University of Helsinki
Dissertationes Universitatis Helsingiensis
18/2025

Variational Regularization Methods and Open Data for Image Reconstruction in Computed Tomography

Alexander Meaney

Doctoral dissertation

To be presented for public discussion with the permission of the Faculty of Science of the University of Helsinki, in Auditorium A111, Exactum building, on the 15th of January, 2025 at 12 o'clock.

Department of Mathematics and Statistics
Faculty of Science
University of Helsinki

HELSINKI 2025

Advisors:

Professor Samuli Siltanen
Department of Mathematics and Statistics
University of Helsinki
Finland

Professor Miika Nieminen
Research Unit of Health Sciences and Technology
University of Oulu
Finland

Pre-examiners:

Research Professor Emil Sidky
Department of Radiology
The University of Chicago
United States

Professor Marta M. Betcke
Department of Computer Science
University College London
United Kingdom

Opponent:

Associate Professor Silvia Gazzola
Department of Mathematics
University of Pisa
Italy

Custos:

Professor Samuli Siltanen
Department of Mathematics and Statistics
University of Helsinki
Finland

Publisher: Helsingin yliopisto
Series: Dissertationes Universitatis Helsingiensis 18/2025

ISBN 978-952-84-0510-8 (print)
ISBN 978-952-84-0509-2 (online)
ISSN 2954-2898 (print)
ISSN 2954-2952 (online)
PunaMusta, Joensuu 2025

Abstract

Inverse problems are concerned with recovering some property of interest of a system from indirect measurement data. Such problems are abundant in mathematics and in natural and applied sciences. Typically, inverse problems are ill-posed, meaning that a unique solution may not exist, or that the solution is highly unstable with respect to noise and modelling error. Ill-posedness can be overcome by using regularization, whereby the original problem is replaced with a similar but stable, well-posed problem.

Variational regularization is a regularization strategy in which the solution of the inverse problem is defined as the minimizer of a loss function, consisting of a data term, which ensures that the solution sufficiently matches the data, and a regularization term, which includes prior knowledge about the model properties and which stabilizes the solution. Such problems can be solved computationally by using tools from mathematical optimization.

One very important application of inverse problems is medical imaging, where the interior of the body is imaged using exterior measurements. Computed tomography (CT) is a medical imaging modality in which cross-sectional images are produced using X-ray measurements taken from different directions. CT imaging also has many applications in biological research, non-destructive testing, and security screening.

The cross-sectional CT images are created by applying a computational process known as a reconstruction algorithm. The standard reconstruction algorithm in CT is filtered backprojection (FBP), which is fast, well understood, and has excellent performance in many settings. FBP fails, however, when the X-ray data is undersampled or acquired using very low amounts of radiation. In such cases, superior CT images can be obtained using reconstruction algorithms based on variational regularization. Also, new reconstruction algorithms may be necessitated by the emergence of new hardware and system developments, such as multi-energy CT.

The design and performance evaluation of CT reconstruction algorithms often begins with using simulated data, which is straightforward to generate. However, no simulation can fully model all the myriad physical and measurement phenomena that occur in a CT scan, and real data can sometimes produce unexpected results. It is thus essential that reconstruction algorithms are developed and tested using real measured X-ray data so as to fully gauge their performance and properties. Unfortunately, algorithm research is often hindered by the limited availability of suitable real CT data.

The purpose of this thesis is twofold. The first is to present new CT reconstruction algorithms based on variational regularization for use in various imaging settings. The second is to document open datasets of real X-ray data that have been made available to the reconstruction algorithm development community for use in their work. The thesis consists of a summarising report, three published scientific articles, and one article manuscript.

The summarising report discusses inverse problems theory, regularization, mathemat-

ical optimization, and computed tomography in the context of the research conducted. It also gives an overview of the open CT datasets that were measured and published during this work.

In the first article, multiple different reconstruction algorithms were developed and compared for low dose, sparse angle multi-energy CT in a 2D setting. The algorithms used an approach known as joint reconstruction, where all the data from the different energy channels are combined into a single reconstruction problem, enforcing structural similarity across the channels to improve image quality.

The second article describes a reconstruction approach to multi-energy CT that combines the data from all the energy channels as side information in order to improve the individual channel reconstructions by applying directional total variation regularization. This article also utilized a sparse angle 2D setting.

The third article describes the Helsinki Tomography Challenge 2022, which was organized to promote reconstruction algorithm development for limited angle tomography, which is a particularly challenging inverse problem. The article presents the challenge setup, the performance of the submitted algorithms, and the dataset that was used to benchmark the algorithms, which has now been released as an open dataset.

The fourth article describes a new approach to total variation regularization in 3D cone beam computed tomography (CBCT) reconstruction. The regularization strength is treated as a dynamic variable, driving the reconstruction towards a pre-determined gradient magnitude sparsity.

Tiivistelmä

Inversio-ongelmassa kiinnostuksen kohteena on jokin suure, jota ei voi suoraan määrittää, mutta jonka vaikutukset ovat havaittavissa epäsuorien mittausten avulla. Inversio-ongelmille on tyypillistä, että ne ovat huonosti asetettuja, jolloin yksikäsitteistä ratkaisua ei ole olemassa, tai ratkaisu on hyvin epävakaa kohinan ja mallinnusvirheiden suhteen. Tällainen ongelma voidaan ratkaista korvaamalla alkuperäinen ongelma samankaltaisella mutta hyvin asetetulla, vakaasti käyttäytyvällä ongelmalla, johon on sisällytetty etukäteistietoa ratkaisusta. Tätä kutsutaan regularisoinniksi.

Tietokonetomografia on kuvantamistekniikka, jossa kohteesta muodostetaan poikkileikkauskuvia käyttämällä eri suunnista tehtyjä röntgenmittauksia sekä rekonstruktioalgoritmiksi kutsuttua laskennallista menetelmää. Standardialgoritmi FBP suoriutuu huonosti kun röntgendata on puutteellista tai erittäin kohinaista. Tällöin voidaan käyttää inversiomatematiikkaan ja regularisointiin perustuvia algoritmeja saavuttamaan parempi kuvanlaatu. Myös uudentyyppiset kuvantamisratkaisut hyötyvät kehittyneemmistä algoritmeista, tärkeänä esimerkikkinä monienergiakuvantaminen, joka antaa uusia työkaluja eri kudosten ja materiaalien analysointiin. Algoritmisuunnittelussa ja -testauksessa on tärkeää käyttää mitattua dataa, koska simuloidulla datalla ei voida kattavasti mallintaa kaikkia mittaustilanteen moninaisia fysikaalisia prosesseja ja virhelähteitä.

Tässä tutkielmassa on kehitetty uusia rekonstruktioalgoritmeja tietokonetomografian eri sovelluksiin, sekä julkaistu ja dokumentoitu avoimia datasettejä algoritmitestausta varten. Ensimmäiset kaksi osatyötä keskittyvät matalan annostason monienergiakuvantamiseen. Kolmas osatyö esittelee julkista tutkimushaastetta, jonka tavoitteena oli kehittää uusia algoritmeja rajatun kulman tomografiaan, jossa röntgendata puuttuu kokonaan tietyistä suunnista. Osatyö toimii myös haasteen myötä julkaistun avoimen datasetin dokumentaationa. Neljäs osatyö keskittyy kolmiulotteiseen kuvantamiseen, jossa laajasti käytettyyn kokonaisvaihtelu-regularisointiin on kehitetty uusi tulokulma. Lisäksi tutkielmassa on yhteenveto, jossa esitellään inversiomatematiikan teoriaa ja laskennallisia menetelmiä, sekä tutkimuksen aikana mitattuja ja julkaistuja avoimia datasettejä. Kokonaisuutena työ asettuu osaksi rekonstruktioalgoritmien suunnittelun aktiivista tutkimuskenttää.

Acknowledgements

Writing a PhD thesis has been compared to a long and winding path filled with obstacles, pitfalls, and unexpected twists and detours. For me, this has certainly been the case. I started this thesis as an experimental physicist walking confidently into the world of applied mathematics, with the intention of designing brilliant new reconstruction algorithms for medical imaging. Science has a way of humbling the unprepared, and I soon found myself dazed and confused in a world of function spaces, operators, ill-posed problems, and any number of other confusing subjects. Little by little, and with no small amount of sweat, I have managed to build a modicum of understanding on these topics. Hopefully, some new science has also been created along the way. I have not walked this path alone, and now it is time to offer thanks.

First and foremost, I want to thank my doctoral advisor, Professor Samuli Siltanen. You brought me into the fascinating world of mathematics and provided me with time and space to develop the mathematical skillset I needed to become an algorithm researcher. Along the way I even dabbled a bit in topology and algebra, but always returning to what is best in life: MATLAB programming. We have shared many memorable experiences, from causing laboratory havoc with liquid nitrogen to travelling across the globe to conferences in places like Rio de Janeiro and Tokyo. Thank you for being a mentor and a friend.

