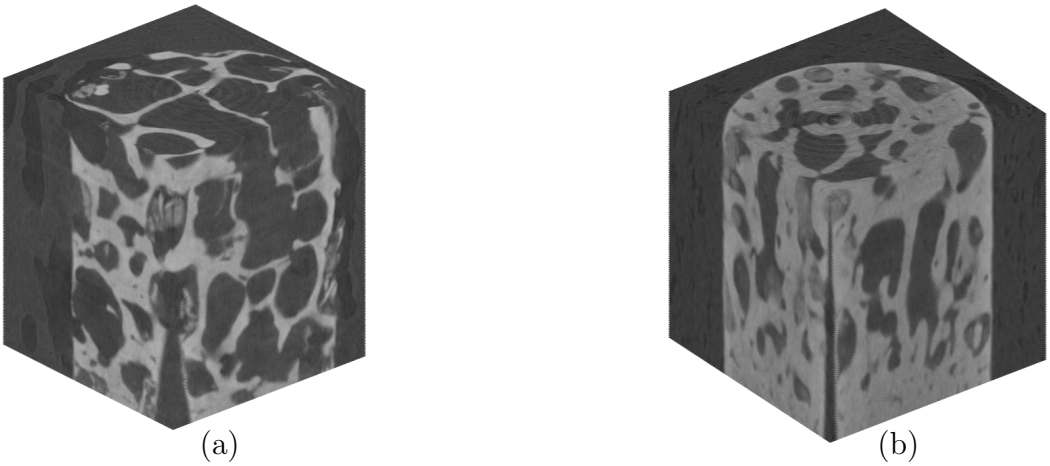
In this thesis, two novel computational inversion methods for solving sparse X-ray tomography problems are proposed:

- MCMC algorithm in which the tomography reconstruction is in terms of non-uniform rational basis spline (NURBS) which is commonly used in computer-aided design (CAD) format. The proposed method assesses the boundary

shape of a homogeneous object (consists of one material) by estimating the coordinates of the control points in the NURBS curve and the attenuation value of the object. MCMC method offers relatively simple algorithm but powerful method to generate samples of the desired distribution.

- Adaptive method for tuning regularization parameter based on a control algorithm driving the sparsity of the reconstruction to an *a priori* known ratio of nonzero versus zero wavelet coefficients in the unknown. A central slice of walnut were tested. In the higher-dimensional case, instead of exploiting wavelets, we use shearlets which have a distinctive feature known as shearings to control directional selectivity. In this work, an automatic regularization method based on enforcing sparsity in three-dimensional shearlet transform domain is implemented. Human subchondral bone samples were tested and morphometric parameters of the bone reconstructions were then analyzed. In real applications, the algorithm seems to be a promising strategy and the end-users could avoid the tuning parameters manually.

The rest of this overview is organized as follows. In Chapter 2, the mathematical model of X-ray tomography and sparse tomography model are described. Introduction to NURBS and the tomographic model using NURBS are given in Chapter 3. Chapter 4 reviews the basics of wavelets and shearlets. Bayesian approach and the different types of MCMC algorithms are discussed in Chapter 5. The results of the publications are then reviewed in Chapter 6 and finally, Chapter 7 presents the conclusion and future outlook.



(a) (b)

Figure 1.1: 3D reconstructions of healthy (a) and osteoarthritis (b) human trabecular bone using FDK method from full projections.

2 X-ray tomography model

Consider a physical domain $\Omega \subset \mathbb{R}^2$ and a density function $f : \Omega \to \mathbb{R}$. The X-rays travel through Ω along straight lines $L \subset \Omega$ and we assume that the initial intensity of the X-ray is $I_0(L)$ and the exiting X-ray intensity is $I_1(L)$. Then the formula for the intensity loss is

$$\int_L f(x)dx = -\log\frac{I_1(L)}{I_0(L)}. \tag{2.1}$$

In X-ray tomographic imaging, the aim is to collect information about f using available measurement data collected from the intensities $I_1(L)$ of X-rays for all lines L through Ω and from different angles of view. The problem can be written as Radon transform:

$$\mathcal{R}f(s,\theta) = \int_L f(\mathbf{x})\mathbf{dx}, \tag{2.2}$$

with $\mathbf{x} = (x, y)$ and $L = \{(x, y) \in \mathbb{R}^2 : x\cos\theta + y\sin\theta = s, \theta \in [0, \pi), s \in \mathbb{R}\}$. In the idealized continuum case, the inverse Radon transform, the so-called filtered back-projection (FBP) algorithm can reconstruct the attenuation coefficient f perfectly [3, 30, 41].

2.1 Discrete tomographic data

In practice, a discrete model is required for numerical computation. Commonly, the measurement geometry can be modelled as a parallel beam, fan beam or cone beam geometry as it can be seen in Figures 2.1 - 2.3. In the case that we have P equiangularly sampled projections from full 360 degrees, then the angles of view are of the form

$$\theta = \{0, \frac{2\pi}{P}, \frac{2(2\pi)}{P}, \frac{3(2\pi)}{P}, \ldots, \frac{(P-1)2\pi}{P}\}. \tag{2.3}$$

In two-dimensional case, let us represent f by a matrix $\mathbf{f} = [\tilde{f}_{ij}] \in \mathbb{R}^{N\times N}$. After calibration, the measurements of line integrals can be modelled as

$$\int_L f(x)ds = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\tilde{f}_{ij}, \tag{2.4}$$

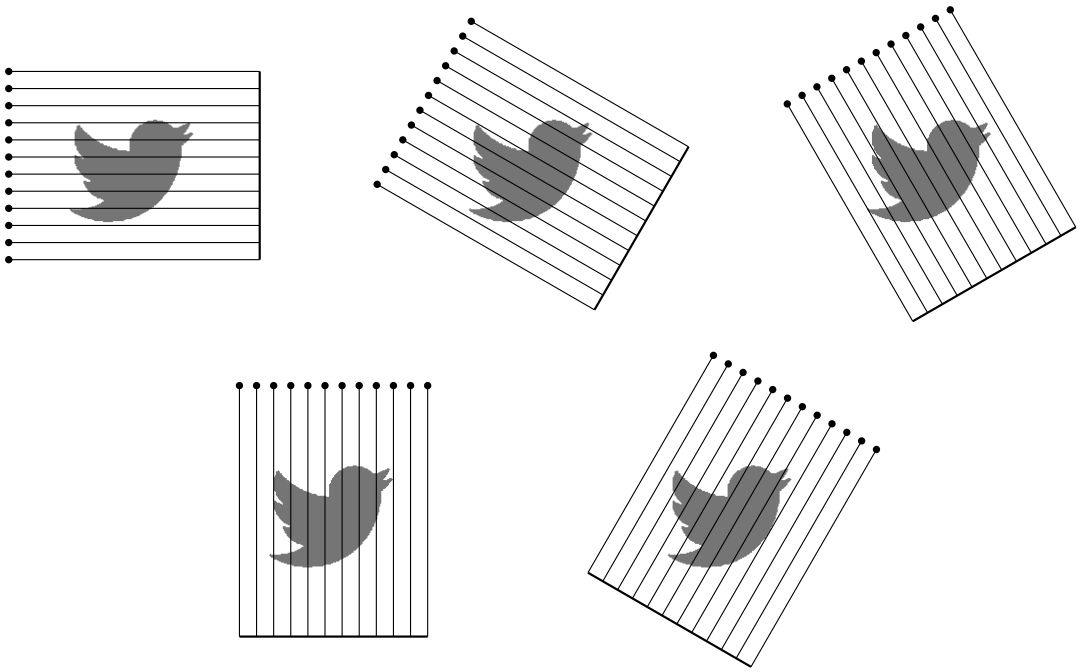


Figure 2.1: Parallel beam X-ray measurement geometry. There are five different directions: $\pi, \frac{5\pi}{6}, \frac{4\pi}{6}, \frac{3\pi}{6}$ and $\frac{2\pi}{6}$ applying 12 ray-lines. Black dots show the locations of the X-ray source at different times of measurement. The thick line represents the detector measuring the intensity of the X-rays after passing through the target.

where a_{ij} is a distance that the lines L travel in the voxel with indices (i, j).

For three-dimensional case, the construction of the measurement matrix is analogous to 2D case. Consider a matrix $\mathbf{f} = [\tilde{f}_{ijk}] \in \mathbb{R}^{N\times N\times N}$. The line integral can be approximated by

$$\int_L f(x)dx = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ijk}\tilde{f}_{ijk}, \tag{2.5}$$

where a_{ijk} is a distance that the line L travels in the voxel with indices (i, j, k).

Then the practical tomographic X-ray measurement data is given as follows:

$$\mathbf{m} = \mathbf{A}\mathbf{f} + \varepsilon, \tag{2.6}$$

where a measurement matrix for 2D case $\mathbf{A} = [a_{ij}]$ (in 3D, $\mathbf{A} = [a_{ijk}]$) containing one row for each voxel in the set of measurements and ε is the measurement noise.

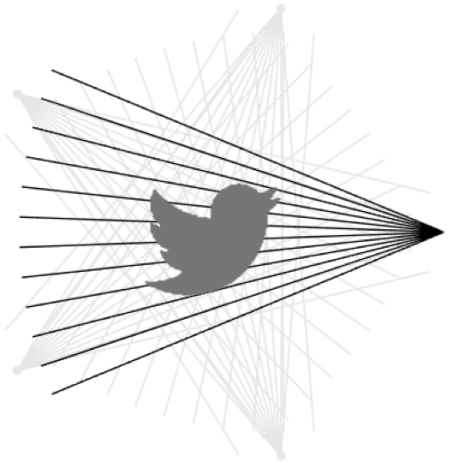


Figure 2.2: Fanbeam X-ray measurement geometry. There are five different directions: 2π, $\frac{8\pi}{5}$, $\frac{6\pi}{5}$, $\frac{4\pi}{5}$ and $\frac{2\pi}{5}$ applying 12 ray-lines. Black dot show the locations of the X-ray source at different times of measurement. In the opposite of the source there is a detector measuring the intensity of the X-rays after passing through the target.

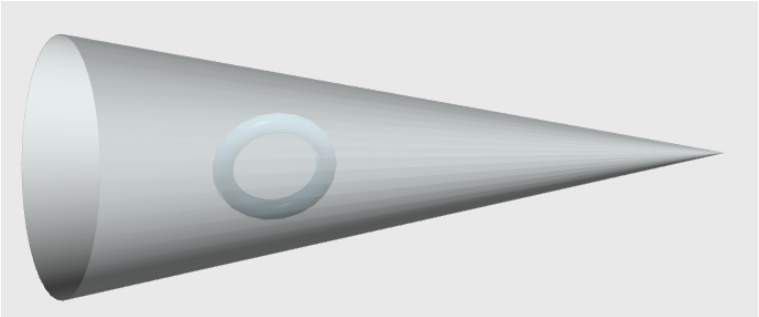


Figure 2.3: An illustration for a cone beam X-ray measurement geometry from one direction: 2π. The point of the cone indicates the X-ray source. The rays penetrates through all the object. In the opposite side of the source, the intensity of the X-rays after passing through the target is recorded by the detector.

2.2 Sparse X-ray tomography

In some applications of X-ray tomography, the datasets are limited. One example is in biomedical imaging such as dental imaging and mammographic imaging, in which the X-ray data acquisition cannot cover all directions [40]. Another example is dynamic tomography, where the target changes in time between the recording of two consecutive projection images. Due to the restriction of the measurement setting, an adjustment in the measurement geometry model is necessary. For instance:

- if the data collection is sparse over full 360 degrees, then the number of projection data P in (2.3) is small, or
- if the data is sampled from limited angle geometry, for instance, the angle of view is 90 degrees only, then

$$\theta = \{0, \frac{\pi}{2P}, \frac{\pi}{P}, \frac{3\pi}{2P}, \ldots, \frac{(P-1)\pi}{2P}\}. \tag{2.7}$$

When only sparse sample data is available, FBP algorithm is not well-suited anymore to perform the reconstruction. As an example, Figure 2.4 shows a real data reconstruction of the central slice using FBP and modern algorithm, so-called Total Variation (TV) method, in which prior information is used, [2, 6, 47]. As it can be seen, the FBP result gives unsatisfying reconstruction, while the TV result yields a much better reconstruction. In general, other computational approaches that enforce *a priori* knowledge or regularization are highly demanded.

2.3 Ill-posed problems

Let us recall the mathematical model (2.6). The inverse problem is to find the vector $\mathbf{f}$ where the measurement $\mathbf{m}$ is given. As mentioned in the introduction, inverse problem is typically *ill-posed* if at least one of the *well-posed* properties of Hadamard condition below is failed:

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

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

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

In this section, let us consider an example in a finite dimensional-setting so called *singular value decomposition (SVD)* of matrix $\mathbf{A}$. We know that any matrix $\mathbf{A} \in \mathbb{R}^{K\times N}$ can be formulated in the following form

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T, \tag{2.8}$$

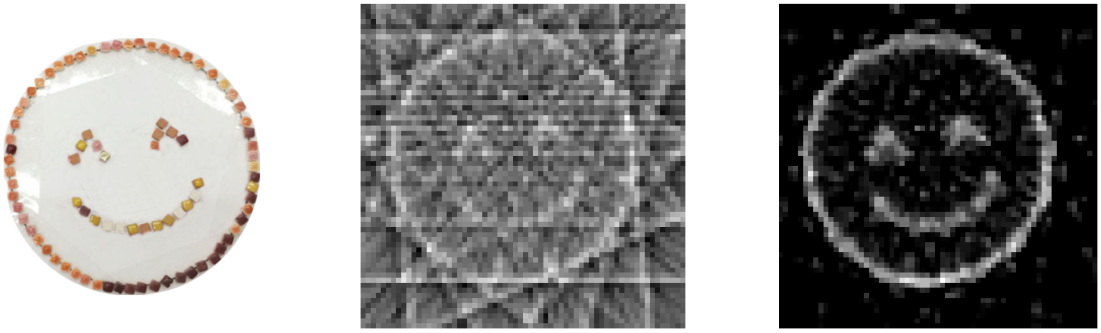


Figure 2.4: An example of tomographic reconstruction of real data. Left: The emoji data built from ceramic stones. Middle: FBP reconstruction from 15 projection data out of 360°. Right: TV reconstruction using the same number of projection data and measurement geometry as FBP setting.

where $\mathbf{U}^{K\times K}$ and $\mathbf{V}^{N\times N}$ are orthogonal matrices and $\Sigma^{K\times N}$ is a diagonal matrix. The formula (2.8) is called SVD of matrix $\mathbf{A}$ and the diagonal elements of matrix Σ are the *singular values* of $\mathbf{A}$. The diagonal elements σ_i have a decreasing property:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(K,N)} \geq 0.$$

Recall the definitions below:

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

Now, let us consider three cases:

(i) For $K > N$ (*overdetermined* problem), then

$$\Sigma^{K\times N} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & \sigma_N \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

It follows that the dimension of Range($\mathbf{A}$) is less than K so that there is a nonzero $\mathbf{m}' \in \text{Coker}(\mathbf{A})$. In the case $\varepsilon = 0$, we cannot find any $\mathbf{f} \in \mathbb{R}^n$ that satisfies $\mathbf{Af} = \mathbf{m}'$. It is even worse for the case $\varepsilon > 0$, even though $\mathbf{Af} \in \text{Range}(\mathbf{A})$, it is possible that $\mathbf{Af} + \varepsilon \notin \text{Range}(\mathbf{A})$. Hence, the existence condition H_1 fails.

(ii) For $K < N$ (*underdetermined* problem), then

$$\Sigma^{N\times K} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots & \vdots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & \sigma_K & 0 & \cdots & 0 \end{bmatrix}.$$

Therefore $\dim(\text{Ker}(A))> 0$ and there exists $\mathbf{f}' \in \text{Ker}(\mathbf{A})$. Even though $\varepsilon = 0$, we have $\mathbf{Af} = \mathbf{m}$ and $\mathbf{A}(\mathbf{f} + \mathbf{f}') = \mathbf{m}$. Hence, the uniqueness condition H_2 fails.