I want to express my gratitude and appreciation to my co-advisor Professor Miika Nieminen for his guidance and encouragement during my PhD path. I still remember fondly our first meeting at a doctoral dissertation's late night afterparty, and I am happy to have ended up collaborating with you. I have greatly enjoyed our many discussions, which have ranged from medical imaging research to guitar gear to fatherhood, and I hope they will continue in the coming years.

I am deeply grateful to my pre-examiners, Research Professor Emil Sidky and Professor Marta M. Betcke, for their thorough work in reviewing my thesis, and for their insightful and helpful feedback. I am equally grateful to Associate Professor Silvia Gazzola for acting as the honoured Opponent in my PhD defense.

Research is a team effort. This thesis would not exist without my research collaborators and co-authors Jussi Toivanen, Professor Ville Kolehmainen, Evelyn Cueva, Matthias J. Ehrhardt, Fernando Silva de Moura, Markus Juvonen, and Mikael Brix. I am deeply grateful to all of you. To my friend Markus I also wish to say that it's been a joy sharing an office with you throughout our concurrent, occasionally intersecting, PhD paths. I could not have asked for a better office mate.

I want to thank Chief Medical Physicist Mika Kortesniemi and Professor Pekka Pankka for their membership in my thesis committee. Mika, you have been a source of encouragement and provided valuable insight into the latest developments in medical imaging, both during my days of working in medical physics and during my PhD research. Pekka, you have patiently helped a physicist trying to learn to swim in the waters of mathematics.

The Inverse Problems Research Group at the Department of Mathematics and Statis-

tics at the University of Helsinki has always fostered a friendly atmosphere of mutual encouragement and support. For this, I want to thank my colleagues both past and present, in no particular order: Professor Matti Lassas, Professor Lauri Oksanen, Tatiana Bubba, Siiri Rautio, Lauri Ylinen, Anna Suomenrinne-Nordvik, Elli Karvonen, Ensio Suonperä, Heli Virtanen, Tommi Heikkilä, Jonatan Lehtonen, Jussi Korpela, Jalo Nousiainen, Andreas Hauptmann, Petri Ola, Zenith Purisha, Antti Kujanpää, Salla-Maaria Latva-Äijö, Hewan Zewde, Luca Ratti, Martin Ludvigsen, Neil Dizon, Teemu Saksala, Tapio Helin, Hanne Kekkonen, and Petteri Piiroinen. Extra special thanks to Topias Rusanen for his invaluable MATLAB codes for X-ray imaging. A big thank you also to Mareike Paul from the Environmental Geochemistry group for friendship and solidarity during our concurrent PhD studies.

My first scientific home was the X-ray laboratory at the Department of Physics at the University of Helsinki, where I ended up working with X-ray microtomography almost by accident, starting my path into the world of imaging science. I want to thank Professor Keijo Hämäläinen for giving me that opportunity, and Aki Kallonen, Jussi-Petteri Suuronen, and Liisa Porra for guiding me during those early days. My experimental work in building an X-ray scanner for inverse problems research would never have got off the ground without the patient help of Senior Engineer Merja Blomberg. Thank you Merja. Many thanks also to Laboratory Engineer Heikki Suhonen for collaborating in imaging projects in the X-ray Micro-CT Laboratory, and to Professor Simo Huotari for many interesting and insightful scientific conversations.

During the course of my PhD research, I have been priviliged to collaborate with many skilled researchers in Professor Nieminen's research group at the University of Oulu: Juuso Ketola, Satu Inkinen, Ella Räinä, Dinidu Jayakody Arachchilage, Mustafa Al-Rubaye, and Samuel Onnela. It has been a pleasure working with you.

I am grateful to Professor Sauli Savolainen and the medical physics community at Helsinki University Central Hospital for the valuable experience I gained working with clinical medical imaging prior to my doctoral research.

I am indebted to the Academy of Finland's Centre of Excellence of Inverse Modelling and Imaging, and to Business Finland (funding decisions 6614/31/2016 and 8132/31/2022) for providing the financial support which made this work possible. I also want to thank The Doctoral Programme in Mathematics and Statistics (Domast) for providing travel funding.

According to credible sources, life exists also outside academia. To all of my friends: thank you. You know who you are. I am especially grateful to Jutta Rahmel for her support during the early years of my PhD research, and during the first difficult months of COVID in 2020.

I would never have reached this point without the endless support of my parents Eija Liisa Sokka-Meaney and Kenneth Meaney, and my brother Aleksis. I cannot thank you enough.

Last but definitely not least, I want to thank my wife Outi for her boundless love and support, and our son Robin for filling my heart with joy every single day. I love you both, you are my world.

Helsinki, December 2024 *Alexander Meaney*

The thesis consists of this overview and the following articles.

List of Publications

I J. Toivanen, A. Meaney, V. Kolehmainen, and S. Siltanen. *Joint reconstruction in low dose multi-energy CT.* Inverse Problems and Imaging, 14(4):607-629, 2020.

II E. Cueva, A. Meaney, S. Siltanen, and M. J. Ehrhardt. *Synergistic multi-spectral CT reconstruction with directional total variation.* Philosophical Transactions of the Royal Society A, 379:20200198, 2021.

III A. Meaney, F. S. de Moura, M. Juvonen, and S. Siltanen. *Helsinki tomography challenge 2022: Description of the competition and dataset.* Applied Mathematics for Modern Challenges, 1(2):170-201, 2023.

IV A. Meaney, M. A. K. Brix, M. T. Nieminen, S. Siltanen. *Image reconstruction in cone beam computed tomography using controlled gradient sparsity.* Manuscript. arXiv:2412.07465.

Author's Contributions

I Generation and preprocessing of simulated data, measurement and preprocessing of experimental data, generation of computational forward models entirely by the author.

Writing has been done in equal shares with the other authors.

II Generation and preprocessing of simulated data, measurement and preprocessing of experimental data entirely by the author. Collaboration with E. Cueva in generation of computational forward models.

The author contributed to writing the article.

III Conceptual development of the challenge and dataset together with the other authors. Measurement and preprocessing of experimental data, generation of reference data and example codes entirely by the author.

Writing has been done in equal shares with the other authors.

IV Theoretical method development, algorithm development, generation and preprocessing of simulated data, reconstruction computations by the author.

Most of the writing has been done by the author.

Nomenclature

Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
AI	Artificial Intelligence
ATV	Anisotropic Total Variation
CBCT	Cone-Beam Computed Tomography
CG	Conjugate Gradient method
CGS	Controlled Gradient Sparsity
CT	Computed Tomography or Computerized Tomography
DL	Deep Learning
dTV	Directional Total Variation
GD	Gradient Descent method
HU	Hounsfield Unit
IQ	Image Quality
IR	Iterative Reconstruction
JTV	Joint Total Variation
MBIR	Model-Based Iterative Reconstruction
ML	Machine Learning
NN	Non-Negativity constraint
PDFP	Primal-Dual Fixed Point algorithm
SD	Steepest Descent algorithm
SNR	Signal-to-Noise Ratio
STV	Smoothed Total Variation
SVD	Singular Value Decomposition
TSVD	Truncated Singular Value Decomposition
TV	Total Variation