(iii) For $K = N$, the matrix Σ is a square-shaped

$$\Sigma^{K\times K} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & \sigma_K \end{bmatrix}.$$

Consider $\sigma_K > 0$. For $\varepsilon = 0$, let us write the solution using SVD:

$$\mathbf{f} = \mathbf{A}^{-1}\mathbf{m} = (\mathbf{U}\Sigma\mathbf{V}^T)^{-1}\mathbf{m} = \sum_{i=1}^{K} \frac{u_i^T m_i}{\sigma_i} v_i \tag{2.9}$$

Hence if σ_K is small, then small perturbations in $\mathbf{A}$ or $\mathbf{m}$ will cause a significant change in $\mathbf{f}$. Consider a *condition number* below:

$$\text{Cond}(\mathbf{A}) = \frac{\sigma_1}{\sigma_K}. \tag{2.10}$$

If the $\min\{\sigma_i\}$ is very small, then $\text{Cond}(\mathbf{A})$ becomes larger and the matrix is ill-conditioned. In inverse problems, the small singular values are understood as an amplification for the noise components.

In the case $\varepsilon > 0$, we can write

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

The error $\mathbf{A}^{-1}\varepsilon$ can be bounded by $\|\mathbf{A}^{-1}\varepsilon\| \leq \|\mathbf{A}^{-1}\|\|\varepsilon\|$. Consequently, even though ε is small, the the error $\mathbf{A}^{-1}\varepsilon$ can be very large if $\|\mathbf{A}^{-1}\|$ is large. Thus the Hadamard condition H_3 fails.

Moreover, let us still consider the case $K = N$ and recall the inversion matrix $\mathbf{A}^{-1}$ in (2.9). Even though σ_1 is several orders of magnitude greater than σ_K and it leads to the large condition number in (2.10), the matrix $\mathbf{A}$ is invertible. However, the numerical inversion of $\mathbf{A}$ is difficult since the Σ^{-1} contains very different sizes of floating point numbers.

Thus, to understand the ill-posedness in this case, we must consider the infinite-dimensional problem. Suppose that we model the continuum measurement by a sequence of matrix $\mathbf{A}_K$ with a size $K \times K$ where $K = K_0, K_0 + 1, K_0 + 2, \cdots$. The approximation to the forward model will be better as K grows. Then if

$$\lim_{K\to\infty} \text{Cond}(\mathbf{A}_K) = \infty,$$

then condition H_3 fails. Hence, the ill-posed behavior of a large class of linear systems cannot be detected by examining a single approximation of matrix A_K, but from the sequence $\{\mathbf{A}_K\}_{K=K_0}^{\infty}$.

.

3 NURBS-based tomographic model

In industrial production, computer numerical control (CNC) machines using computer-aided design (CAD) have been widely used in reverse engineering. Non-uniform rational basis splines (NURBS) representation has been commonly used in CAD system to describe the shape of an object [43, 46, 48]. In this section, we present the measurement model of X-ray in terms of a NURBS curve.

3.1 Non-uniform rational basis splines

A NURBS curve as a piecewise rational function $\mathcal{S}$ defined on $t \in [0, 1]$, is introduced briefly here. We introduce the curve representation as a function of the parameter t as follows:

$$\mathcal{S}(t) = \sum_{i=0}^{n} \mathbf{p}_i R_{i,p}(t), \quad 0 \leq t \leq 1 \tag{3.1}$$

where the $n+1$ points $\mathbf{p}_i \in \mathbb{R}^2$ configure the curve shape. The $n+1$ points are called control points and $R_{i,p}(t)$ is the following rational function of degree p:

$$R_{i,p}(t) = \frac{\omega_i N_{i,p}(t)}{\sum_{i=0}^{n} \omega_i N_{i,p}(t)},$$

where $\{\omega_i\}$ are nonnegative weights for all i and $\{N_{i,p}(t)\}$ are basis functions that describe how strongly the control points $\{\mathbf{p}_i\}$ attract the NURBS curve. For $p > 0$, they are defined recursively as

$$N_{i,p}(t) = \frac{t - t_i}{t_{i+p} - t_i} N_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} N_{1+i,p-1}(t), \tag{3.2}$$

where $\frac{0}{0} = 0$ by definition, and for $p = 0$ as

$$N_{i,0}(t) = \begin{cases} 1 & \text{if } t_i \leq t < t_{i+1}, \\ 0 & \text{otherwise.} \end{cases}$$

A collection of K *breakpoints* is then called knot vector:

$$\mathbf{t} = [t_1, t_2, ..., t_K]^T, \tag{3.3}$$

where $K = p+n+2$ and the values of t_i satisfy the relation $0 = t_1 \leq t_2 \leq ... \leq t_K = 1$.

If the breakpoints $0 = t_1 < t_2 < ... < t_K = 1$ are evenly spaced, then $\mathbf{t}$ is called a periodic uniform knot vector. We implement a closed NURBS curve using periodic uniform knot vector to recover the boundary of the object. By repeating the first p control points after the last point, an unclamped closed NURBS curve is obtained.

Two examples of closed NURBS curves of 3rd-degree basis functions with seven control points are given in Figure 3.1. It illustrates a closed curve with the basis functions defined on the same periodic uniform knot vector:

$$[0, \frac{1}{13}, \frac{2}{13}, \frac{3}{13}, \frac{4}{13}, \frac{5}{13}, \frac{6}{13}, \frac{7}{13}, \frac{8}{13}, \frac{9}{13}, \frac{10}{13}, \frac{11}{13}, \frac{12}{13}, 1]^T.$$

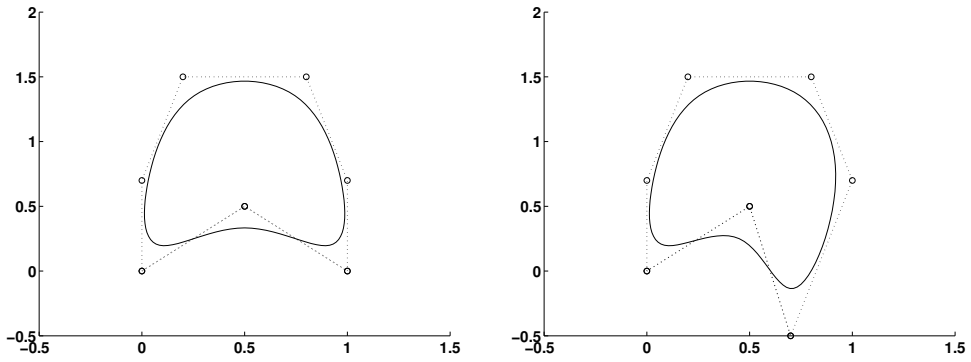


Figure 3.1: Left: NURBS curve with control points $(0, 0)$, $(0.5, 0.5)$, $(1, 0)$, $(1, 0.7)$, $(0.8, 1.5)$, $(0.2, 1.5)$, $(0, 0.7)$. Right: NURBS curve with control points $(0, 0)$, $(0.5, 0.5)$, $(0.7, -0.5)$, $(1, 0.7)$, $(0.8, 1.5)$, $(0.2, 1.5)$, $(0, 0.7)$.

In these examples and in our computations, we assume that the weights corresponding to all of the control points are the same.

3.2 NURBS-based measurement model

Consider a physical domain $\Omega \subset \mathbb{R}^2$ and a continuous tomography model $f \colon \mathbb{R}^2 \to \mathbb{R}$ where $f(x, y) \geq 0$ is defined as

$$f(x, y) = \begin{cases} c, \text{ for } (x, y) \in \Omega \\ 0, \text{ for } (x, y) \in \mathbb{R}^2 \setminus \Omega, \end{cases} \tag{3.4}$$

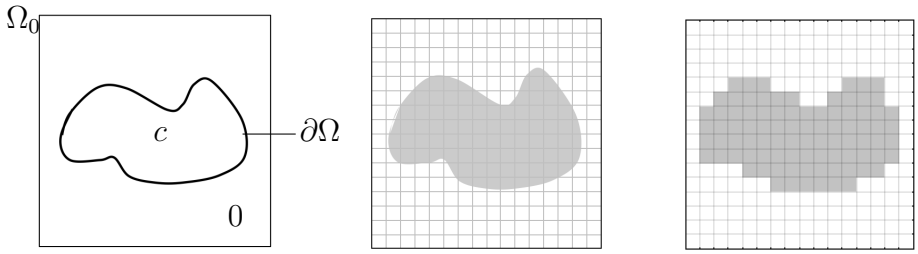


Figure 3.2: Left: the NURBS-based object model where $\partial\Omega$ is a NURBS curve. Middle: The attenuation function f defined in (3.4). The N × N pixel grid is indicated in gray. Right: the pixel-based attenuation model (3.6).

with constant X-ray attenuation $c > 0$.

Assume that the boundary $\partial\Omega$ is given by a NURBS curve as it is shown in the leftmost of Figure 3.2. Let us define

$$\int_L f(x,y)ds = c\,\ell(L \cap \Omega), \tag{3.5}$$

where ℓ is one-dimensional Lebesgue measure and L denotes a line in $(x,y)-$plane. The X-ray tomography model discussed in Section 2 enables an approximate computational method for evaluating the right-hand side of (3.5). In this work, we define the control points of the NURBS curve in polar coordinates, $\mathbf{p}_i = (r_i \cos\theta_i, r_i \sin\theta_i)$, $i = 1, \ldots, n$, where r is the radius and θ is the angle. We put together the radius and the angles for each control points as $v = [r_1, \theta_1, ..., r_n, \theta_n, c]^T \in \mathbb{R}^{2n+3}$, where n is the number of control points of the NURBS curve and c is the attenuation parameter.

We construct the interior of Ω_0 into $N \times N$ square-shaped pixels with indices (j,k) as shown in Figure 3.2 with j for the row index and k for the column. Define a nonlinear operator $\mathcal{B} : \mathbb{R}^{\mathbb{R}^{2n+3}} \to \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ by

$$\mathcal{B}(v)_{jk} = \begin{cases} c, & \text{if the center of pixel } (j,k) \text{ is inside the NURBS curve,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.6}$$

where v is the parameters defining the homogeneous object in the NURBS form, see Figure 3.2. Recalling $\mathbf{A}$ from (2.6), the measurement model can now be written as a composition:

$$\mathbf{m} = \mathbf{A}(\mathcal{B}(v)) + \varepsilon. \tag{3.7}$$

From the model (3.7), we can then recover the quantity of interest v by using MCMC as discussed in Section 6.1.

4 Multiscale framework: Wavelets and Shearlets

4.1 A brief introduction of frame theory

In this section, a brief introduction of frame theory is given. Let us denote $\mathcal{H}$ is a Hilbert space.

Definition 4.1.1. A sequence $(\varphi_i)_{i\in I}$ in $\mathcal{H}$ is called a *frame* for $\mathcal{H}$, if there exists constants $0 < A \leq B < \infty$ such that

$$A\|x\|^2 \leq \sum_{i\in I} |\langle x, \varphi_i\rangle|^2 \leq B\|x\|^2,$$

for all $x \in \mathcal{H}$. If $A = B$, then the frame is called a *tight frame*, and if $A = B = 1$ is possible, then $(\varphi_i)_{i\in I}$ is a *Parseval frame*. A frame is called *equal norm* if there exists some $c > 0$ such that $\|\varphi_i\| = c$ for all $i \in I$. If $c = 1$, then the frame is called *unit norm*.

Frame provides redundant expansions and serves as an analysis tool. If $(\varphi_i)_{i\in I}$ in $\mathcal{H}$ is a frame for $\mathcal{H}$, it allows the analysis of data via the study of the associated *frame coefficients* $(\langle x, \varphi_i\rangle)_{i\in I}$ and the operator T is defined as

$$T : \mathcal{H} \to \ell^2(I), \quad x \mapsto (\langle x, \varphi_i\rangle)_{i\in I}$$

and it is called the *analysis operator*. The adjoint of T^* of the analysis operator is called the *synthesis operator* if it satisfies

$$T^* : \ell^2(I) \to \mathcal{H}, \quad ((c_i)_{i\in I}) \mapsto \sum_{i\in I} c_i\varphi_i$$

The main operator associated with a frame (which provides a stable reconstruction process) is the *frame operator* defined as follows

$$F = T^*T : \mathcal{H} \to \mathcal{H}, \quad x \mapsto (\langle x, \varphi_i\rangle)\varphi_i,$$

with F is a positive, self-adjoint invertible operator on $\mathcal{H}$ where $A \cdot I_{\mathcal{H}} \leq F \leq B \cdot I_{\mathcal{H}}$ and $I_{\mathcal{H}}$ denotes the identity operator on $\mathcal{H}$. In the case of a Parseval frame, $F = I_{\mathcal{H}}$.

Consider a signal $x \in \mathcal{H}$. Generally, it can be recovered from its frame coefficients through the following formula

$$x = \sum_{i \in I} \langle x, \varphi_i \rangle F^{-1} \varphi_i.$$

Regarding a frame as a means for expansion in the system $(\varphi_i)_{i \in I}$, for each vector $x \in \mathcal{H}$ can be written as

$$x = \sum_{i \in I} \langle x, F^{-1} \varphi_i \rangle \varphi_i.$$

If the frame $(\varphi_i)_{i \in I}$ does not constitute a basis, then the coefficient sequence $(\langle x, F^{-1} \varphi_i \rangle)_{i \in I}$ of this expansion is not unique, and the frame is called *redundant*.

4.2 Wavelets

One of the popular mathematical tools that can be used to extract information from the large datasets is *wavelets*. The major reason for the success of wavelets is their ability to provide the sparse approximation of signals and to represent singularities efficiently. Wavelets are very suitable to approximate signals with sharp discontinuities and have been successfully used in many applications such as the algorithm of JPEG2000 and FBI fingerprint compression [22].

4.2.1 One-dimensional case

We start with a *discrete wavelet systems* in $L^2(\mathbb{R})$ that can represent efficiently the signal through the action of dilation and translation operator

$$\{\psi_{j,k} = 2^{j/2} \psi(2^j \cdot -k) : j, k \in \mathbb{Z}\}, \tag{4.1}$$

where $\psi \in L^2(\mathbb{R})$, j and k are the scale and the position index, respectively. The associated *Discrete Wavelet Transform* of a function f is given by

$$\mathcal{W}_{j,k}(f) = \langle f, \psi_{j,k} \rangle, \tag{4.2}$$

where $\langle f, \psi_{j,k} \rangle$ denotes the so-called *wavelet coefficients* [13].

The system (4.1) is called an orthonormal wavelet system if it is an orthonormal basis of $L^2(\mathbb{R})$. There is a general tool to construct orthonormal wavelet bases called multiresolution analysis (MRA).

Definition 4.2.1. In one-dimension, an MRA of $L^2(\mathbb{R})$ is a sequence of closed subspaces of $L^2(\mathbb{R})$

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots$$

with properties

(i) $\bigcup_n V_n$ is dense in $L^2(\mathbb{R})$.

(ii) $\bigcap_n V_n = \{0\}$.

(iii) $f(x) \in V_n \iff f(2x) \in V_{n+1}$ for all $n \in \mathbb{Z}$.

(iv) $f(x) \in V_n \iff f(x - 2^{-n}k) \in V_n$ for all $n, k \in \mathbb{Z}$.

(v) There exists a function $\varphi \in L^2(\mathbb{R})$ so that $\{\varphi(x-k) : k \in \mathbb{Z}\}$ is an orthonormal basis for V_0.

The basis function φ is called the *scaling function*. If φ is orthogonal, the MRA is called orthogonal.