Mathematical Notation

$\mathcal{A}$	Forward operator in the continuum model $m = \mathcal{A}\mathcal{F} + \epsilon$.
A	Forward operator in the discretized model $m = Af + \varepsilon$.
$B(x, r)$	Open ball in metric space (X, d), centered at x with radius $r > 0$. Defined as $B(x, r) = \{y \in X \mid d(y, x) < r\}$.
$\overline{B}(x, r)$	Closed ball in metric space (X, d), centered at x with radius $r > 0$. Defined as $\overline{B}(x, r) = \{y \in X \mid d(y, x) \leq r\}$.
$\mathcal{C}_f$	Sparsity level of $f \in \mathbb{R}^N$. Defined as $\mathcal{C}_f = (\#_\kappa f)/N$.
D	Discrete gradient matrix.
$\mathcal{F}$	Quantity of interest in the continuum model $m = \mathcal{A}\mathcal{F} + \epsilon$.
f	Discretization $f \in \mathbb{R}^N$ of the continuum quantity $\mathcal{F}$.
f^*	Solution of a variationally regularized inverse problem.
(i, j)	Pixel coordinates in $N_1 \times N_2$ image representation of discretized 2D data $f \in \mathbb{R}^N$, such that $N = N_1 N_2$.
(i, j, k)	Voxel coordinates in $N_1 \times N_2 \times N_3$ volume representation of discretized 3D data $f \in \mathbb{R}^N$, such that $N = N_1 N_2 N_3$.
$L(f)$	Objective function $L : \Omega \mapsto \mathbb{R}$, $\Omega \subset \mathbb{R}^N$. Also known as a loss function or cost function.
l	Element index $l \in \{1, \ldots, N\}$ in vectorized representation of the discretization $f \in \mathbb{R}^N$.
M	Number of data points in the measurement data vector $m \in \mathbb{R}^M$.
m	A vector $m \in \mathbb{R}^M$ containing the measurement data in a given inverse problem.
N	Number of elements in the discretized model $f \in \mathbb{R}^N$.
n	Dimensionality of the underlying quantity of interest $\mathcal{F} : \Omega \mapsto \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, in the continuum model $\mathcal{A}\mathcal{F} = m + \epsilon$. That is to say, $n = \dim \Omega$.
$\|\|x\|\|_p$	p-norm or ℓ^p-norm of vector $x \in \mathbb{R}^N$. Defined as $\|\|x\|\|_p = \left(\sum_{l=1}^{N} \|x_l\|^p \right)^{1/p}$.
Q	Number of channels in multispectral data.
q	Channel index number $q \in \{1, \ldots, Q\}$ in multispectral data.
$R(f)$	Regularization term in the cost function $L : \Omega \mapsto \mathbb{R}$, $\Omega \subset \mathbb{R}^N$, used in variational regularization.
S	1. Descent direction in gradient-based optimization. 2. Soft thresholding operator $S_\tau(f) = \operatorname{sgn}(f) \cdot \max\{\|f\| - \tau, 0\}$, $\tau \geq 0$.
α	Regularization parameter in variational regularization.
β	1. Smoothing parameter in smoothed total variation regularization. 2. CG update parameter in the nonlinear conjugate gradient method.
γ	Step length in optimization algorithm.

ϵ	Modeling error and noise in the continuum model $m = \mathcal{A}\mathcal{F} + \epsilon$.
ε	1. Modeling error and noise in the discretized model $m = Af + \varepsilon$. 2. Tolerance for stopping criterion in numerical optimization.
κ	1. Condition number of a matrix. 2. Tolerance around zero in the counting function $\#_\kappa$.
ν	Iteration index $\nu \in \mathbb{N}$ in iterative algorithms.
Ω	Subset of Euclidean space. $\Omega \subset \mathbb{R}^n$ in the continuum model. $\Omega \subset \mathbb{R}^N$ in the discretized model. Ω may or may not be a proper subset.
$\#_\kappa$	The counting function $\#_\kappa f = \#\{l \mid 1 \leq l \leq N,\ \lvert f_l \rvert < \kappa\}$, where $f \in \mathbb{R}^N$.

Contents

1 Introduction **1**

2 Inverse Problems **5**
2.1 The Mathematical Model of Inverse Problems 5
2.2 Regularization 8
2.3 Sparsity 9
2.4 Inverse Crime 11
2.5 Images and Volumes in Computational Inverse Problems 11
2.5.1 Image Data 12
2.5.2 Volume Data 13
2.5.3 Multispectral Data 13

3 Total Variation Regularization **17**
3.1 The Total Variation Functional 17
3.2 Numerical Implementation 17
3.3 Variants 20
3.3.1 Smoothed Total Variation 21
3.3.2 Anisotropic Total Variation 21
3.3.3 Joint Total Variation 22
3.3.4 Directional Total Variation 22
3.4 Discrete Gradient Matrices 23
3.4.1 Discrete Gradient Matrix Operator Norm 25

4 Mathematical Optimization **27**
4.1 Mathematical Optimization Problems 27
4.2 Basic Concepts 27
4.3 Convexity 28
4.4 Gradient-Based Algorithms 31
4.4.1 Gradient Descent and Steepest Descent 32
4.4.2 Conjugate Gradient Methods 33
4.5 Proximal Algorithms 36
4.5.1 Primal-Dual Fixed Point Algorithm 37

5 Image Reconstruction in Computed Tomography **41**
5.1 Projections and X-ray Measurements 41
5.1.1 The Radon Transform 41
5.1.2 X-ray Projections 42
5.1.3 Multispectral X-ray Imaging 44
5.2 Analytical Reconstruction 45

5.2.1 Filtered Backprojection Reconstruction . . . 45
5.2.2 Reconstruction in Fan-beam Geometry . . . 46
5.2.3 Cone Beam Reconstruction . . . 47
5.3 Iterative Reconstruction . . . 48
5.3.1 Discretized CT Imaging Model . . . 48
5.3.2 Model-Based Iterative Reconstruction . . . 49
5.4 Image Quality, Noise, and Representation . . . 51
5.5 Proximal Algorithms for CBCT Reconstruction . . . 52
5.5.1 ATV-NN-PDFP . . . 53
5.5.2 TV-NN-PDFP . . . 53
5.5.3 Experimental Results for ATV-NN-PDFP and TV-NN-PDFP . . . 54
5.5.4 TV-CGS . . . 58

6 Open Software and Data for CT Reconstruction 67
6.1 Reconstruction Algorithms in the Real World . . . 67
6.2 Reconstruction Tools . . . 67
6.3 Open CT Projection Datasets . . . 68

7 Summary of Results 71
7.1 Publication I . . . 71
7.2 Publication II . . . 72
7.3 Publication III . . . 72
7.4 Publication IV . . . 73

References 75

1 Introduction

The mathematical science of *inverse problems* is focused on solving indirect problems. A direct problem or forward problem follows the natural causal direction: *given the cause, determine the effects*. Conversely, an inverse problem works in the other direction: *given the effects, determine the cause*. Typically the effects are measurement data, and the cause is the model for some underlying quantity of interest. Inverse problems are abundant in mathematics, physics, and engineering. Some examples of practical inverse problems include determining the structure of an object from its X-ray projections, recovering a sharp image from a blurred image, and determining the electrical conductivity properties of an object based on surface electode measurements.

Research on inverse problems is usually focused on problems that are *ill-posed*. This means that they are, in some sense, difficult to solve. There is an inherent asymmetry in ill-posed inverse problems: if one knows the solution, it is usually easy to calculate or simulate the measurement data, but if one only has the data, obtaining the solution is hard. One way of interpreting ill-posedness is that the *forward model* from the model space to the data space is typically stable, well-defined, and continuous, whereas the *inverse model* from the data space to the model space is highly unstable and sensitive to modelling error and noise, poorly defined, and/or not continuous. The relationship between the model and the data in direct problems and inverse problems is visualized in Figure 1.1.

A family of techniques known as *regularization methods* is key to obtaining useful and meaningful solutions to ill-posed inverse problems. In regularization, the original ill-posed problem is replaced with a similar but well-posed problem that should produce an approximate solution close to the true solution of the original problem. Usually there is some prior knowledge available on the properties of the solution. In such cases *variational regularization* is a particularly useful approach, as it provides a flexible and powerful framework for designing and incorporating strong priors to the regularized problem.

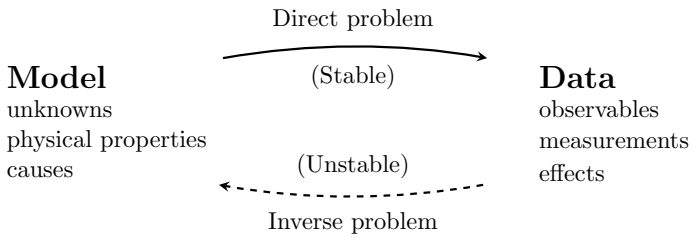


Fig. 1.1. The model-to-data relationship in direct problems and inverse problems. The solid line indicates the forward model and the dashed line indicates the inverse model.

Inverse problems feature prominently in medical imaging, where the interior of the human body is imaged using indirect means for diagnosis or treatment. There exist multiple

modalities of medical imaging, each suitable for studying different structural or functional properties of the body. Each modality uses some form of energy capable of penetrating tissues, leading to its own mathematical model of the imaging process. For instance, radiography and computed tomography (CT) use X-rays, magnetic resonance imaging (MRI) uses strong magnetic fields and radiofrequency electromagnetic waves, and nuclear medicine (NM) imaging uses gamma rays emitted by the decay of radioactive isotopes. Most modalities require using a computerized *reconstruction algorithm* to produce the desired image from the indirect measurements. When viewing medical imaging through the lense of inverse problems, the terms reconstruction and solution can be used more or less synonymously.

The modality of interest in this work is computed tomography. The purpose of CT imaging is to reconstruct cross-sectional images of the patient from X-ray projection data taken from different directions. The mathematics underlying CT reconstruction were first developed by the Austrian mathematician Johann Radon in 1917 [1, 2]. Despite this significant achievement, the importance of Radon's work went unnoticed for almost half a century [3]. The mathematics of CT were further advanced by the South African physicist Allan MacLeod Cormack in two landmark papers published in the early 1960s [4, 5]. Concurrently, and unaware of the works of Radon and Cormack, the English electrical engineer Godfrey Hounsfield, working for Electric and Musical Industries (EMI), developed the first CT scanner in the early 1970s, and collaborated with radiologist James Ambrose to publish the first clinical results [6, 7]. In recognition of their role in the development of clinical CT scanning, Cormack and Hounsfield were awarded the 1979 Nobel Prize in Physiology or Medicine [8].

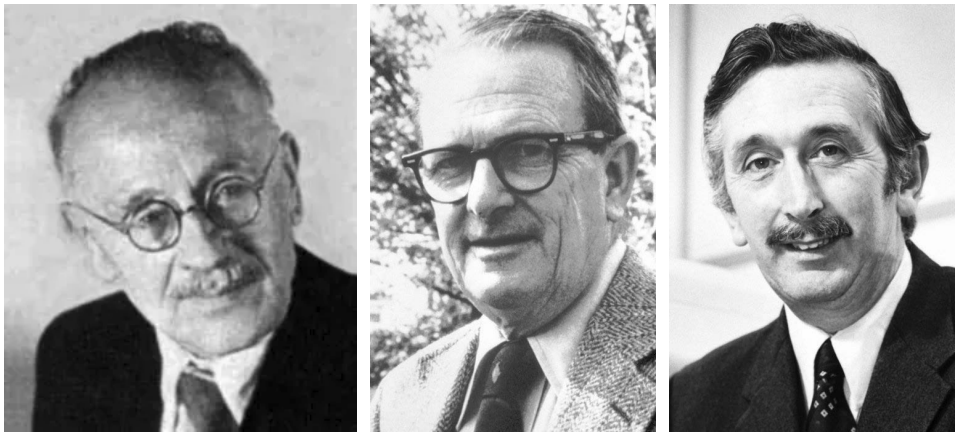
The significance of computed tomography in the practice of medicine is difficult to overstate, and CT scanners have become an indispensable tool in diagnostics [9–11]. Throughout half a century, CT imaging has continued to evolve, driven forward by new hardware developments and algorithmic innovations, often in interplay with each other [12–14]. New imaging geometries and hardware have necessitated development of new algorithms, and conversely, algorithmic inventions have enabled previously unapplicable hardware solutions to be taken into use. Multiple complementary CT technologies have been developed, including helical (spiral) CT [15, 16], cone beam computed tomography (CBCT) [17, 18], microtomography [19], synchrotron radiation CT [20, 21], and energy-sensitive photon-counting detector CT (PCD-CT) [22–25], with broad applications in medicine and beyond [26–29]. Nevertheless, these technologies are all variations on a theme, and the underlying mathematics and physics remain the same.

During recent years, model-based iterative reconstruction (MBIR) has become increasingly popular in CT image reconstruction [30, 31]. MBIR can be thought of as applying inverse problems methodology to CT reconstruction, and it has significant potential in improving image quality and reducing radiation dose [32]. Even more recently, CT reconstruction methods based on machine learning (ML) and artificial intelligence (AI) have also gained significant traction [33–36]. MBIR methods and ML/AI methods have significant overlap, and in both methodologies broad availability of open source software and open data plays a central role in algorithm development and testing [37–39].

This work focuses on development of 1) new MBIR methods based on variational regularization, and 2) open data for image reconstruction in CT imaging. Publications I and II deal with iterative reconstruction algorithms for spectral (multi-energy) CT in a 2D setting, Publication III deals with open data and open software for algorithm development, specifically in the context of engaging researchers in an algorithm development challenge for limited-angle CT reconstruction, and Publication IV introduces a new type of sparsity-

based approach to 3D image reconstruction in CBCT. Publications I-III involve using real, measured X-ray data.

This overview is organized as follows. Chapter 2 discusses the general mathematical framework of inverse problems and regularization, with an added emphasis on computational representation of imaging data. Chapter 3 discusses total variation regularization, a widely applied regularization strategy in inverse problems. Chapter 4 discusses mathematical optimization, which is a key tool in practical implementation of many reconstruction algorithms. Chapter 5 discusses image reconstruction specifically in the context of computed tomography, and gives a brief comparison of traditional analytical methods and newer inverse problem techniques. Chapter 6 discusses open data and open software for CT reconstruction algorithm development. Throughout chapters 2-6, the methodologies are discussed in the context of how they relate to the research in the publications that form the bulk of this work. Finally, Chapter 7 gives a concise summary of the results of Publications I-IV.



(a) Johann Radon **(b)** Allan M. Cormack **(c)** Godfrey N. Hounsfield

Fig. 1.2. Pioneers of CT imaging. (Image credits: (a): Johann Radon Institute for Computational and Applied Mathematics; (b) and (c): The Nobel Foundation.)

2 Inverse Problems

2.1 The Mathematical Model of Inverse Problems

The mathematical framework of this work is focused on problems of the form

$$m = \mathcal{A}\mathcal{F} + \epsilon, \tag{2.1}$$

where $m \in \mathbb{R}^M$ is measurement data given by a measurement device, $\mathcal{F} : \Omega \mapsto \mathbb{R}$ is a function modelling some physical quantity of interest defined on the subset $\Omega \subset \mathbb{R}^n$, $\mathcal{A} : \Omega \mapsto \mathbb{R}^M$ is an operator which models the measurement process, and $\epsilon \in \mathbb{R}^M$ is a perturbation term representing measurement noise and modelling error. In this work, it is assumed that $\mathcal{A}$ can be modelled as, or at least approximated as being linear. We are interested in recovering $\mathcal{F}$, but only have access to it via indirect measurements, which is why we label Eq. (2.1) as an *inverse problem*: we wish to determine the cause from the effect, manifested in the measurement data. Typically Eq. (2.1) is difficult to solve in the sense of being an *ill-posed problem*.

In order to discuss ill-posed problems, it is first necessary to define what a *well-posed problem* is. This was done in 1902 by the French Mathematician Jacques Hadamard in the *Hadamard conditions* [40]. A well-posed problem has the following properties:

1. The solution exists.
2. The solution is unique.
3. The solution depends continuously on the data.

If a problem is not well-posed, it is ill-posed. An English overview of the Hadamard conditions can be found in [41].

This work is primarily concerned with numerically solving computational inverse problems, for which it is necessary to discretize the continuum model in Eq. (2.1) as

$$m = Af + \varepsilon, \tag{2.2}$$

where $A \in \mathbb{R}^{M\times N}$ is the discretized linear measurement model, $f \in \mathbb{R}^N$ is a discretization of the unknown continuum quantity $\mathcal{F}$ we wish to determine, and $\varepsilon \in \mathbb{R}^M$ models measurement noise as well as modelling and discretization error. The number of measurement data points M and the resolution (if applicable) of m are determined by the dimensions and properties of the measurement apparatus and by the data collection procedure. When discretizing $\mathcal{F}$, it may be possible in some cases to choose the size N and the resolution (if applicable) independently of m. The size and structure of A are determined by the measurement setup, the mathematical model for the underlying physical problem, and by the discretization f. A is usually called the *forward model* or the *system matrix*.

In general, A cannot be assumed to be a square matrix. Even if this is the case, attempting the naive reconstruction

$$f \approx A^{-1}m \tag{2.3}$$

will yield unsatisfactory results in ill-posed problems [42, 43]. Because Eq. (2.2) expresses the inverse problem as set of linear equations, it is useful to reflect on what failure to meet the Hadamard conditions means when viewed through the lens of linear algebra.

Failure to meet Hadamard condition 1 means that m does not belong to the column space of A. In such cases, a best estimate of the solution can be obtained using the least squares approximation

$$f^*_{\text{lsq}} = \underset{f\in\mathbb{R}^N}{\arg\min}\, ||Af - m||_2^2. \tag{2.4}$$

Finding the minimizer of the least squares problem is equivalent to solving the *normal equations* [44]

$$A^TAf = A^Tm. \tag{2.5}$$

For most practical inverse problems A will be so large that direct computation of $(A^TA)^{-1}$ is infeasible, and an iterative method needs to be used to minimize Eq. (2.4).

Failure to meet Hadamard condition 2 means that A has a non-trivial null space, so that A^TA will be singular (non-invertible) and the normal equations have an infinite number of solutions. In such cases, the best approximate solution is the *minimum length least squares* solution f^*_{mllsq}, *i.e.* the solution to the normal equations with the minimum length measured in the ℓ^2 norm. The unique minimum length least squares solution to Eq. (2.5) is given by

$$f^*_{\text{mllsq}} = A^+m, \tag{2.6}$$

where A^+ is the *pseudoinverse* or *Moore-Penrose inverse* of A, computed using the *singular value decomposition* (SVD) of A [44].