With this approach, the decomposition of functions into different resolution level with respect *wavelet spaces* W_j, $j \in \mathbb{Z}$ and it is defined to be the orthogonal complement:

$$V_{j+1} = V_j \oplus W_j, \quad j \in \mathbb{Z},$$

where $\oplus$ is the direct sum. It decomposes a function $f_{j+1} \in V_{j+1}$ as $f_{j+1} = f_j + g_j \in V_j \oplus W_j$ with f_j contains the lower frequency component of f_{j+1}, while g_j contains its higher frequency component. It follows that

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j.$$

Another result from the MRA approach is that there always exists a function $\psi \in L^2(\mathbb{R})$ such that $\{\psi_{j,k} \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(\mathbb{R})$. Also, with MRA approach, an orthonormal basis involves both the wavelet and the scaling functions, of the form

$$\{\varphi_k = T_k\varphi = \varphi(\cdot - k) : k \in \mathbb{Z}\} \cup \{\psi_{j,k} : j \geq 0, k \in \mathbb{Z}\}.$$

The translates of the scaling function take care of the low-frequency region and the wavelet systems of the high-frequency region.

4.2.2 Two-dimensional case

In this study, we work on the image reconstruction from the CT data, so the two-dimensional case of the wavelet transform is considered. In 2D, we consider the standard tensor-product extension of the 1D wavelet transform. In detail, for the scaling function $\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k)$, the 2D wavelet system is spanned by four types of functions:

$$\begin{aligned} &\varphi_{jk}(x)\psi_{jk}(y), \\ &\psi_{jk}(x)\varphi_{jk}(y), \\ &\psi_{jk}(x)\psi_{jk}(y), \\ &\varphi_{jk}(x)\varphi_{j_0k}(y) \end{aligned} \tag{4.3}$$

where j_0 is the coarsest scale and $j, k \in \mathbb{Z}, j \geq j_0$.

The associated *2D discrete wavelet transform* of f is given by

$$C_\varphi = \frac{1}{\sqrt{XY}} \sum_{x=1}^{X} \sum_{y=1}^{Y} f(x,y)\varphi_{j_0k}(x)\varphi_{j_0k}(y) \tag{4.4}$$

$$D_\psi^h = \frac{1}{\sqrt{XY}} \sum_{x=1}^{X} \sum_{y=1}^{Y} f(x,y)\varphi_{jk}(x)\psi_{jk}(y) \tag{4.5}$$

$$D_\psi^v = \frac{1}{\sqrt{XY}} \sum_{x=1}^{X} \sum_{y=1}^{Y} f(x,y)\psi_{jk}(x)\varphi_{jk}(y) \tag{4.6}$$

$$D_\psi^d = \frac{1}{\sqrt{XY}} \sum_{x=1}^{X} \sum_{y=1}^{Y} f(x,y)\psi_{jk}(x)\psi_{jk}(y) \tag{4.7}$$

$$\tag{4.8}$$

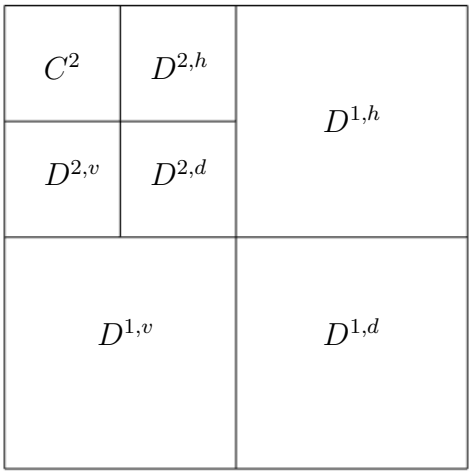
where $X, Y \in \mathbb{Z}$. An illustration of 2D discrete wavelet transform is presented in Figure 4.1.

For a function $f \in L^2(\mathbb{R}^2)$, the *best N-term approximation* f_N of f with respect to a wavelet basis is obtained by approximating f from its N largest wavelet coefficients in magnitude and it is defined as

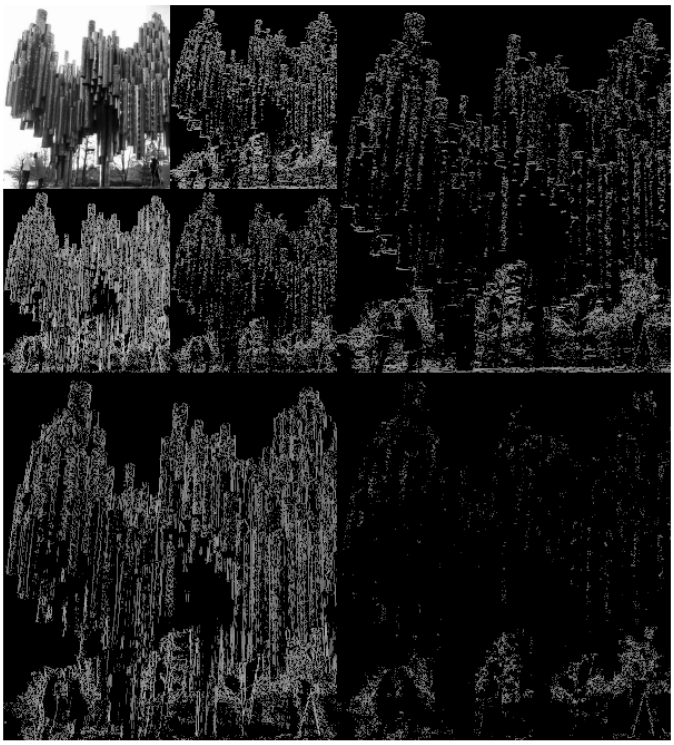
$$f_N = \sum_{\lambda \in \Lambda_N} \langle f, \psi_\lambda \rangle \psi_\lambda,$$

with Λ_N is the index set corresponding to the N largest wavelet coefficients $|\langle f, \psi_\lambda \rangle|$ assosiated with some wavelet basis $(\psi_\lambda)_{\lambda \in \Lambda}$.

(a)



(b)



(c)

Figure 4.1: An illustration of decomposition obtained using the discrete wavelet transform. The image (a) is the original Sibelius monument image, and it is denoted as C^0. Image (b) shows the schematic representation of the two-dimensional wavelet transform. The original image is high-pass filtered, yielding the three large images, as it can be seen from (c). It is then low-pass filtered and downscaled, yielding an approximation image (C^1); then this image is high-pass filtered to produce three smaller detailed images, and low-pass filtered to produce the final approximation image (C^2) in the upper-left corner of (c).

4.3 Shearlets

In 2005, a new multi scale framework called *shearlets* was introduced [23, 36]. Shearlets promote directional representation systems for multidimensional data. They have a distinctive feature to control directional selectivity or orientations by means of the shearing parameter. In order to achieve optimal sparse approximations of images that contain anisotropic singularities, the elements of shearlet systems must be distributed over various scales, locations and orientations [33]. However, the shearlet system only constitutes a frame, not a basis.

To obtain a simplified model of the object, which emphasizes anisotropic features and most notably edges, a class of functions so-called *cartoon-like images class* was introduced.

Definition 4.3.1. For fixed $\mu > 0$ and $d = \{2, 3\}$, the *class* $\mathscr{E}^2(\mathbb{R}^d)$ *of cartoon-like image* shall be the set of functions $f : \mathbb{R}^d \to \mathbb{C}$ of the form

$$f = f_0 + f_1 \chi_B, \tag{4.9}$$

where $B \subset [0,1]^d$ is a set with ∂B being a closed C^2-curve (for $d = 2$) or C^2-surface (for $d = 3$) with bounded curvature/surface and $f_i \in C^2(\mathbb{R}^d)$ with $\text{supp} f_0 \subset [0,1]^d$ and $\|f_i\|_{C^2} \leq \mu$ for each $i = 0, 1$.

For dimension $d = 2$, ∂B is assumed being a closed C^2-curve with bounded curvature, and for $d = 3$, the discontinuity ∂B shall be a closed C^2-surface with bounded principal curvature.

Let us start with the two dimensional discrete shearlet system.

Definition 4.3.2. In two-dimensional case, let $\psi \in L^2(\mathbb{R}^2)$, a *(regular) discrete shearlet system* assosiated with ψ is defined by

$$\{\psi_{j,k,m} = 2^{3/4} \psi(S_k A_{2^j} \cdot - m) : j, k \in \mathbb{Z}, m \in \mathbb{Z}^2\}, \tag{4.10}$$

where $A_2^j = \begin{bmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{bmatrix}$, is *parabolic scaling matrix*, $S_k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$, is shearing matrix, and being j is the scale index, k is the orientation index and m is the position index.

Definition 4.3.3. For $\psi \in L^2(\mathbb{R}^2)$, the *discrete shearlet transform* of $f \in L^2(\mathbb{R}^2)$ is defined as follows

$$f \quad \to \quad \mathcal{SH}(f) = \langle f, \psi_{j,k,m} \rangle, \quad (j, k, m) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2. \tag{4.11}$$

To avoid the shear parameters to become arbitrarily large, which leads to a biased treatment of directions, *cone-adapted 2D discrete shearlet systems* are introduced by dividing the frequency plane into four cones as depicted in Figure 4.2.

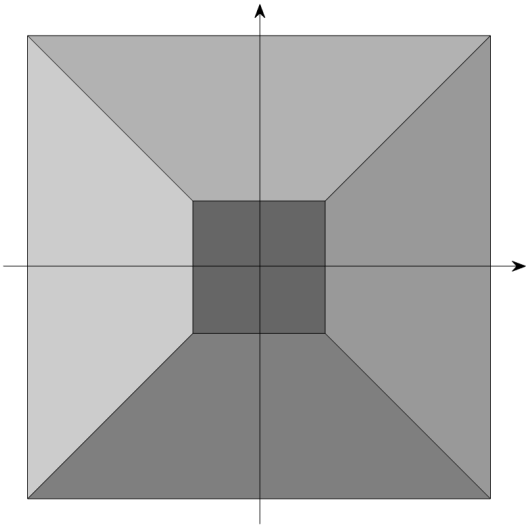


Figure 4.2: The frequency plane is partitioned into four cones. The $\Phi(\phi; c_1)$ is associated with the square (the low-frequaency region). The $\Psi(\psi; c)$ is associated with the horizontal cones, while the $\tilde{\Psi}(\psi; c)$ is associated with the vertical cones

Definition 4.3.4. For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ and $c = (c_1, c_2) \in (\mathbb{R}_+)^2$, the (regular) cone-adapted discrete shearlet system $SH(\phi, \psi, \tilde{\psi}; c)$ is defined by

$$SH(\phi, \psi, \tilde{\psi}; c) = \Phi(\phi; c_1) \cup \Psi(\psi; c) \cup \tilde{\Psi}(\tilde{\psi}; c)$$

where

$$\begin{aligned}
\Phi(\phi; c_1) =& \{\phi_m = \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2\}, \\
\Psi(\psi; c) =& \{\psi_{j,k,m} = 2^{\frac{3}{4}j}\psi(S_k A_{2^j} \cdot - M_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2)\}, \\
\tilde{\Psi}(\psi; c) =& \{\tilde{\psi}_{j,k,m} = 2^{\frac{3}{4}j}\tilde{\psi}(S_k^T A_{2^j} \cdot - \tilde{M}_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2)\},
\end{aligned}$$

with $M_c = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}$ and $\tilde{M}_c = \begin{bmatrix} c_2 & 0 \\ 0 & c_1 \end{bmatrix}$.

Figure 4.3 presents an illustration of 2D discrete shearlet transform. In three-dimensional case, the pyramid-adapted discrete shearlets are scaled according to the *paraboloidal scaling matrices* defined by

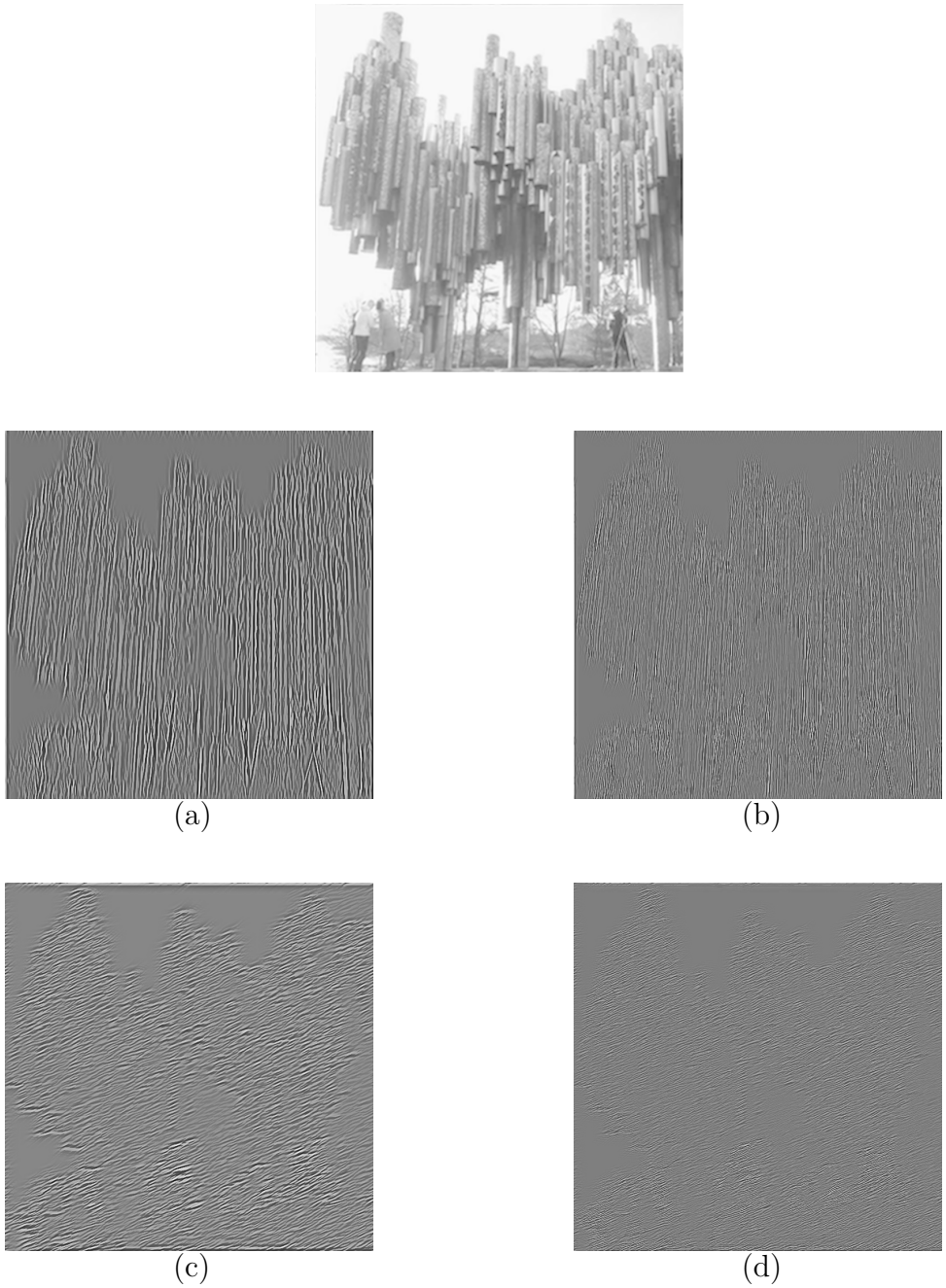


Figure 4.3: An illustration of decomposition obtained using the discrete shearlet transform from the same Sibelius monument image in Figure 4.1. The top image is the coarse-scale image. The shearlet corresponding to the coefficients in (a) has a scale parameter $j = 1$ a shearing parameter $k = 0$, and in (b) $j = 2$ and $k = 0$ and both located on the vertical frequency cones. The shearlet corresponding to the coefficients in (c) has a scale parameter $j = 1$ a shearing parameter $k = -1$, and in (d) $j = 2$ and $k = -1$ and both located on the horizontal frequency cones.

$$A_{2^j} = \begin{bmatrix} 2^j & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^{j/2} \end{bmatrix}, \tilde{A}_{2^j} = \begin{bmatrix} 2^j/2 & 0 & 0 \\ 0 & 2^j & 0 \\ 0 & 0 & 2^{j/2} \end{bmatrix}, \check{A}_{2^j} = \begin{bmatrix} 2^{j/2} & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^j \end{bmatrix},$$

where $j \in \mathbb{Z}$. The directionality is represented by the shearing matrices

$$S_k = \begin{bmatrix} 1 & k_1 & k_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tilde{S}_k = \begin{bmatrix} 1 & 0 & 0 \\ k_1 & 1 & k_2 \\ 0 & 0 & 1 \end{bmatrix}, \check{S}_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k_1 & k_2 & 1 \end{bmatrix},$$

where $k_1, k_2 \in \mathbb{Z}^2$.