The singular values $\sigma_1, \ldots, \sigma_N$ of a matrix $A \in \mathbb{R}^{M\times N}$ are the square roots of the eigenvalues of A^TA, arranged in descending order so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$. Now, let A have the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_S > 0$ and $\sigma_{S+1} = \sigma_{S+2} = \ldots = \sigma_N = 0$, where $0 \leq S \leq N$. We define the $S \times S$ diagonal matrix

$$D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_S \end{bmatrix} \tag{2.7}$$

and the $M \times N$ matrix

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.8}$$

where each 0 indicates a zero matrix of necessary size. Now, for *every* matrix $A \in \mathbb{R}^{M\times N}$ there exists the factorization

$$A = U\Sigma V^T, \tag{2.9}$$

where $U \in \mathbb{R}^{M\times M}$ and $V \in \mathbb{R}^{N\times N}$ are both orthogonal matrices. The pseudoinverse is given by

$$A^+ = V\Sigma^+U^T, \tag{2.10}$$

where

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \tag{2.11}$$

It is important to emphasize that the pseudoinverse exists even for singular matrices.

Failure to meet Hadamard condition 3 means that the matrix A is *ill-conditioned*. This is, in some sense, the most challenging obstacle to obtaining a useful solution to Eq. (2.2), and it is also a characteristic feature of ill-posed inverse problems. A matrix is ill-conditioned if small changes in the entries of A produce large changes to the solutions of $Af = m$. In a well-conditioned matrix, small changes in A lead to only small changes in the solution. The *condition number* κ of a matrix is given by

$$\kappa(A) = ||A||\ ||A^{-1}||. \tag{2.12}$$

For the ℓ^2 matrix norm, this can be equivalently stated as

$$\kappa_2(A) = \frac{\sigma_1}{\sigma_N}, \tag{2.13}$$

where σ_1 is the largest singular value of A and σ_N is the smallest singular value of A. The higher the condition number is, the worse ill-conditioned A is. In the case where $\sigma_N = 0$, or when A is non-invertible, the condition number is defined to be $\kappa(A) = \infty$. Alternatively, the condition number of a non-invertible or non-square matrix is sometimes defined as the ratio between the largest and smallest non-zero singular values [45].

Computers represent real numbers using *floating-point arithmetic*. Each real number is stored using a finite number of bits which depends on the computer system. This means that there will invariably be *round-off error* when using computers to make numerical computations. Computing matrix factorizations such as SVD will therefore be only approximations of the exact ones. If A has a high condition number, this can lead to dramatic errors when using the SVD to solve the problem. It should be noted that in most practical inverse problems the forward operator A is so large that computing the SVD becomes infeasible, and an iterative method must be used to solve the problem. However, even if no factorization is used to compute a solution, in practical applications measurement device inaccuracies will invariably lead to some modelling error in A, and using an iterative method to solve Eq. (2.4) will result in accumulation of errors during iteration, which can dramatically affect the outcome in ill-conditioned systems.

Ill-conditioning also makes solving Eq. (2.2) highly susceptible to perturbations from noise in m. Using the pseudoinverse to solve the system results in multiplication of vectors with the inverses of the singular values. When the last non-zero singular values σ_S, σ_{S-1}, σ_{S-2}, etc. are very small, their inverses become exceedingly large, leading to severe amplification of round-off errors and noise. In iterative least-squares solvers the presence of noise and modelling error in m will eventually also begin to severely degrade the results.

Finally, in situations where the discretization of f is chosen so that A is an (in principle) invertible $M \times M$ matrix, it can be difficult to identify ill-posedness in Hadamard condition 3 for the general case of the inverse problem in question. In such cases, ill-posedness in the sense of Hadamard condition 3 occurs when [42]

$$\lim_{M\to\infty} \kappa(A) = \infty. \tag{2.14}$$

In practical settings, this type of scenario usually occurs when there are few measurement points in m and a low-dimensional discretization of f leads to A having relatively few entries.

It is evident that failure of Hadamard condition 3 can lead to highly challenging problems, making the solution exceedingly susceptible to modelling error and noise. In such settings, the solution obtained can be so degraded by numerical unstability and noise as to be useless. These challenges can be tackled by employing *regularization*.

2.2 Regularization

Regularization is an essential tool when searching for meaningful and useful solutions to ill-posed inverse problems [45, 46]. It serves two purposes. Firstly, regularization is used to counter the ill-posedness arising from failure to meet the Hadamard conditions, and especially to stabilize the high sensitivity to amplification of noise and modelling error arising from ill-conditioning in the sense of Hadamard condition 3. Secondly, regularization can be used to include *prior information* about the solution into the reconstruction problem. Based on the physics of the underlying model, one often has some kind of knowledge on the general properties of the solution, relating for example to the solution norm, smoothness, structure, sparsity, or proximity to some already known model. Incorporating prior information into the solution method is especially important when working with limited or undersampled measurement data, which is often the case in applied inverse problems.

In regularization, the ill-posed problem of attempting to invert Eq. (2.2) is replaced with a nearby well-posed problem. This regularized problem should be designed in such a way that it produces a useful approximation f^* of the true solution f. There are several different regularization methods available for working with computational inverse problems.

One common approach to regularization is using *truncated singular value decomposition* (TSVD) [47–52]. In TSVD, the $K < S$ largest nonzero singular values are retained in the pseudoinverse, and the rest are set to zero. We define the $S \times S$ matrices

$$D_K = \mathrm{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_K, 0, \ldots, 0,) \tag{2.15}$$

and

$$D_K^+ = \mathrm{diag}\left(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_K^{-1}, 0, \ldots, 0,\right), \tag{2.16}$$

and the $M \times N$ matrix

$$\Sigma_K^+ = \begin{bmatrix} D_K^+ & 0 \\ 0 & 0 \end{bmatrix}. \tag{2.17}$$

The TSVD-regularized pseudoinverse A_K^+ is obtained by replacing Σ^+ in Eq. (2.10) with Σ_K^+. It is possible to compute the TSVD solution without first determining the computationally expensive full SVD [52]. The central question in TSVD is choosing K so that a suitable balance is found between retaining important features in the reconstruction while sufficiently suppressing noise and oscillations. For very large-scale inversion TSVD becomes challenging, and iterative methods are usually preferred.

Many ill-posed inverse problems exhibit a behaviour known as *semi-convergence*: iterative methods for solving the least-squares problem in Eq. (2.4) initially approach a useful approximate reconstruction before the iterates start becoming excessively degraded by noise and oscillations [43, 53–55]. Sufficiently early interruption of the iterations therefore acts as a form of regularization, and the solution is determined by both the iterative method used and the *stopping criterion* for interrupting the iterations. This approach is also known as *iterative regularization*, and it was utilized in Publication II.

In *variational regularization* we define the *loss function* or *cost function* $L : \Omega \mapsto \mathbb{R}$,

$$L(f) = \frac{1}{2}||Af - m||_2^2 + \alpha R(f), \tag{2.18}$$

where $\Omega \subset \mathbb{R}^N$, $\frac{1}{2}||Af - m||_2^2$ is the *data fidelity term*, $R(f)$ is the *regularization term*, also known as the *regularizer* or *penalty*, and $\alpha \in \mathbb{R}_+$ is the *regularization parameter* that

controls the strength of the regularization [46, 56, 57]. The solution f^* to the regularized inverse problem is defined as the minimizer of the cost function:

$$f^* = \arg\min_{f \in \Omega} \left\{ \frac{1}{2} ||Af - m||_2^2 + \alpha R(f) \right\}. \tag{2.19}$$

The regularization term $R(f)$ is used to stabilize the solution with regards to the data m, as it penalizes arbitrarily large fluctuations in f. The regularization term can also be used to be used to include *a priori* information on the solution. Because finding f^* as defined in Eq. (2.19) usually requires a computational optimization algorithm to minimize the cost function, such methods are sometimes called *optimization-based reconstruction methods* in the image reconstruction literature. The expression *model-based iterative reconstruction* (MBIR) is also used in the context of medical imaging, as the minimization problem includes both the imaging model A and prior knowledge modelling of f^* via the regularizer $R(f)$, and optimization algorithms are based on iterative processes. Variational regularization is the main approach to solving inverse problems throughout this work, and it was used in Publications I, II, and IV.

The choice of the regularizer $R(f)$ will have a strong impact on f^*. This is often guided by knowledge on the underlying physics of the problem. The choice of the regularization parameter α is similarly important. Numerous methods for choosing α have been developed, but no catch-all solution exists, and a good choice of the regularization parameter is highly context-dependent on the forward model A, on the noise and modelling error ε in the data m, and on the regularizer $R(f)$ [56, 58–71]. It needs to be noted that K in the TSVD method and the number of iterations taken in iterative regularization should also be interpreted as regularization parameters in their own contexts [72].

A widely used approach in ill-posed inverse problems is to use the loss function

$$L(f) = \frac{1}{2} ||Af - m||_2^2 + \alpha ||Bf||_p^p, \tag{2.20}$$

where B is a linear transformation used to emphasize some *a priori* chosen properties of f. When $p = 2$, Eq. (2.20) is known as *Tikhonov regularization* [42, 43, 46]. When $p = 1$, the regularizer is called a *sparsity inducing prior* [45, 46].

2.3 Sparsity

A vector or matrix is said to be *sparse* when most of its elements are zero-valued. When $p = 1$ in Eq. (2.20), the use of this loss function in variational regularization promotes solutions where the vector Bf^* is sparse. A qualitative understanding of the cause of this phenomenon can be gained by studying the properties of the norms involved. The contour lines of the ℓ^1 norm have a diamond-like shape, centered around the origin, whereas the contour lines of the ℓ^2 norm are shaped like circles or balls centered around the origin. The minimizer f^* will be found at some intersection point of the data fidelity term $\frac{1}{2}||Af - m||_2^2$ and the regularizer $\alpha||Bf||_p^p$. When $p = 1$, this intersection point is more likely to occur at one of the tips or along an edge of the ℓ^1 diamond, so that some of the components of Bf^* are zero. The higher the value of α, the stronger is this sparsity-inducing effect of ℓ^1 regularization, with more elements of Bf^* pushed to zero.

As an illustrative case of the mechanism of sparsity-inducing regularization, consider the very simple, low-dimensional toy problem $L : \mathbb{R}^2 \mapsto \mathbb{R}$ where $B = I$, $\alpha = 11$ and

$$A = \begin{bmatrix} -3 & 1 \\ -1 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} -4 \\ 8 \end{bmatrix}.$$

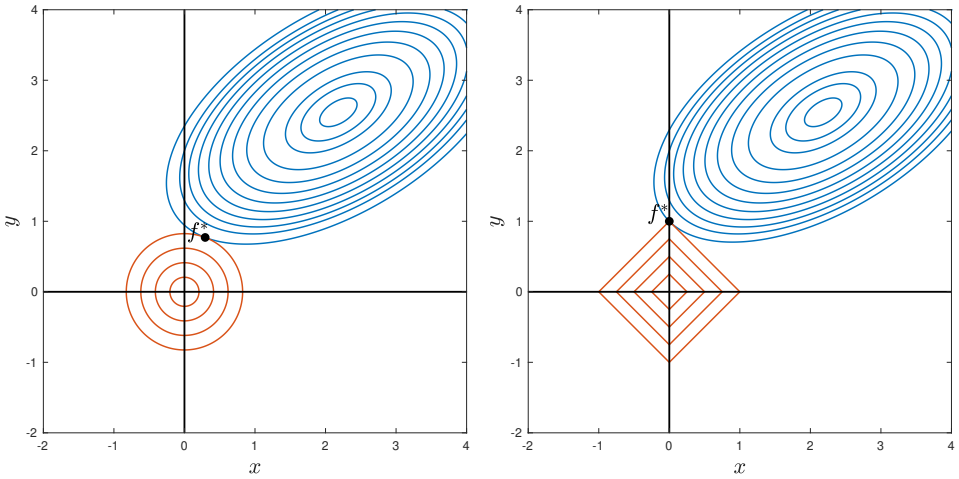


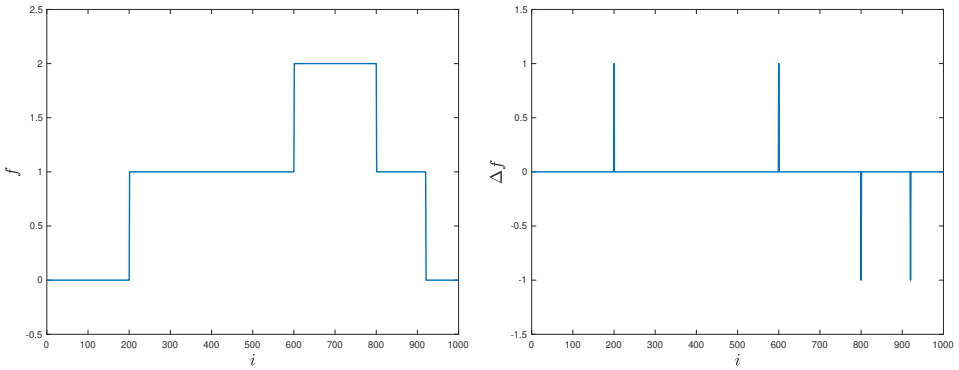
Fig. 2.1. Contour lines for the data fidelity term $\frac{1}{2}||Af-m||_2^2$ (blue) and the regularization term $\alpha||f||_p^p$ (red) when **(a)** $p = 2$, and **(b)** $p = 1$ for a regularized variation problem $L : \mathbb{R}^2 \mapsto \mathbb{R}$. The marker labeled f^* indicates the unique minimizer of $L(f) = \frac{1}{2}||Af - m||_2^2 + \alpha||f||_p^p$ in each case.

Figure 2.1 shows the contour lines for the data fidelity term and the regularizer for both $p = 2$ and $p = 1$. For the sparsity-inducing regularizer that uses the ℓ^1 norm, the minimizer f^* of Eq. (2.20) occurs at a location where one of the coordinates is zero, and the solution has 50% sparsity. When the value of the regularization parameter is reduced, f^* moves away from the sparse solution, but increasing α maintains sparsity in f^*.

Why should one be interested in sparse solutions? The motivation is that in many practical cases, the essential features of the solution to the inverse problem can be represented using relatively few nonzero coefficients, either in f^* or - as is more often the case - after some change of basis is applied using the linear transformation B. Hence, using sparsity-based regularization allows one to enforce very strong priors in the regularized solution. It is therefore essential to find suitable priors which accurately reflect the properties of the true solution. Common approaches include sparsity in the derivative or gradient of f^* (discussed further in Chapter 3), or in a wavelet basis, a Fourier basis, or a trigonometric basis, to name but a few [73]. Of course, even with a suitable sparsity-based regularizer, the choice of an appropriate regularization parameter α remains an open question, but the solutions are at least heading in the right direction and allow experimenting with different values of α. Publications I, II, and IV of this work use ℓ^1 regularization, and Publication IV is especially focused on quantifying sparsity and utilizing it in the image reconstruction process.

The potentially remarkable effectiveness of using a change of basis to promote sparsity can be straightforwardly demonstrated via another toy example. Assume that the signal in Figure 2.2 (a) represents the true solution f in some unspecified inverse problem. The signal contains 1000 data points, out of which 720 are nonzero. Applying the difference operator

$$\Delta f_i = \begin{cases} f_{i+1} - f_i, & i \in \{1, \ldots, N-1\} \\ 0, & i = N, \end{cases} \tag{2.21}$$



(a) **(b)**

Fig. 2.2. **(a)** a 1D signal and **(b)** its numerical derivative, illustrating how a change of basis can promote sparsity.

which can be interpreted as a numerical derivative of sorts, yields the signal in Figure 2.2 (b). This signal has only four nonzero coefficients, allowing an extremely sparse representation of the original signal.

Sparsity-based regularization is closely related to *compressed sensing*, a signal processing technique which allows accurate recovery of a signal from fewer samples than what is required by conventional signal processing theory [74–76].

2.4 Inverse Crime

In the development of numerical methods for solving computational inverse problems, a natural starting point is to computationally create a synthetic true solution f, then create the *simulated data* m (possibly with some simulated noise added afterwards), and proceed with implementing and testing suitable solution methods. Here lurks a great pitfall in inverse problems method development: the committing of *inverse crime* [77–79]. Inverse crime means that the same discretization and forward model A are used for both generating the simulated data and for the solution method, resulting in a situation where the existence of modelling error and discretization error are neglected.

Committing inverse crime will lead to a problem that is less ill-posed than the true underlying problem, and it will result in unrealistically good numerical solutions. For this reason, the development and validation of robust and reliable solution methods for inverse problems require that the model used to generate the simulated data and the model used to solve the inverse problem are independent of each other.

2.5 Images and Volumes in Computational Inverse Problems

Inverse problems often deal with two-dimensional (2D) or three-dimensional (3D) functions. In order to work with such problems computationally, they must be appropriately discretized into 2D data, which can be thought of as images, or 3D data, which can be thought of as volumes. Because numerical inverse problems are most often solved using the tools of linear algebra and matrix computations, the discretizations need to be further

flattened into one-dimensional vectors. The discretized data needs to be meticulously labeled in both the standard 2D or 3D form and the vectorized form so that shifting between the different representations can be done in a consistent and unambiguous manner.