In the 3D shearlet systems, we will make use of the vector notation $|k| \leq K$ for $k = (k_1, k_2)$ and $K > 0$ to denote $|k_1| \leq K$ and $|k_2| \leq K$.

Definition 4.3.5. For $c = (c_1, c_2) \in (\mathbb{R}_+)^2$, the *pyramid-adapted discrete shearlet system* $SH(\phi, \psi, \tilde{\psi}, \check{\psi}; c)$ generated by $\phi, \psi, \tilde{\psi}, \check{\psi}; c \in L^2(\mathbb{R}^3)$ is defined by

$$SH(\phi, \psi, \tilde{\psi}, \check{\psi}; c) = \Phi(\phi; c_1) \cup \Psi(\psi; c) \cup \tilde{\Psi}(\tilde{\psi}; c) \cup \check{\Psi}(\check{\psi}; c),$$

where

$$\begin{aligned}
\Phi(\phi; c_1) =& \{\phi_m = \phi(\cdot - m) : m \in c_1\mathbb{Z}^3\}, \\
\Psi(\psi; c) =& \{\Psi_{j,k,m} = 2^j \psi(S_j A_{2^j} \cdot - m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in M_c\mathbb{Z}^3)\} \\
\tilde{\Psi}(\tilde{\psi}; c) =& \{\tilde{\psi}_{j,k,m} = 2^j \tilde{\psi}(\tilde{S}_j \tilde{A}_{2^j} \cdot - m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in M_c\tilde{\mathbb{Z}}^3)\} \\
\check{\Psi}(\check{\psi}; c) =& \{\check{\psi}_{j,k,m} = 2^j \check{\psi}(\check{S}_j \check{A}_{2^j} \cdot - m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in M_c\check{\mathbb{Z}}^3)\}
\end{aligned}$$

where $j \in \mathbb{N}_0$ and $k \in \mathbb{Z}^2$ with $M_c = \mathrm{diag}(c_1, c_2, c_2)$, $\tilde{M}_c = \mathrm{diag}(c_2, c_1, c_2)$, and $\check{M}_c = \mathrm{diag}(c_2, c_2, c_1)$ for $c_1 > 0$ and $c_2 > 0$.

The frequency plane into four cones is presented in Figure 4.4.

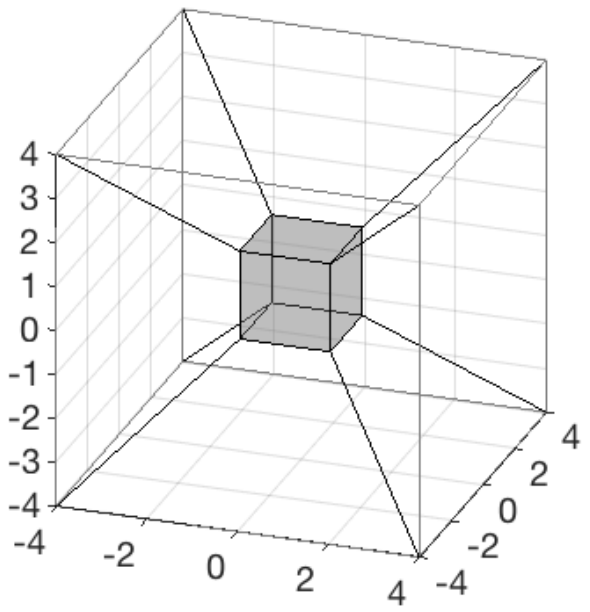


Figure 4.4: The frequency plane is partitioned into six pyramids. The $\Phi(\phi; c_1)$ is associated with the cube (the low-frequency region). The arrangement of the six pyramids is indicated by the *diagonal* lines.

5 Bayesian framework

Bayesian inversion arises from the need of interpreting real-world observations, which often contain only incomplete information is available. In consequence, the outcome is not fully predictable. The *randomness* appears due to the lack of information. Fortunately, we usually have prior knowledge or *belief* about the circumstances and it can compensate the incompleteness of the information.

The objective of Bayesian approach is to extract information and obtain the uncertainty about the unknown quantity based on:

1. the information of the measurement process, and
2. the knowledge about the unknown quantity prior to the measurement.

Both the measurement and our *belief* about the outcome are modelled as random variables. Instead of producing a single estimate as in deterministic inversion, Bayesian inversion approach provides probability distribution as a solution that contains information of the variables.

Let us first briefly discuss random variable and probability.

Definition 5.0.1. Given Ω as a *sample space*. A *random variable* X in $\mathbb{R}^n$ is defined by

$$X \colon \Omega \mapsto \mathbb{R}^n.$$

Here random variables are denoted by capital letters and their realizations by lowercase letters, *i.e.* $X(\omega) = x \in \Omega$. The probability distribution function $P : \mathbb{R}^n \mapsto [0\ 1]$ of the random variable X is defined as

$$P(x) = \Pr(X \leq x),$$

with $\Pr(X \leq x)$ denotes the probability of the event $X \leq x$, where $X \leq x := \{X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n\}$. Consider continous random variables $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$. The joint probability distribution function $P : \mathbb{R}^{n+m} \mapsto [0\ 1]$ of X and Y is defined as

$$P(x, y) = \Pr(X \leq x, Y \leq y).$$

By assuming that the joint probability distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{n+m}$, there exists a measurable function $\pi(x, y)$, so-called the joint probability density of X and Y such that

$$P(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \pi(a,b)\,\mathrm{d}a\mathrm{d}b.$$

We introduce the conditional probability density of X given $Y = y$ as follows:

$$\pi(x|y) = \frac{\pi(x,y)}{\pi(y)},$$

where a normalization constant $\pi(y) > 0$ is marginal density of y. By constructing the conditional probability density of Y given $X = x$, *i.e.*, $\pi(x,y) = \pi(y|x)\pi(x)$, we can form the *Bayes formula*:

$$\pi(x|y) = \frac{\pi(y|x)\pi(x)}{\pi(y)}, \quad \pi(y) > 0.$$

5.1 Bayesian model for inverse problem

This section discusses the formulation of our model (2.6) in terms of a Bayesian framework. Recall that in statistical inverse problems, all parameters are viewed as random variables. We write the measurement model containing an additive noise E:

$$M = \mathbf{A}F + E,$$

where M, F and E are random variables and $\mathbf{A}$ is the measurement matrix.

Let us consider Gaussian case where $E = \varepsilon$ and $\varepsilon \sim \mathcal{N}(0,\Sigma)$. Hence, up to normalization constant, the probability distribution of the noise with covariance Σ is

$$\pi(\mathbf{m}\,|\,\mathbf{f}) \propto \exp\Big(-\frac{1}{2}(\mathbf{A}\mathbf{f}-\mathbf{m})\Sigma^{-1}(\mathbf{A}\mathbf{f}-\mathbf{m})\Big), \tag{5.1}$$

where $\mathbf{f}$ and $\mathbf{m}$ are the realizations of F and M, respectively. The formula (5.1) is called the *likelihood* function. If $\Sigma = \sigma^2 I$, where I is identity matrix and variance σ^2, then

$$\pi(\mathbf{m}\,|\,\mathbf{f}) \propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{f}-\mathbf{m}\|_2^2). \tag{5.2}$$

Prior to the measurement, some information about the unknown needs to be exploited to compensate the missing information in the measurement [29, 31, 50]. The prior construction is a crucial step in the statistical inverse problem. Typically, our prior knowledge of the unknown is *qualitative* in nature. The challenging task is to translate this qualitative information into a *quantitative* form that can be postulated into a mathematical form which is written down as a density $\pi(\mathbf{f})$. It is quite common that Gaussian probability density is used for the prior, called *Gaussian prior*. This is because it is easy to construct.

The solution of the inverse problem is the *posterior* distribution which has the following form:

$$\pi(\mathbf{f} \,|\, \mathbf{m}) \propto \pi(\mathbf{f})\pi(\mathbf{m} \,|\, \mathbf{f}).$$

There are well-known approaches to estimate the unknown parameter: *maximum a posteriori* (MAP) and *conditional mean* (CM) estimates. Both are point estimators.

5.2 Conditional mean estimate

A popular method to get an estimate of $\mathbf{f}$ is to consider conditional mean as it is defined as follows:

$$\mathbf{f}^{\mathrm{CM}} = \int_{\mathbb{R}^N} \mathbf{f}\, \pi(\mathbf{f} \,|\, \mathbf{m}) \mathbf{df}. \tag{5.3}$$

High-dimensional integration problem appears in (5.3), hence we shall employ an MCMC technique for solving it [9, 20, 21]. In the following section, details of MCMC method are discussed.

5.2.1 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is one of the most popular techniques for sampling from probability distributions, and is based on the construction of a Markov chain that has the desired distribution as its equilibrium distribution.

The basic idea of Monte Carlo integration is to generate samples $\{x^{(1)}, x^{(2)}, \cdots, x^{(\ell)}\} \subset \mathbb{R}^M$ from $\pi(x)$. Monte Carlo estimate is obtained by sampling a set of realizations from the distribution $\pi(x)$ *i.e.* $x^{(\ell)} \overset{i.i.d}{\sim} \pi(x)$ and then approximate the integral $f(x)$ by the sample mean

$$\int f(x)\pi(x)dx \approx \frac{1}{K} \sum_{\ell=1}^{L} f(x^{\ell}). \tag{5.4}$$

With the increasing availability of computer power, Monte Carlo techniques are being increasingly used. Monte Carlo methods are especially useful for simulating systems with many coupled degrees of freedom.

In general, generating samples $\{x^{(\ell)}\}$ independently from $\pi(\cdot)$ is often not feasible. However, if the sequence $\{x^{(\ell)}\}$ is dependent and as long as it is *Markov chain*, then Monte Carlo integration still converges. This holds since *Markov chain* has $\pi(\cdot)$ as its equilibrium distribution [29]. This technique is then called *Markov chain* Monte Carlo.

In detail, let $X = \{X_1, X_2, ...X_M\}$ be a real valued random variable in $\mathbb{R}^M$. A *Markov chain* is a sequence of random variables $\{X^{(\ell)}\} \subset \mathbb{R}^M$ which has property that when all values of the sequence is available, the distribution of the state $X^{(\ell+1)}$ only depends on the state $X^{(\ell)}$. In other words, at each time $\ell \geq 1$, the next state is sampled which depends only on the current state, and it is written as below:

$$\pi(X^{(\ell+1)} \leq y | X^{(1)} = x^{(1)}, X^{(2)} = x^{(2)}, ..., X^{(\ell)} = x) = \pi(X^{(\ell+1)} \leq y | X^{(\ell)} = x). \tag{5.5}$$

When the probability in (5.5) does not depend on ℓ or time-independent, the chain is called *homogeneous*. For a homogeneous Markov chain, the transition kernel from state ℓ to state $\ell + 1$ is defined by

$$\mathrm{T}(x, y) = \pi(X^{(\ell+1)} \leq y | X^{(\ell)} = x).$$

The homogeneous Markov chain has a property that as the chain runs for a long time, $\ell \to \infty$, the chain will reach an equilibrium called *stationary distribution*. Few concepts concerning transition kernels is still needed to be introduced. The transition kernel T is *irreducible* if there exists finite number m of transitions such that for any $x, y \in \Omega$, $\mathrm{T}^m(x, y) > 0$. In other words, irreducibility means regardless the present state, the chain can reach any other state in finite time. Now let $\mathcal{M}(x) = \{m \geq 1 : \mathrm{T}^m(x, x) > 0\}$ as the set of times when it is possible for the chain to return to starting position x. The *period* of state x is defined as the greatest common divisor (gcd) of $\mathcal{M}(x)$. The irreducible T is called *aperiodic* if all the states have period 1 [54].

Back to the integral of the form (5.4), the convergence is guaranteed if the irreducibility and aperiodicity of the Markov chain are hold.

5.2.1.1 Metropolis-Hastings

In recent years, statisticians have been increasingly drawn to MCMC methods to simulate nonstandard multivariate distributions. The Gibbs sampling algorithm is one of the best known of these methods but a considerable attention is now being devoted to the Metropolis-Hasting algorithm [8, 9, 20, 21, 51].

Consider the proposal distribution or *candidate-generating density* denoted by $q(x, y)$ with the property $\int q(x, y)dy = 1$. The density is to be interpreted as the probability of a *candidate* value y generated from $q(x, y)$ when the current state is x. The candidate y can be either accepted or rejected. It brings us to the so-called *acceptance ratio* or *probability of move*. If the move is not made, then the next state remains the same as the current state, *i.e* v, instead of y. As it is studied in [8], to obtain an invariant distribution of $\pi(\cdot)$, the *reversibility condition* (also called *detailed balance*) must be hold:

$$\pi(x)q(x, y) = \pi(y)q(y, x) \tag{5.6}$$

However, we might end up in the situation where the process moves from x to y is more often than from y to x:

$$\pi(y)q(y,x) < \pi(x)q(x,y).$$

Thus a correction factor, α, needs to be determined. Our aimed is to choose α so that

$$\pi(y)\alpha(y,x)q(y,x) = \pi(x)\alpha(x,y)q(x,y),$$

and it can be achieved by setting

$$\alpha(y,x) = 1, \quad \alpha(x,y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} < 1.$$

By reversing the x and y, to obtain the reversibility in (5.6) then

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}.$$

As a remark, the calculation of $\alpha(x,y)$ does not require the value of normalizing constant of $\pi(\cdot)$, since it appears both in the denominator and numerator. Hence, formidable task can be avoided. We point out if the proposal distribution $q(\cdot)$ is symmetric, *i.e* $q(x,y) = q(y,x)$, then the acceptance ratio reduces to $\frac{\pi(y)}{\pi(x)}$, which means if $\pi(y) \geq \pi(x)$, the candidate y is accepted as the next state, otherwise it is accepted with the probability $\frac{\pi(y)}{\pi(x)}$. The Metropolis-Hastings algorithm takes the form:

1. Initialize value $x^{(\ell)} \in \mathbb{R}^N$, and set $\ell = 1$
2. Draw $y \in \mathbb{R}^N$ from the proposal distribution $q(x^{(\ell)}, y)$ and calculate the acceptance ratio
$$\alpha(x^{(\ell)}, y) = \min\left\{1, \frac{\pi(y)q(y,x^{(\ell)})}{\pi(x^{(\ell)})q(x^{(\ell)},y)}\right\}.$$
3. Draw $t \in [0,1]$ from uniform probability density
4. If $\alpha(x^{(\ell)}, y) \geq t$, accept the candidate by setting $x^{(\ell+1)} = y$, else $x^{(\ell+1)} = x^{(\ell)}$
5. If $\ell = L$ (the number of iteration) then stop, else set $\ell = \ell + 1$ and return to step 2.

A good rule of thumb of the candidate-states ys should be accepted is around $20\% - 30\%$ [45].

5.2.1.2 Delayed Rejection Adaptive Metropolis

Combination of two ideas, Delayed rejection and Adaptive Metropolis, was proposed in [24] to improve efficiency of the resulting MCMC estimators.