2.5.1 Image Data

Let $\Omega \subset \mathbb{R}^2$. The continuous physical or mathematical model underlying the problem can be represented as the function $\mathcal{F} : \Omega \mapsto \mathbb{R}$. The model could represent, for example, the contrast in a black-and-white film photograph, the distribution of X-ray attenuation coefficients in a single cross-sectional plane of an object, or the X-ray power flux density incident on a flat-panel radiation detector.

The 2D discretization of $\mathcal{F}$ is the matrix $f \in \mathbb{R}^{N_1 \times N_2}$, whose elements are called *pixels* (picture elements). Each element is labeled $f_{i,j}$ with $i \in \{1, \ldots, N_1\}$ to indicate the row and $j \in \{1, \ldots, N_2\}$ to indicate the column, consistent with standard matrix notation. Additionally, each element is also given a unique index $l \in \{1, \ldots, N\}$, where $N = N_1 N_2$, so that the 2D image can alternatively be expressed as the column vector $f \in \mathbb{R}^N$. The index l grows first from top to bottom along rows, then from left to right along columns. When the 2D data is flattened, the columns of f are simply stacked downwards into one column vector, from left to right. An example of the vector element indexing and the flattening process for 2D data is shown in Figure 2.3. In this particular discretization, for instance, the element $f_{1,3}$ in the image representation corresponds to the element f_7 in the vector representation.

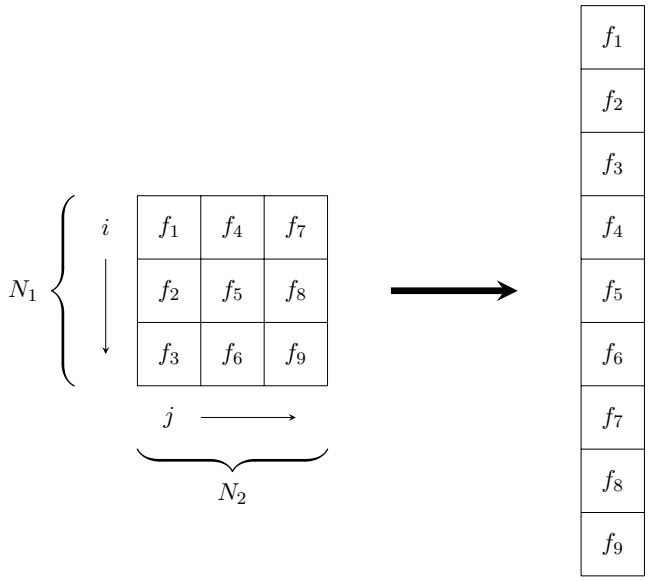


Fig. 2.3. Example of image data indexing and flattening in a discretized 2D problem with $N_1 = N_2 = 3$.

The elements of f can be referred to either as $f_{i,j}$ when using the 2D coordinates, or as f_l when using the vector indexing. The difference between these two is implicitly clear from the number of indices used. It needs to be emphasized that the image coordinate system differs from the standard Cartesian coordinate system, and to avoid confusion it is

important keep track of which system is being used when studying the 2D data in different software applications.

2.5.2 Volume Data

Let $\Omega \subset \mathbb{R}^3$ and let $\mathcal{F} : \Omega \mapsto \mathbb{R}$ represent the underlying continuum model. This model could represent, for example, the full three-dimensional distribution of X-ray attenuation coefficients in an object.

The 3D discretization of $\mathcal{F}$ is the 3D array $f \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, whose elements are called *voxels* (volume elements). Each element is labeled $f_{i,j,k}$, where $i \in \{1, \ldots, N_1\}$, $j \in \{1, \ldots, N_2\}$, and $k \in \{1, \ldots, N_3\}$ indicate the row, the column, and the z-plane, respectively. Each element is once again also given a unique index $l \in \{1, \ldots, N\}$, with $N = N_1 N_2 N_3$, allowing the vector representation $f \in \mathbb{R}^N$. The index l grows identically to the 2D case in each individual imaging plane, and then along the z-direction. When the 3D data is flattened, each plane is flattened as in the 2D case, and the resulting column vectors are stacked downwards, proceeding in the z-direction. An example of the vector element indexing and the flattening process for 3D data is shown in Figure 2.4. In this particular discretization, for instance, the element $f_{1,3,2}$ in the volume representation corresponds to the element f_{16} in the vector representation.

Similarly to the 2D case, the elements of f can be referred to either as $f_{i,j,k}$ when using the 3D coordinates, or as f_l when using the vector indexing, with the difference implicitly clear from the context. The volume coordinate system is also significantly different from the Cartesian coordinate system, which is illustrated in Figure 2.5.

2.5.3 Multispectral Data

In multispectral data, information on the same object is available in multiple different energies, frequencies, or wavelengths. Multispectral imaging data is usually highly spatially correlated along the different channels. An archetypical example of multispectral data is the RGB representation of colour images, where the image is split into red, green, and blue colour channels. In the context of CT imaging, multispectral data is produced when more than one X-ray energy spectra are used in the X-ray projection measurements [80–84]. In the literature this is usually referred to as spectral CT, multispectral CT, dual-energy CT, or multienergy CT. Publications I and II of this work deal with image reconstruction in spectral CT imaging.

In the mathematical model for multispectral data, the measurements are conducted using Q different channels, each with one with its own discretized measurement model

$$m_q = A_q f_q + \varepsilon_q, \tag{2.22}$$

where $q \in \{1, \ldots, Q\}$ is the channel index. In the imaging sciences, f_q in each individual channel is usually 2D or 3D data, as detailed in Sections 2.5.1 - 2.5.2. The data from the different energy channels are usually not flattened together into a single vector. Rather, because the data does represent different properties the same underlying object, the different energy channels are contained together in some kind of data structure. There is no universal standardized way of referring to the individual elements in multispectral data, but one approach is to use $f_{i,j,q}$ for image data, $f_{i,j,k,q}$ for volume data, and $f_{l,q}$ for vectorized data. An example of this particular indexing, with flattening, of multispectral 2D is shown in Figure 2.6. The different channels are stored together in a data structure.

In this particular discretization, for instance, the element $f_{1,3,2}$ in the image representation corresponds to the element $f_{7,2}$ in the vector representation of channel 2 (the green channel).

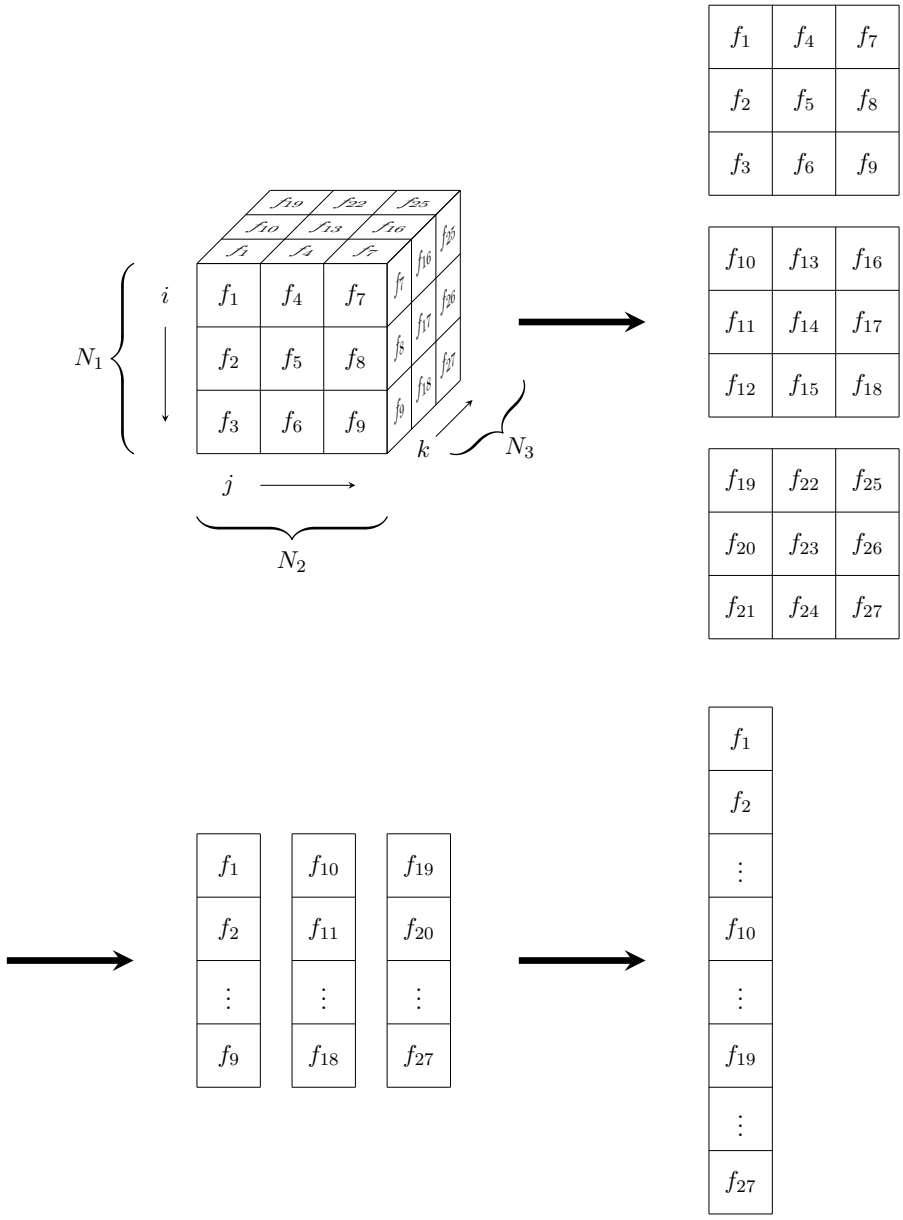


Fig. 2.4. Example of volume data indexing and flattening in a discretized 3D problem with $N_1 = N_2 = N_3 = 3$.