- **Delayed rejection (DR).** DR is an improvement of Metropolis Hastings by reducing the number of rejected proposals [39]. Suppose that the current state of the Markov chain is $X^{(\ell)} = x$, and a candidate move y_1 is generated from a proposal $q_1(x, \cdot)$ with acceptance probability

$$\alpha_1(x, y_1) = \min\left\{1, \frac{\pi(y_1)q_1(y_1, x)}{\pi(x)q_1(x, y_1)}\right\}. \tag{5.7}$$

  In DR, if the proposal state is rejected, instead of remaining in the same position, a second stage is proposed. The second stage proposal depends not only on the current state but also on the one we have just proposed and rejected. The acceptance probability of the second stage proposal:

$$\alpha_2(x, y_1, y_2) = \min\left\{1, \frac{\pi(y_2)q_1(y_2, y_1)q_2(y_2, y_1, x)[1 - \alpha_1(y_2, y_1)]}{\pi(x)q_1(x, y_1)q_2(x, y_1, y_2)[1 - \alpha_1(x, y_1)]}\right\}. \tag{5.8}$$

  This process can be iterated further for higher stage proposals.

- **Adaptive Metropolis (AM)**. It is well known that in MCMC algorithm, the choice of proposal distribution is very crucial. In one hand, if the proposal distribution is too large, the convergence is slow since too many candidate points are rejected. On the other hand, if the proposal distribution is too small, too many candidate points are accepted. Therefore it is very important to choose the proposal distribution so that the computational efficiency is obtained. Adaptive Metropolis algorithm was proposed to tune the proposal distribution suitably updating the covariance from all of the previous states [25, 26, 52]. The core idea in AM is to construct a Gaussian proposal distribution with a covariance matrix calibrated using the history of the MCMC chain. In detail, after generating initial ℓ_1 samples using standard process (such as MH or DR), the Gaussian proposal to be centered at the current state of the Markov chain and its covariance is set to be

$$\Sigma_\ell = \begin{cases} \Sigma_0, & \ell \leq \ell_1 \\ s_d \mathrm{Cov}(X^0, \cdots, X^{\ell-1}) + s_d \varepsilon I_d, & \ell > \ell_1 \end{cases}$$

  with s_d is the scaling factor which depends on the dimension d of the state space on which the posterior distribution is defined, the small number of $\varepsilon > 0$ to prevent the covariance matrix becoming singular, and I_d denotes the d-dimensional identity matrix.

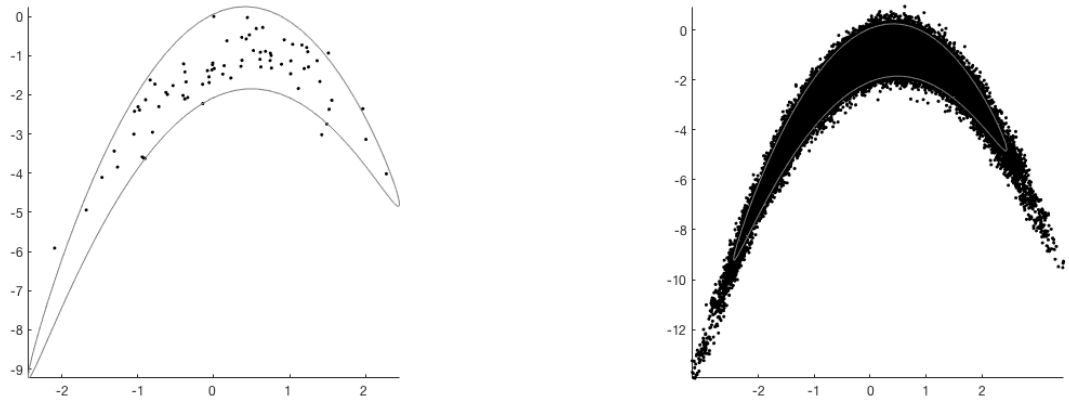


Figure 5.1: A demonstration of Metropolis-Hasting (MH) (Left) and Delayed Rejection Adaptive metropolis (DRAM) (Right) for drawing chains in a banana shaped distribution. Each of them were run for 100 000 samples. It is clearly seen that the samples produced by the DRAM method is well mixed, while the samples from MH has a very slow movement to cover the important region.

In general, MCMC can perform well if the proposal distribution fits the target distribution. In DR strategy, a given number of fixed proposals is used for different stages. Thus, it is very important to have at least one of the proposals is successfully calibrated. The idea of AM is to efficiently tune the proposals based on the information obtained during the run of the chain. The DR and AM strategies are combined to obtain more efficient method. The outline of implementing the method is given as follows:

- Assume that we use m-stages DR algorithm. The AM process is applied at the first stage of DR: the Σ_ℓ^1 is computed from the sampled chain (which is independent with m)
- The covariance Σ_n^i of the proposal for the i-th stage, where $i = 2, \cdots, m$ is computed as $\Sigma_\ell^i = \gamma_i \Sigma_\ell^1$. The γ_i is freely chosen.

A demonstration of drawing samples using MH and DRAM techniques is given in Figure 5.1. With 100 000 iterations, samples produced by DRAM mix very well to the banana-area, while for MH, the candidate-samples are too often rejected.

5.3 Maximum a posteriori estimate

Maximum a posteriori (MAP) estimate is one of the most popular statistical estimates that equals the highest mode of the posterior distribution. The MAP estimate $\mathbf{f}_{\text{MAP}}$ satisfies

$$\mathbf{f}_{\text{MAP}} = \arg\max_{\mathbf{f}} \pi(\mathbf{f} \mid \mathbf{m}).$$

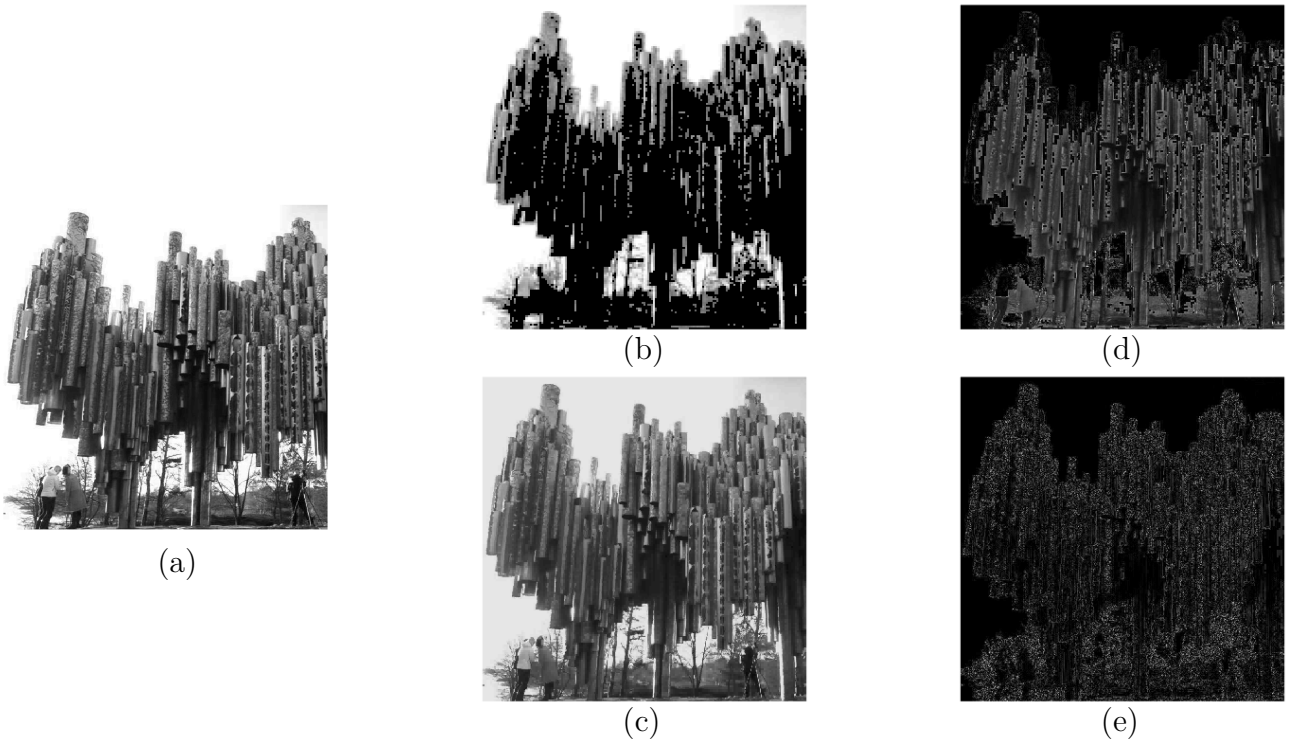
Often, optimization methods are used to find the MAP estimate. Typically, the optimization strategies such as iterative methods are carried out to search the maximizer and they lead to the same computational problem as with the deterministic regularization methods [18, 29, 40, 53].

5.3.1 Sparsity-promoting regularization

It is quite often that we need to deal with complicated signal or big data in which the computational task becomes very challenging. However, actually, some dominant or important features in the signal or data can be exploited as a representation of the signal. This idea promotes sparsity in the sense that the less dominant features are not taken into account in the signal extraction. With this sparsity-promoting idea, we now only face the smaller data. One of well-known example is *Compressed Sensing* [16, 4, 17] and they study that a finite-dimensional signal can be recovered or well-approximated as a linear combination of just a few elements from a known basis or dictionary. If the representation is exact, the signal is-called *sparse*. Sparsity can be thought of as *Occam's razor* principle: when many possible representations can predict equally well, the simplest choice (in this case, the fewest number of elements) should be selected.

Now, we go back to our example using a multiscale wavelet transform [38]. As it can be seen in Figure 4.1, the lowest frequency components provide a coarse scale approximation of the image, and the higher frequency components provide the detail and resolve the edges of the image. In the same figure, we see that many of the coefficients are very small, thus we can obtain a good approximation of the image by only using a smaller number of wavelet coefficients, or in other words, by setting the small coefficients to zero. Figure 5.2 shows the sparse recovery images using two different numbers of nonzero wavelet coefficients. By only keeping the largest 10% of the wavelet coefficients, we can still achieve a good recovery image, for instance, the edges are recovered well.

Now we move to another inverse problem application, such as X-ray tomography where the X-ray measured data is not only noisy but also the complete data is not available. In this case, the problem is, of course, *ill-posed*. We are interested in



(a) (b) (c) (d) (e)

Figure 5.2: An illustration of sparse representation using the wavelet transform. Image (a) is the original image, images (b) and (c) show 3% and 10% of the largest wavelet coefficients which are kept to represent the approximation images, respectively. Images (d) and (e) present the difference between the original image and the representation images (b) and (c), respectively. By compressing the original image by using only 10% of the largest wavelet coefficients, the compressed image is still presenting a good quality image.

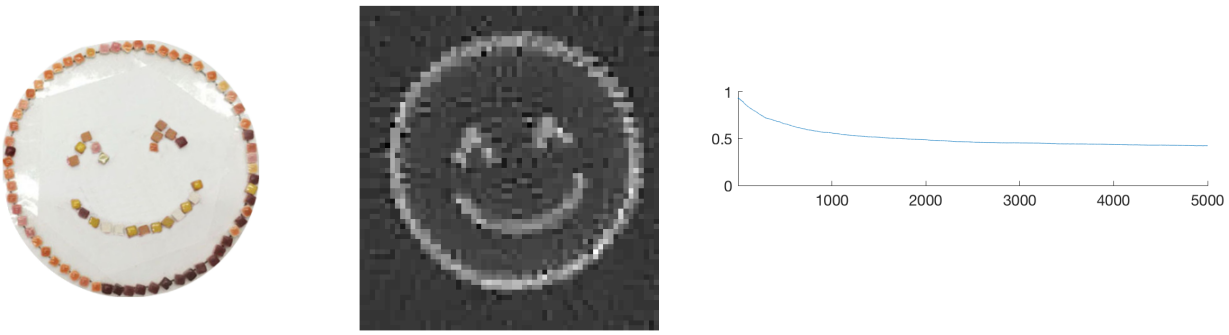


Figure 5.3: The implementation of ISTA algorithm to X-ray tomography. Left: The emoji data built from ceramic stones. Middle: ISTA reconstruction using 60 projection images out of 360° with fix value of thresholding parameter μ. Right: The ratio of the non-zero wavelet coefficients against the number of iterations.

finding the vector $\mathbf{f}$ that minimizes the variational regularization functional

$$\mathbf{f}_{\mathrm{S}} = \underset{\mathbf{f}\in\mathbb{R}^{N^2}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \mu \sum_{\gamma} |\langle \mathbf{f}, \psi_\gamma \rangle| \right\}, \tag{5.9}$$

where $\mathbf{A}$ and $\mathbf{m}$ are modelled as in Section 2, and the parameter μ describes a trade-off between emphasizing the data fidelity term and the regularizing penalty term. The term $\sum_\gamma |\langle \mathbf{f}, \psi_\gamma \rangle|$ could be replaced by $\|\mathbf{W}\mathbf{f}\|_1$, where $\mathbf{W}$ could be the wavelet or shearlet transform. For orthonormal basis $\{\psi_\gamma\}$, one of popular techniques to solve the problem above is the iterative shrinkage/thresholding algorithm (ISTA) introduced by [14]. It iterates

$$\mathbf{f}^n = \mathbf{S}_\mu(\mathbf{f}^{n-1} + \mathbf{A}^T(\mathbf{m} - \mathbf{A}\mathbf{f}^{n-1})),$$

where n is the current iteration and S_μ is the soft-thresholding function defined by

$$\mathcal{T}_\mu(x) = \begin{cases} x + \frac{\mu}{2} & \text{if} \quad x \leq -\frac{\mu}{2} \\ 0 & \text{if} \quad |x| < \frac{\mu}{2} \\ x - \frac{\mu}{2} & \text{if} \quad x \geq -\frac{\mu}{2}. \end{cases}$$

Figure 5.3 illustrates the tomographic reconstruction using the ISTA algorithm. Nevertheless, convergence rate for a constrained problem, such as non-negativity

constraints, is not taken into account in [14, 37]. However, in tomographic problems, enforcing non-negativity on the attenuation coefficients is highly desired. This is based on the physical fact that the X-ray radiation can only attenuate inside the target, not strengthen. Thus, the problem we need to solve reads as:

$$\mathbf{f}_{\mathrm{S}} = \underset{\mathbf{f}\in\mathbb{R}^{N^2},\,\mathbf{f}\geq 0}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{A}\mathbf{f}-\mathbf{m}\|_2^2 + \mu\|\mathbf{W}\mathbf{f}\|_1 \right\}, \tag{5.10}$$

where the inequality $\mathbf{f} \geq 0$ is meant component-wise. In their seminal paper [7], Peijun Chen, Jianguo Huang, and Xiaoqun Zhang showed that the minimizer of (5.10) can be computed using the primal-dual fixed point (PDFP) algorithm:

$$\begin{aligned} \mathbf{y}^{(i+1)} &= \mathbb{P}_C\Big(\mathbf{f}^{(i)} - \tau\nabla g(\mathbf{f}^{(i)}) - \lambda\mathbf{W}^T v^{(i)}\Big) \\ \mathbf{v}^{(i+1)} &= \Big(I - \mathcal{T}_\mu\Big)\Big(\mathbf{W}\mathbf{y}^{(i+1)} + v^{(i)}\Big) \\ \mathbf{f}^{(i+1)} &= \mathbb{P}_C\Big(\mathbf{f}^{(i)} - \tau\nabla g(\mathbf{f}^{(i)}) - \lambda\mathbf{W}^T v^{(i+1)}\Big) \end{aligned} \tag{5.11}$$

where τ and λ are positive parameters, $g(\mathbf{f}) = \frac{1}{2}\|\mathbf{A}\mathbf{f}-\mathbf{m}\|_2^2$, and $\mathcal{T}$ is the soft-thresholding operator. Here $\mu > 0$ represents the thresholding parameter, while τ and λ are parameters that need to be suitably chosen to guarantee convergence. In detail, $0 < \lambda < 1/\lambda_{\max}(\mathbf{W}\mathbf{W}^T)$, where $\lambda_{\max}$ denotes the maximum eigenvalue, and $0 < \tau < 2/\tau_{\mathrm{lip}}$, being τ_{lip} the Lipschitz constant for $g(\mathbf{f})$. Furthermore, in (5.11) the non-negative "quadrant" is denoted by $C = \mathbb{R}_+^{N^2}$ and $\mathbb{P}_C$ is the Euclidean projection. In other words, $\mathbb{P}_C$ replaces any negative elements in the input vector by zero.

5.3.2 Automated regularization parameter tuning

In this work, a novel tomographic reconstruction strategy using sparsity-based regularization parameter is presented. The method requires *a priori* knowledge about the number of nonzero wavelet/shearlet coefficients of the unknown quantity. The penalty term promotes sparsity of expansion of $\mathbf{f}$ with respect to ψ_γ, which in this research ψ_γ can be wavelet systems or shearlet systems. Choosing a suitable regularization parameter is a difficult problem. Here, a novel method for automatically choosing the parameter based on sparsity is introduced. The main idea of the method is that the parameter μ in (5.10) is automatically tuned based on a control algorithm which drives the sparsity of the reconstruction to *a priori* known ratio $0 \leq \mathcal{C}_{pr} \leq 1$ of nonzero coefficients in $\mathbf{f}$. Our approach is based on the following idea: *in sparsity-promoting regularization, it is natural to assume that the* a priori *information is given as the percentage of nonzero coefficients in the unknown.*

If $\mathcal{C}^{(i)}$ is the degree of sparsity at the current iterate, we change $\mu^{(i)}$ adaptively as follows:

$$\mu^{(i+1)} = \mu^{(i)} + \beta(\mathcal{C}^{(i)} - \mathcal{C}_{pr}), \tag{5.12}$$

where $\beta > 0$ is a parameter used to tune the controller. Ideally, $\mathcal{C}_{pr}$ is computed from many similar objects.

6 Articles Summary

The goals of the thesis are:

- to reconstruct an attenuation value and a boundary shape of a homogeneous target in terms of a NURBS curve using MCMC algorithm for sparse X-ray data,
- to recover inner structures of a target using adaptive methods for tuning regularization parameter for sparse X-ray data. In 2D, the automatic regularization method based on enforcing sparsity in the two-dimensional wavelet transform domain is implemented. While in 3D case, the three-dimensional shearlet transform domain is used as it promotes directional representation systems.

To resolve the incomplete measurement data, a good prior knowledge needs to be enforced.

In Article I, II and III, tomographic reconstructions using NURBS-MCMC method are proposed to recover the shape of the object from very sparse X-ray tomography data. In Article I, a stricter prior distribution is constructed as a preliminary study, while in Article II, a Gaussian prior is used along with the hard prior. In both articles, the method is tested using simulated data and Metropolis Hastings is implemented as the MCMC strategy. In Article III, the method is implemented to reconstruct the shape of the object from two physical X-ray data sets with the Gaussian prior and a more suitable hard prior is used. To improve the efficiency of the method, DRAM strategy is applied as the MCMC method.

Article IV and V present novel reconstruction techniques where the regularization parameter is tuned automatically based on a prior knowledge of a sparsity level of the reconstruction. The article IV proposes a new strategy so-called controlled wavelet domain sparsity (CWDS) and in the Article V controlled shearlets domain sparsity (CSDS). Both methods seem to be promising strategies, especially in real life applications where end-users could avoid manually tuning the parameters.

6.1 NURBS-MCMC strategy

As mentioned above, we assume that our target is one homogeneous object. It means that we could build prior knowledge based on the fact that we only have one

material in the object (one attenuation value) and one boundary object. Formulating a qualitative information of the prior to a quantitative form as a prior distribution is a challenging task. Therefore, spending considerable amount of time to design the prior distribution can be necessary. In this research, the design of the prior distribution is improved gradually in each article.

Let us recall the measurement model in (3.7) with measurement matrix $\mathbf{A}$ in the following form:

$$\mathbf{m} = \mathbf{A}(\mathcal{B}(v)) + \varepsilon.$$

where $\mathcal{B}$ is defined in (3.6), $\mathbf{m}$ is the X-ray measurement vector, v is our quantity of interest and ε is the measurement error. The vector v contains the radii, angles and attenuation parameters, *i.e.* $v = [r_1, \theta_1, ..., r_n, \theta_n, c]^T$, where n is the number of control points of the NURBS curve and c is the attenuation parameter. The *likelihood* is constructed as:

$$\pi(\mathbf{m} \,|\, v) \propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{A}(\mathcal{B}(v)) - \mathbf{m}\|_2^2).$$

In each article, the prior distribution is constructed in order to:

- avoid self-intersecting NURBS-curve, and
- avoid the uncontrollable rotational movement of the control point set.

Prior distribution in Article I With respect to the parameter angles, the prior is constructed as follows

$$\mathbf{A}(\theta_i) = \begin{cases} 1 - \frac{|\theta_i - \theta_i'|}{45} & \text{for } \theta_i' - 45 \leq \theta_i \leq \theta_i' + 45 \\ \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta_i' = 45i$, where i is the index of the i-th control point.

With respect to the radius parameter, the prior is constructed as follows

$$\mathbf{R}(r_i) = \begin{cases} 1 - \frac{|r_i - 1|}{15} & \text{for } 0 \leq r_i \leq 15 \\ \\ 0, & \text{otherwise.} \end{cases}$$

Then we write

$$\pi(v) = \mathbf{A}(\theta_i) \cdot \mathbf{R}(r_i).$$

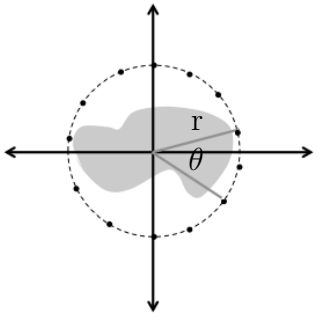


Figure 6.1: An illustration of the vector mean $\widetilde{v}$ in (6.2). The dashed circle is the NURBS curve produced by the control points in $\widetilde{v}$ with a fixed uniform periodic knot vector.

Prior distribution in Article II We formulate the prior formula as follows

$$\pi(v) = \begin{cases} \exp(-\frac{1}{2\sigma_2^2}\|v-\tilde{v})\|_2^2), & \text{for } 0 \le r_i \le r_i^{\max} \text{ and } 0.1 < c < c_{\max} \\ & \text{and } \theta_i^{\min} \le \theta_i \le \theta_i^{\max}, \\ 0 & \text{otherwise,} \end{cases} \tag{6.1}$$

with the mean $\widetilde{v} = [\widetilde{r_1}, \widetilde{\theta_1}, ..., \widetilde{r_n}, \widetilde{\theta_n}, \widetilde{c}]^T$.

Prior distribution in Article III We assume that the attenuation function f is supported in the disc $\{(x,y) \,|\, \sqrt{x^2+y^2} \le r\} \in \mathbb{R}^2$ with a known radius $r > 0$. In terms of control points, this condition means that $|\mathbf{p}_i| \le 2r$.

Assume that the prior information is Gaussian distributed with variance σ_0. Then a prior information has the following quantitative form

$$\pi(v) \propto \exp(-\frac{1}{2\sigma_0^2}\|v-\widetilde{v}\|_2^2), \tag{6.2}$$

with the mean $\widetilde{v} = [\widetilde{r_1}, \widetilde{\theta_1}, ..., \widetilde{r_n}, \widetilde{\theta_n}, \widetilde{c}]^T$. Figure 6.1 depicts the prior information.

In addition, to avoid uncontrollable rotational movement of the control point set, we impose hard constraints on the angles θ_i:

$$\max\left\{\theta_{i-2}, \Gamma_i^{\mathrm{m}}\right\} \le \theta_i \le \min\left\{\theta_{i+2}, \Gamma_i^{\mathrm{M}}\right\}, \tag{6.3}$$

where Γ_i^{m} is a lower bound for θ_i and Γ_i^{M} is an upper bound for θ_i. The radii are restricted as well as: $0 \le r_i \le r^{\mathrm{M}}$, where $\Gamma_i^{\mathrm{m}}, \Gamma_i^{\mathrm{m}}, r^{\mathrm{M}} \in \mathbb{R}$.

To minimize oscillations in the curve, the following condition should be satisfied

$$||\mathbf{d} - \mathbf{p}_i|| \le k||\mathbf{p}_{i-1} - \mathbf{p}_{i+1}||,$$

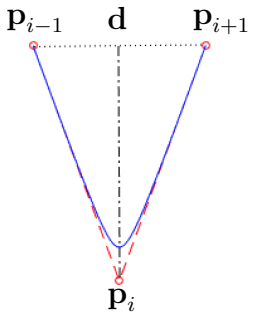


Figure 6.2: The illustration of hard prior is given, where $\mathbf{p}_{i-1}, \mathbf{p}_i$ and $\mathbf{p}_{i+1}$ are the control points and $\mathbf{d} = (\frac{x_{i-1}+x_{i+1}}{2}, \frac{y_{i-1}+y_{i+1}}{2})$.

where $\mathbf{p}_{i-1}, \mathbf{p}_i$ and $\mathbf{p}_{i+1}$ (recall that $\mathbf{p}_i = (x_i, y_i)$) are the control points as in Figure 6.2, $\mathbf{d} = (\frac{x_{i-1}+x_{i+1}}{2}, \frac{y_{i-1}+y_{i+1}}{2})$ and k is a constant. Finally, another hard prior is added to avoid an intersection of the control polygons.

The *posterior* distribution is then:

$$\pi(v \,|\, \mathbf{m}) \propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{A}(\mathcal{B}(v)) - \mathbf{m}\|_2^2)\pi(v) \tag{6.4}$$

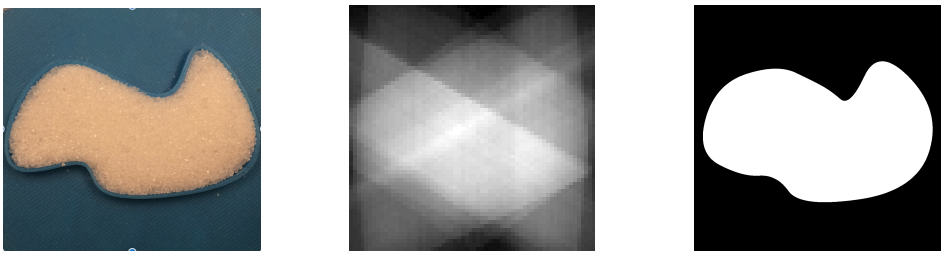
with respect to corresponding prior distribution.

In Article I and II, Metropolis Hastings algorithm is implemented to generate samples from posterior distribution. Objects with different shapes are simulated to test the method. For Article III, simulated and real data of crystal sugar are used and tomographic reconstructions are computed using DRAM strategy. In all results, the point estimate from conditional mean is calculated to obtain the estimation of v, the control points of the NURBS curve and the attenuation value of the object.

6.1.1 Computational results of NURBS-MCMC method

An example of computational results of physical phantom using NURBS-MCMC for formulation in Article III is presented in Figure 6.3, the image (a) is the phantom filled by crystal sugar. We set the number of control points to be 12. We use the following measurement setup: sparse-angle (6 angles) from full 360° angles of fan beam geometry are employed. Thus, the total number of our parameters of interest are 25 ($2n+1$, where n is total number of control points). We used DRAM package, available at `http://helios.fmi.fi/~lainema/mcmc/`. We set ℓ_1 to be 100.

FBP reconstruction (in the middle of the picture) is computed as well for comparison. It is clearly seen that using very sparse projection images, the NURBS-MCMC method outperforms FBP, as evaluated by visual inspection. The NURBS-MCMC result recovers the boundary of the target better. The non-convex shapes of the target are represented well by the proposed method. The binary image is automatically generated since the reconstruction is in the NURBS form.



(a) (b) (c)

Figure 6.3: Reconstruction of a physical phantom with 6 projections from full 360° directions: (a) the ground truth. (b) filtered backprojection reconstruction and (c) NURBS-MCMC reconstruction.

Figure 6.4 presents the evolutions of the MCMC chain of the radii. It shows that after a long run, the chains are mixing properly. The multimodality of marginal distribution of angle parameters is presented in Figure 6.5. The final result is obtained by computing the conditional mean (CM estimate) of the chains of radii and angles (control points) and the attenuation coefficient.

6.2 Automatically tuning parameter methods in sparsity domain for X-ray tomography

In Article IV and V, X-ray tomographic reconstructions are done using adaptive methods for tuning the regularization parameter using wavelets and shearlets. In Article IV, we propose a method so-called controlled wavelet domain sparsity (CWDS) and implement it to 2D X-ray tomography problem. In Article V, instead of using wavelets, we propose the controlled shearlet domain sparsity (CSDS) algorithm to 3D X-ray tomography case. We use primal-dual fixed point (PFDP) algorithm introduced by [7] and adapt regularization parameter as it is discussed in Subsection 5.3.2 using the algorithm (5.11).

For Article IV, we consider the matrix underlying the wavelet transform, which we shall denote by $\mathbf{W} \in \mathbb{R}^{N^2 \times N^2}$. If $\mathbf{f} \in \mathbb{R}^{N^2}$, the vector collecting all the wavelet coefficients is given by:

$$\mathbf{Wf} \in \mathbb{R}^{N^2}, \tag{6.5}$$

where it is clear that the matrix product $\mathbf{Wf}$ is the digital counterpart of (4.2).

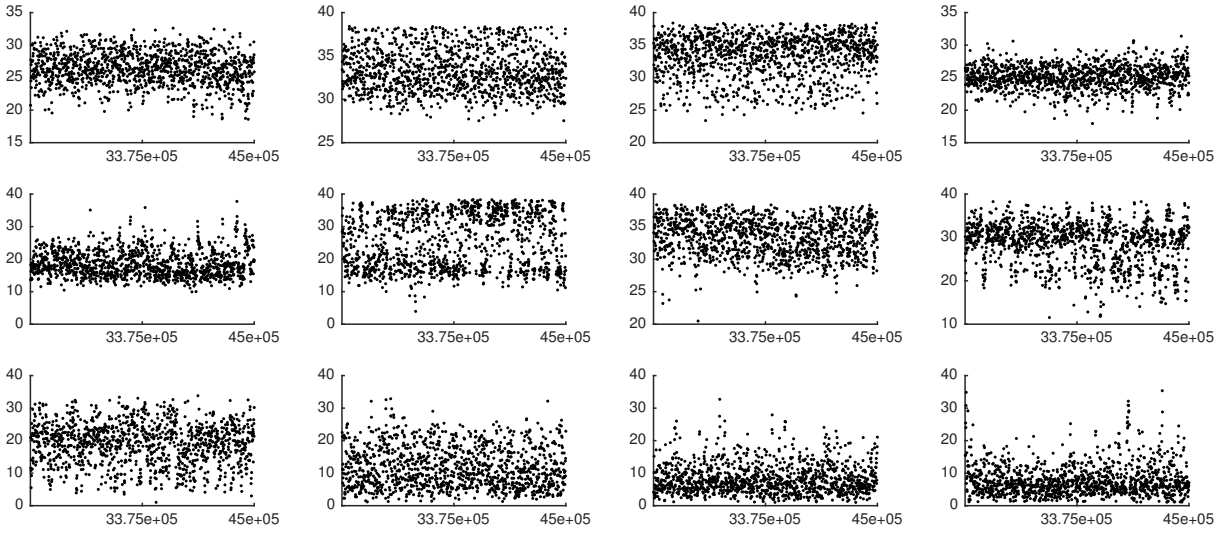


Figure 6.4: The Markov chain Monte Carlo chains of the radii.

In Article V, we implement Digital 3D Shearlet Transform **SH** generated by ShearLab3D [35, 34], available at `http://www.shearlab.org/`. When we consider $\mathbf{f} \in \mathbb{R}^{N^2 \times T}$, the shearlet coefficients are then given by:

$$\mathbf{SH}(\mathbf{f}) \in \mathbb{R}^{N^2 \times T \times K}, \tag{6.6}$$

where K is the number of 3D shearlets.

6.2.1 Computational results of adaptive methods for choosing regularization parameter

Review of Article IV Article IV presents 2D tomographic reconstruction of Shepp-Logan phantom and a middle slice of a walnut using the CWDS. We use the tomographic X-ray real data of the walnut, consisting of a 2D cross-section of the real 3D walnut measured with a custom-built CT device available at the University of Helsinki (Finland). The open-access datasets are available at `http://fips.fi/dataset.php`. For a detailed documentation of the acquiring setup, including the specifications of the X-ray systems, see [27]. We tested the sparse projections: 120 and 30 out of full 360° angles.

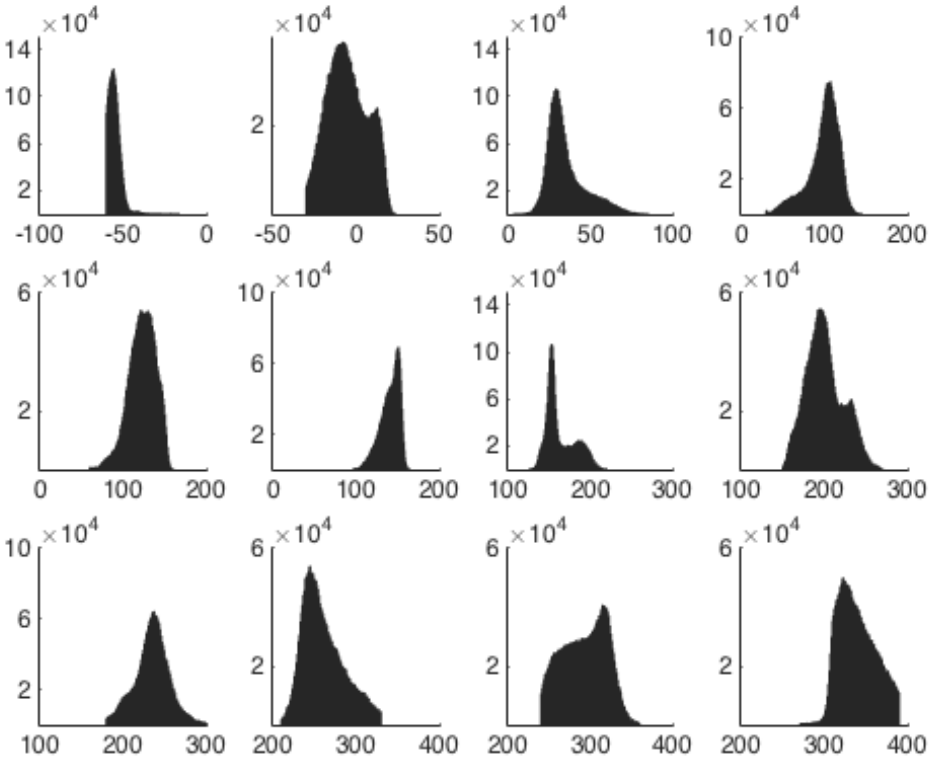


Figure 6.5: Histograms of the 1-d marginal posterior distribution of angles in the Markov chain Monte Carlo chain.

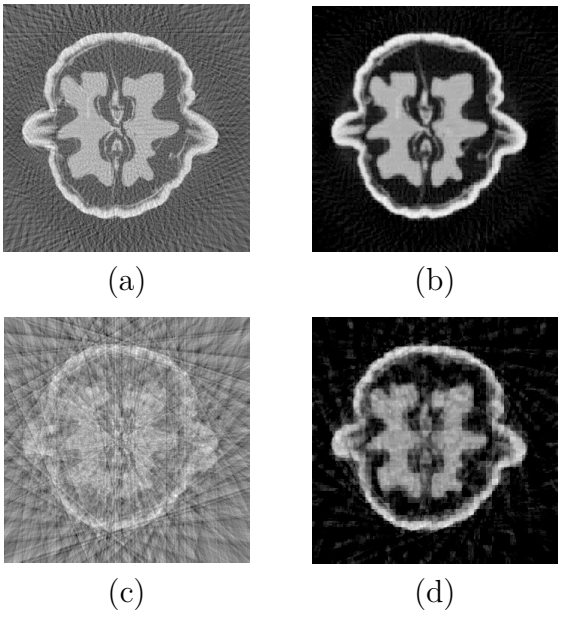
We assume that we have a prior object $\mathbf{f}_{pr}$ similar to the one we are imaging. Given $\kappa \geq 0$, for a vector $w \in \mathbb{R}^{N^2}$, we define the number of elements larger than κ in absolute value as follows:

$$\#_\kappa w := \#\{\, i \mid 1 \leq i \leq N^2,\ |w_i| > \kappa\},$$

with w_i is the i-th Haar wavelet coefficient. Now, the prior sparsity level is defined by

$$\mathcal{C}_{pr} = \frac{\#_\kappa\{\mathbf{W}\mathbf{f}_{pr}\}}{N^2},$$

where N^2 is the total number of coefficients. In practical computations the value of κ is set to be small. In Article IV, the sparsity level is computed from two different walnut photographs. In addition, the reconstruction of Shepp-Logan phantom (generated by MATLAB) is computed as well. The method is tested for sparse X-ray data and the results outperform the conventional FBP method via visual inspection. As it can be seen in Figure 6.6, the reconstructions using CWDS produce less streak artefacts. Figure 6.7 illustrates the ratio of nonzero wavelet coefficient for the walnut case converge to $\mathcal{C}_{pr}$ value relatively fast.



(a) (b)

(c) (d)

Figure 6.6: Reconstructions of the middle slice of walnut using FBP with (a) 120 projections, and (c) 30 projections. Reconstructions of the walnut using the wavelet-based method with (b) 120 projections, and (d) 30 projections.

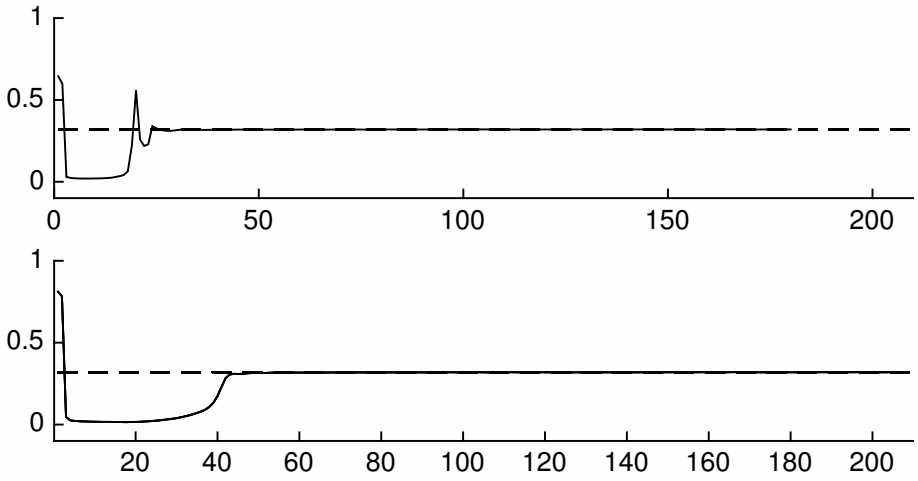


Figure 6.7: The ratio of nonzero wavelet coefficients as the iteration progresses for the walnut case. Top: 120 projections. Bottom: 30 projections. The dashed line shows the desired sparsity $\mathcal{C}_{pr}$.

Review of Article V In Article V, 3D human trabecular bone tomographic reconstruction using controlled shearlets domain sparsity is applied. X-ray data from two osteochondral samples were acquired. Two samples (diameter = 4 mm) were harvested from weight bearing area of tibial plateus from two cadavers under the approval of The Research Ethics Committee of the Northern Savo Hospital District, Kuopio, Finland (approval no 134/2015). The X-ray tomography data was acquired with SkyScan 1272 high-resolution μCT scanner (Bruker microCT, Kontich, Belgium). We collected 300 projection images acquired over a full 180 degree rotation with uniform angular step of 0.6 degrees between projections and used the Feldmann, David and Kress (FDK) reconstruction as baseline reconstructions. We picked 50 and 30 projection images from the measured data with uniform angular sampling and tested in our experiments.

These samples have full 300 projections from half-circle 180° of the healthy and osteoarthritis (OA) human trabecular bone. The shearlets coefficients of a baseline reconstruction, i.e. FDK method is computed [3]. We compute the nonlinear approximation as the best $\kappa-$term approximation of the baseline (FDK reconstruction from full 300 projection images) image where κ is ranging between 5% to 95%. Once it is done, morphometric parameters from the trabecular bone of each approximation images are computed:

1. Percentage of bone volume (BV/TV). BV refers to volume of the region segmented as bone and BV/TV refers to the ratio of the segmented bone volume to the total volume of the volume of interest (VOI);

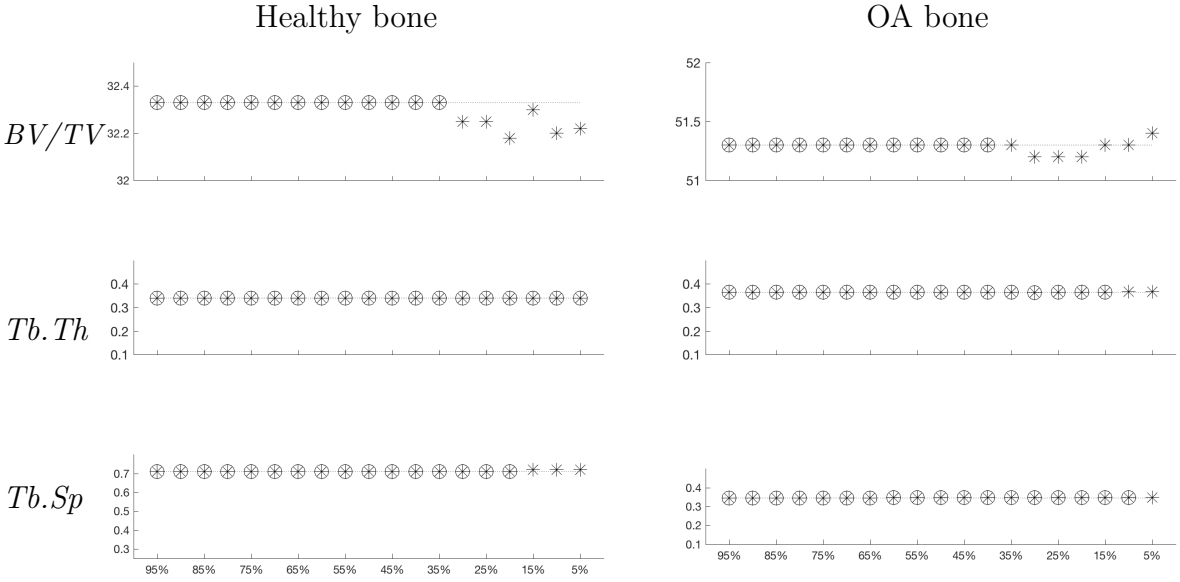


Figure 6.8: The value of bone parameters from inverse 3D shearlet transform of the healthy and OA bone using different sparsity levels. Stars with circles are the stable/unchanged values.

2. Trabecular thickness (*Tb.Th*): the diameter of the largest sphere which is entirely bounded within the solid surfaces (mm);

3. Trabecular separation (*Tb.Sp*): the thickness of the spaces as defined by binarization within the VOI (mm).

At a particular level, as κ decreases, the morphometric parameters start to deteriorate. The sparsity level $\mathcal{C}_{pr}$ is chosen at the stage before the trabeculae parameters start deteriorating as shown in Figure 6.8. In this case we can choose $\mathcal{C}_{pr}$ between $35\% - 40\%$.

We apply the FDK and CSDS methods to reconstruct the samples. When the number of projection images was reduced, the significant streak artefacts overwhelms the FDK reconstruction images while in the CSDS reconstructions the appearance of the streak artefacts is less. The non-negativity constraint and the enforcement of the penalty term ℓ_1-norm combined with the sparsity transform which acts as denoising process in the CSDS method give significant contribution to produce better reconstructions as it shown in Figure 6.9. Figure 6.10 presents the ratio of nonzero shearlets coefficients in each iteration. We stress that the proposed method CSDS offers an automatic method in choosing regularization parameter.

For these particular samples, we also compute the morphometrics parameter of

the trabecular bone of each reconstruction. Standard steps are applied to compute the parameters. Once we obtain the reconstruction images, we choose the inner area of the reconstruction images to avoid the artefacts caused by the dust due to the drilling process. Then, the images are segmented using CTAn software to produce binary images.

When the number of projection images is reduced, the quality of the binary images from the FDK reconstructions were relatively poor due to appearance of noise speckles in the binary images. It can be seen from Figure 6.11 that many of trabeculae become also broken. As a result, morphometric parameters of the trabecular bone for the FDK reconstructions contain large deviations of its baseline parameters as it is confirmed in Table 6.1 and Table 6.2. However, CSDS method performs better in the sense that there is no presence of noise speckles. Thus, the morphometric parameters values are closer to the baseline values. For instance Table 6.3 and 6.4 show that for both samples, the FDK reconstructions using 50 projections had differences in the $Tb.Sp$ parameter of $9.86\% - 17.1\%$, while for 30 projections it was $28.17\% - 40\%$. Other parameters such as $Tb.Th$ was affected significantly as well for the two different numbers of projections. The $Tb.Th$ decreased by up to 43.2% of its baseline values (from 0.37 mm to 0.21 mm or from 0.34 mm to 0.24 mm). The results from the CSDS algorithm show that the differences in the parameters are relatively smaller than those of the FDK method. For instance Tb.Sp increased by up to 14.43% difference of its baseline value for 50 projections and $5.71\% - 11.27\%$ for 30 projections. It is reported as well that $Tb.Th$ increased only up to 5.88%. The BV/TV increased by a relatively small amount: from CSDS method it increased up to 6.06%, not significant difference compared to the FDK method for which the deviation was up to 5.66%. Overall, the results show that the CSDS approach to recover the inner structure of the human trabecular bone samples from sparse projection images outperforms the conventional FDK approach.

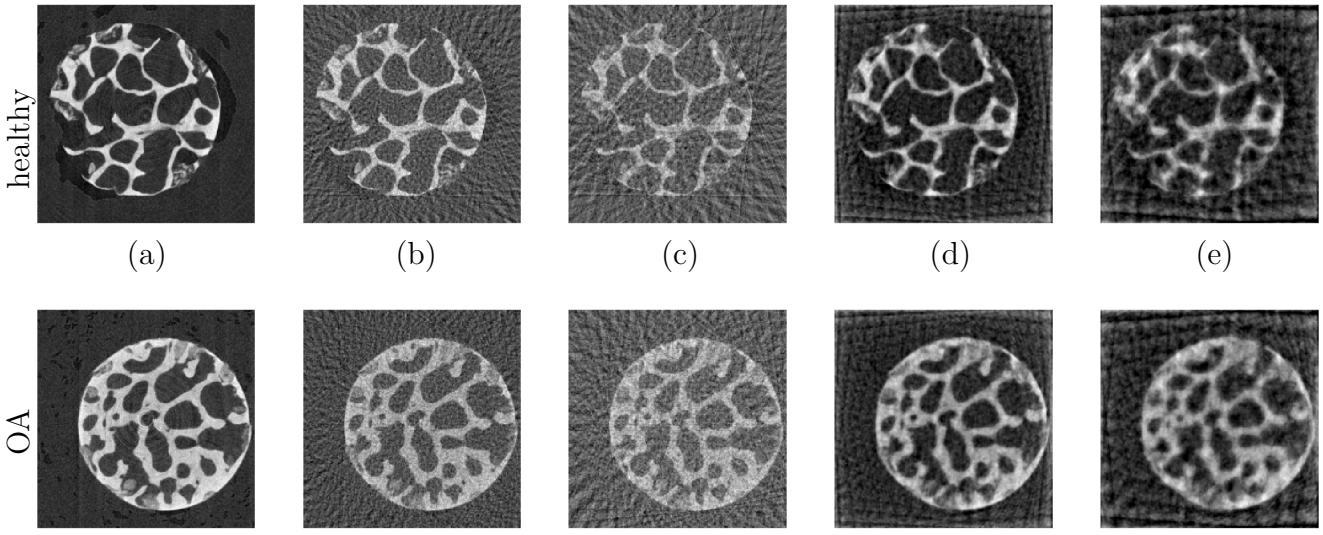
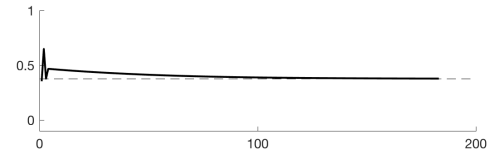
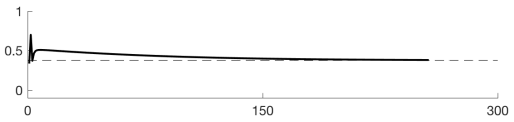


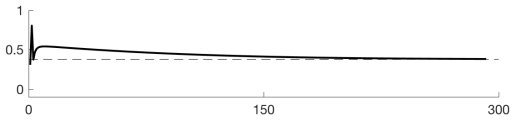
Figure 6.9: Axial micro-CT cross-section images of the 3D reconsructions. The baseline images (FDK reconstruction from 300 projections) are given in (a) and (f), FDK reconstructions from 50 projections are shown in (b) and (g), FDK reconstructions from 30 projections are shown in (c) and (h), CSDS reconstructions from 50 projections are shown in (d) and (i) and CSDS reconstructions from 30 projections are shown in (e) and (j).



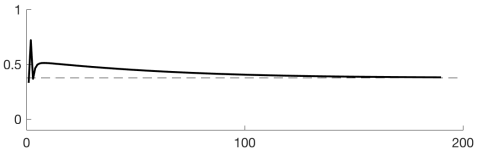
(a)



(b)



(c)



(d)

Figure 6.10: The ratio of nonzero shearlet coefficients as the iteration progresses from healthy sample using: (a) 30 projections and (b) 50 projections and from OA sample using: (c) 30 projections and (d) 50 projections.

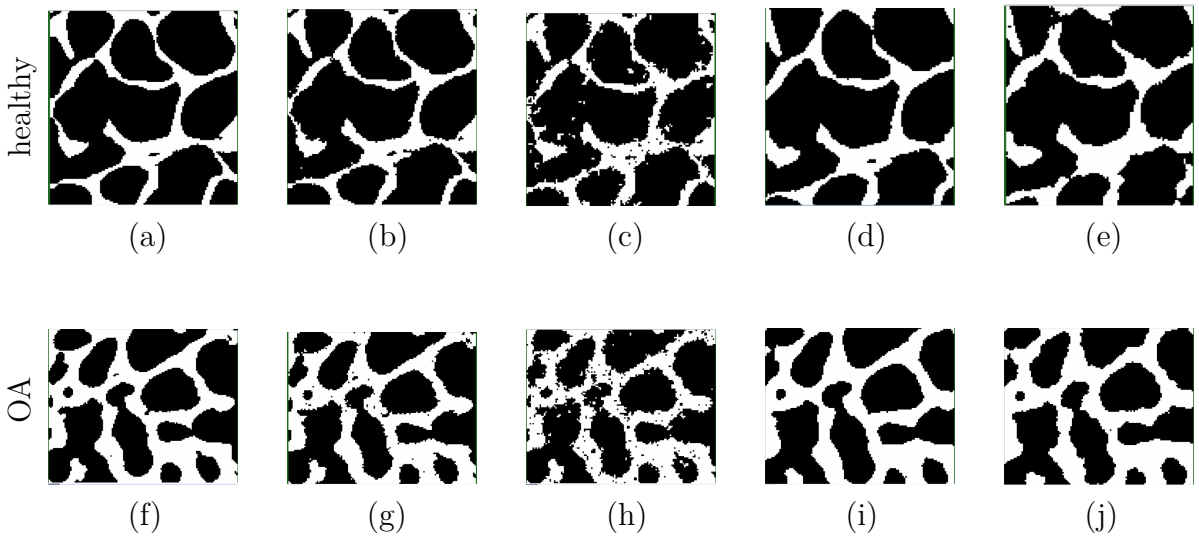


Figure 6.11: Axial cross-section images of thresholded images or binary images of the volume of interest which correspond to Figure 6.9.

Table 6.1: The trabecular bone morphometric parameters calculation for the healthy sample reconstruction from different number of projection images.

Method	Number of projections	BV/TV	$Tb.Th$ (mm)	$Tb.Sp$ (mm)
Baseline	300	32.33%	0.34	0.71
FDK	50	30.50%	0.28	0.64
	30	32.17%	0.24	0.51
CSDS	50	33.77%	0.36	0.70
	30	34.29%	0.33	0.63

Table 6.2: The morphometric parameters calculation for the OA sample reconstruction from different number of projection images.

Method	Number of projections	BV/TV	$Tb.Th$ (mm)	$Tb.Sp$ (mm)
Baseline	300	51.30%	0.37	0.35
FDK	50	50.60%	0.30	0.29
	30	48.57%	0.21	0.21
CSDS	50	53.69%	0.36	0.31
	30	52.79%	0.37	0.33

Table 6.3: The relative error of the morphometric parameters of the healthy trabecular bone from different number of projection images.

Method	Number of projections	BV/TV	$Tb.Th$	$Tb.Sp$
FDK	50	5.66%	17.65%	9.86%
	30	0.50%	29.41%	28.17%
CSDS	50	4.45%	5.88%	1.41%
	30	6.06%	2.94%	11.27%

Table 6.4: The relative error of the morphometric parameters of the OA trabecular bone from different number of projection images.

Method	Number of projections	BV/TV	$Tb.Th$	$Tb.Sp$
FDK	50	1.36%	18.92%	17.14%
	30	5.32%	43.24%	40.00%
CSDS	50	4.66%	2.70%	14.43%
	30	2.90%	0%	5.71%

7 Conclusion and future outlook

This thesis presents novel methods in X-ray computed tomography imaging from limited data:

(i) The NURBS-MCMC strategy to recover the boundary of a homogeneous object (an object that contains only one material) in terms of NURBS curve and its attenuation value.

(ii) Adaptive methods for choosing regularization parameter. The first approach so-called the controlled wavelet domain sparsity (CWDS) is based on enforcing sparsity in the two-dimensional wavelet transform domain, and the second so-called the controlled shearlet domain sparsity (CSDS) in the three-dimensional shearlet transform domain.

Novel algorithms above have been successfully implemented for real measured X-ray data and the results using under-sampled data outperform the baseline method.

Method (i) exploit MCMC strategy to solve the problem, and instead of a single output, the solution in the Bayesian framework is the posterior distribution. In addition, the results of the method (i) are conveniently in CAD-compatible format. In more detail, in this work, the knot vector of the NURBS curve is fixed, the future work is to consider it as random variables as well. The method has heavy computation, hence the future outlook can be to further speed-up the heavy computation such as optimizing the covariance of the sampling strategy. The choice of initial value could be improved by exploiting spline fitting to the baseline reconstruction.

The proposed method (ii) offers a strategy to automatically choosing regularization parameter where the end-users could avoid manually tuning the parameters. A known *a priori* sparsity level calculated from some available objects/samples is required. The future outlook is to study and to collect more objects that can be used to obtain the desired sparsity level. In addition, the strategy to choose the *a priori* sparsity level is problem-dependent. Also, the choice of the wavelet in CWDS method can vary depends on the applications.

The CSDS method can be implemented in dynamic tomography as well, for example in spatio-temporal CT imaging in $(2 + 1)$ dimensions or combining different regularization methods: 2D-shearlets for the spatial part and 1D-wavelet transform in time. Moreover, for the trabecular bone samples case, a statistical comparison is

one of the future studies. Another future research is to do *in vivo* experiments as well.

Finally, implementing parallelization strategy could save the computation time. Especially for the shearlet-based algorithm, computation of shearlet decomposition in a serialized manner can be another acceleration strategy.

Bibliography

[1] S. Akin and A. Kovscek, Computed tomography in petroleum engineering research, Geological Society, London, Special Publications, 215 (2003), pp. 23–38.

[2] K. Bredies, Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty, in Efficient algorithms for global optimization methods in computer vision, Springer, 2014, pp. 44–77.

[3] T. M. Buzug, Computed tomography: from photon statistics to modern cone-beam CT, Springer Science & Business Media, 2008.

[4] E. J. Candes, The restricted isometry property and its implications for compressed sensing, Comptes Rendus Mathematique, 346 (2008), pp. 589–592.

[5] L. Cartz, Nondestructive testing, (1995).

[6] A. Chambolle, An algorithm for total variation minimization and applications, Journal of Mathematical imaging and vision, 20 (2004), pp. 89–97.

[7] P. Chen, J. Huang, and X. Zhang, A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions, Fixed Point Theory and Applications, 2016 (2016), p. 54.

[8] S. Chib and E. Greenberg, Understanding the metropolis-hastings algorithm, The american statistician, 49 (1995), pp. 327–335.

[9] R. P. Christian and G. Casella, Monte carlo statistical methods, 2007.

[10] A. M. Cormack, Representation of a function by its line integrals, with some radiological applications, Journal of applied physics, 34 (1963), pp. 2722–2727.

[11] ——, Representation of a function by its line integrals, with some radiological applications. ii, Journal of Applied Physics, 35 (1964), pp. 2908–2913.

[12] N. R. Council et al., Mathematics and physics of emerging biomedical imaging, National Academies Press, 1996.

[13] I. Daubechies, Ten lectures on wavelets, SIAM, 1992.

[14] I. Daubechies, M. Defrise, and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on pure and applied mathematics, 57 (2004), pp. 1413–1457.

[15] L. De Chiffre, S. Carmignato, J.-P. Kruth, R. Schmitt, and A. Weckenmann, Industrial applications of computed tomography, CIRP Annals-Manufacturing Technology, 63 (2014), pp. 655–677.

[16] D. L. Donoho, Compressed sensing, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.

[17] Y. C. Eldar and G. Kutyniok, Compressed sensing: theory and applications, Cambridge University Press, 2012.

[18] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of inverse problems, vol. 375, Springer Science & Business Media, 1996.

[19] D. Fanelli and O. Öktem, Electron tomography: a short overview with an emphasis on the absorption potential model for the forward problem, Inverse Problems, 24 (2008), p. 013001.

[20] D. Gamerman and H. F. Lopes, Markov chain Monte Carlo: stochastic simulation for Bayesian inference, CRC Press, 2006.

[21] W. R. Gilks, S. Richardson, and D. Spiegelhalter, Markov chain Monte Carlo in practice, CRC press, 1995.

[22] A. Graps, An introduction to wavelets, IEEE computational science and engineering, 2 (1995), pp. 50–61.

[23] K. Guo, G. Kutyniok, and D. Labate, Sparse multidimensional representations using anisotropic dilation and shear operators, 2006.

[24] H. Haario, M. Laine, A. Mira, and E. Saksman, Dram: efficient adaptive mcmc, Statistics and computing, 16 (2006), pp. 339–354.

[25] H. Haario, E. Saksman, and J. Tamminen, Adaptive proposal distribution for random walk metropolis algorithm, Computational Statistics, 14 (1999), pp. 375–396.

[26] ——, An adaptive metropolis algorithm, Bernoulli, (2001), pp. 223–242.

[27] K. Hämäläinen, L. Harhanen, A. Kallonen, A. Kujanpää, E. Niemi, and S. Siltanen, Tomographic x-ray data of a walnut, arXiv preprint arXiv:1502.04064, (2015).

[28] D. Hofmann, P. Sembdner, S. Holtzhausen, C. Schöne, and R. Stelzer, Parametric design in reverse engineering based on ct data, The e-Journal of Nondestructive Testing, 21 (2016).

[29] J. Kaipio and E. Somersalo, Statistical and computational inverse problems, vol. 160, Springer Science & Business Media, 2006.

[30] A. C. Kak and M. Slaney, Principles of computerized tomographic imaging, SIAM, 2001.

[31] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo, Statistical inversion for medical x-ray tomography with few radiographs: Ii. application to dental radiology, Physics in Medicine and Biology, 48 (2003), p. 1465.

[32] P. Kuchment, The Radon transform and medical imaging, SIAM, 2013.

[33] G. Kutyniok and D. Labate, Shearlets: Multiscale analysis for multivariate data, Springer Science & Business Media, 2012.

[34] G. Kutyniok, W.-Q. Lim, and R. Reisenhofer, Shearlab 3d: Faithful digital shearlet transforms based on compactly supported shearlets, ACM Transactions on Mathematical Software (TOMS), 42 (2016), p. 5.

[35] G. Kutyniok, M. Shahram, and X. Zhuang, Shearlab: A rational design of a digital parabolic scaling algorithm, SIAM Journal on Imaging Sciences, 5 (2012), pp. 1291–1332.

[36] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, Sparse multidimensional representation using shearlets, Optics & Photonics, 2005 (2005), pp. 59140U–59140U.

[37] I. Loris and C. Verhoeven, On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty, Inverse Problems, 27 (2011), p. 125007.

[38] S. Mallat, A wavelet tour of signal processing, Academic press, 1999.

[39] A. Mira et al., On metropolis-hastings algorithms with delayed rejection, Metron, 59 (2001), pp. 231–241.

[40] J. L. Mueller and S. Siltanen, Linear and nonlinear inverse problems with practical applications, SIAM, 2012.

[41] F. Natterer, The mathematics of computerized tomography, SIAM, 2001.

[42] L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore, et al., Digital tomosynthesis in breast imaging., Radiology, 205 (1997), pp. 399–406.

[43] L. Piegl and W. Tiller, The nurbs book. 1997, Monographs in Visual Communication, (1997).

[44] M. Rantala, S. Vanska, S. Jarvenpaa, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen, Wavelet-based reconstruction for limited-angle x-ray tomography, IEEE transactions on medical imaging, 25 (2006), pp. 210–217.

[45] G. O. Roberts, A. Gelman, W. R. Gilks, et al., Weak convergence and optimal scaling of random walk metropolis algorithms, The annals of applied probability, 7 (1997), pp. 110–120.

[46] D. Rogers, An introduction to nurbs: With historical perspective. morgan kaufmann, San Francisco, CA, (2001).

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

[48] M. Sarfraz, Computer-aided reverse engineering using simulated evolution on nurbs, Virtual and Physical Prototyping, 1 (2006), pp. 243–257.

[49] L. A. Shepp and J. Kruskal, Computerized tomography: the new medical x-ray technology, American Mathematical Monthly, (1978), pp. 420–439.

[50] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. 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 (2003), p. 1437.

[51] L. Tierney, A note on metropolis-hastings kernels for general state spaces, Annals of applied probability, (1998), pp. 1–9.

[52] L. Tierney and A. Mira, Some adaptive monte carlo methods for bayesian inference, Statistics in medicine, 18 (1999), pp. 2507–2515.

[53] C. R. Vogel, Computational methods for inverse problems, SIAM, 2002.

[54] E. Wilmer, D. A. Levin, and Y. Peres, Markov chains and mixing times, American Mathematical Soc., Providence, (2009).

[55] T. Wu, R. H. Moore, E. A. Rafferty, and D. B. Kopans, A comparison of reconstruction algorithms for breast tomosynthesis, Medical physics, 31 (2004), pp. 2636–2647.

[56] Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M. M. Goodsitt, L. M. Hadjiiski, J. Ge, and C. Zhou, A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis, Medical physics, 33 (2006), pp. 3781–3795.