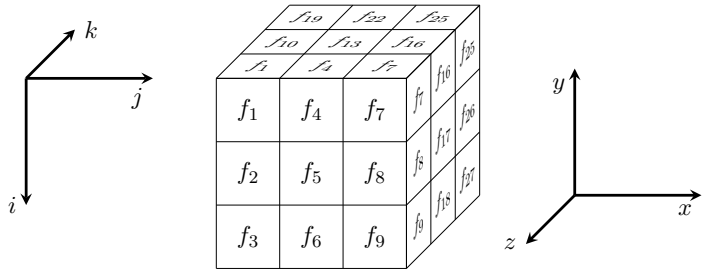


Fig. 2.5. Orientation of the voxel coordinate system and the Cartesian coordinate system with respect to discretized volumetric data.

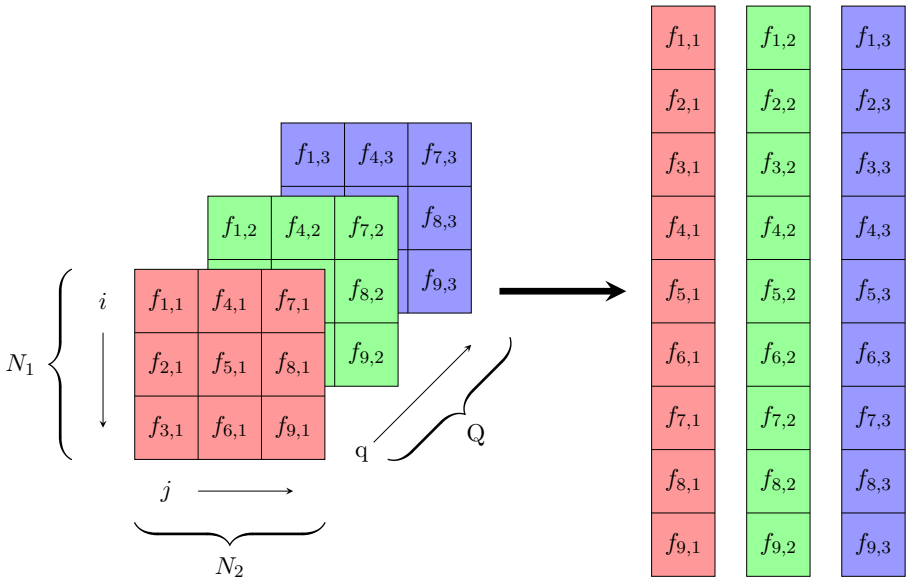


Fig. 2.6. Example of multispectral image data indexing and flattening in a discretized 2D problem with $N_1 = N_2 = 3$ and $Q = 3$.

3 Total Variation Regularization

3.1 The Total Variation Functional

Consider the regularized variational problem

$$\min_{f \in \mathbb{R}^N} \frac{1}{2} ||Af - m||_2^2 + \alpha R(f), \tag{3.1}$$

where f is the discretization of the continuous model $\mathcal{F} : \Omega \mapsto \mathbb{R}$, $\Omega \in \mathbb{R}^n$. An often employed choice for the regularizer $R(f)$ is the *total variation* (TV) functional, usually defined as

$$R_{\mathrm{TV}}(f) = \int_\Omega |\nabla f| \, d\Omega, \tag{3.2}$$

where ∇ should be understood as the discretized gradient operator, and the integral as numerical integration over the problem domain [85, 86]. Total variation is equivalent to the ℓ^1-norm of the gradient magnitude, and using the TV functional as a regularizer in Eq. (3.1) favours edge-preserving solutions that have sparse gradients and are, thus, approximately piecewise constant. Due to this property TV is well suited to signal denoising [85–90].

Total variation regularization has been extensively used in tomographic reconstruction. Regularization based on image gradient magnitude was first applied to CT reconstruction using simulated data [91]. Early studies using simulated data also applied TV regularization to image reconstruction in emission tomography [92, 93]. The first results using real data were presented in [94], and several further studies have followed [95–108]. Total variation and variants thereof have been utilized in image reconstruction in cone beam CT [109–123], helical CT [124], and photon-counting detector CT [125–136].

Publications I, II, and IV of this work utilize TV regularization to varying degrees. This chapter will give a brief overview of some aspects of the numerical implementation of TV, as well as introducing some variations on the standard TV formulation given in Eq. (3.2).

3.2 Numerical Implementation

Discrete derivatives can be implemented using *finite forward differences*. For a discretized 2D problem, as described in Section 2.5.1, they are defined as

$$\Delta_x f_{i,j} = f_{i,j+1} - f_{i,j} \tag{3.3}$$

and

$$\Delta_y f_{i,j} = f_{i+1,j} - f_{i,j}. \tag{3.4}$$

Using Eqs. (3.3) and (3.4) on an image of size $N_1 \times N_2$ will result in two images of size $(N_1-1)\times(N_2-1)$. For computational purposes, it is often desireable that transformations such as differentiation leave the data size unchanged. To this end, the discrete partial derivatives for 2D data can be defined as

$$\partial_x f_{i,j} = \begin{cases} f_{i,j+1} - f_{i,j}, & j \in \{1, \dots, N_2 - 1\} \\ 0, & j = N_2, \end{cases} \tag{3.5}$$

$$\partial_y f_{i,j} = \begin{cases} f_{i+1,j} - f_{i,j}, & i \in \{1, \dots, N_1 - 1\} \\ 0, & i = N_1, \end{cases} \tag{3.6}$$

where the last column or row of the resulting image is filled with zero-valued dummy variables [89]. Eqs. (3.5) and (3.6) can be used to define the 2D discrete gradient $\nabla : \mathbb{R}^{N_1 \times N_2} \mapsto \left(\mathbb{R}^{N_1 \times N_2}\right)^2$, a vector-valued linear map. The gradient magnitude at $f_{i,j}$ is now given by

$$||\nabla f_{i,j}||_2 = \sqrt{(\partial_x f_{i,j})^2 + (\partial_y f_{i,j})^2}. \tag{3.7}$$

Using these element-wise discrete definitions of the partial derivatives and the gradient magnitude, the notations $\partial_x f$, $\partial_y f$, and $||\nabla f||_2$ should each be understood as referring to image data in $\mathbb{R}^{N_1 \times N_2}$. Finally, the total variation of f is given by

$$TV(f) = \sum_{i}^{N_1} \sum_{j}^{N_2} \sqrt{(\partial_x f_{i,j})^2 + (\partial_y f_{i,j})^2}. \tag{3.8}$$

This definition for total variation was used in Publication II of this work.

Examination of Eq. (3.8) reveals that it consists of two nested ℓ^p norms: an inner map $\left(\mathbb{R}^{N_1 \times N_2}\right)^2 \mapsto \mathbb{R}^{N_1 \times N_2}$ computes the ℓ^2 norm of the gradient, *i.e.* the gradient magnitude, for each pixel, and an outer map $\mathbb{R}^{N_1 \times N_2} \mapsto \mathbb{R}$ computes the ℓ^1 norm of the gradient magnitude. This can be expressed concisely using the $\ell^{2,1}$ *mixed norm*:

$$TV(f) = ||\nabla f||_{2,1}. \tag{3.9}$$

Figure 3.1 shows an example of 2D image data f, its partial derivatives $\partial_x f$ and $\partial_y f$, and its gradient magnitude $||\nabla f||_2$, as defined above.

It is straightforward to extend the definitions above to 3D data $f \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ as defined in Section 2.5.2. The partial derivatives are now

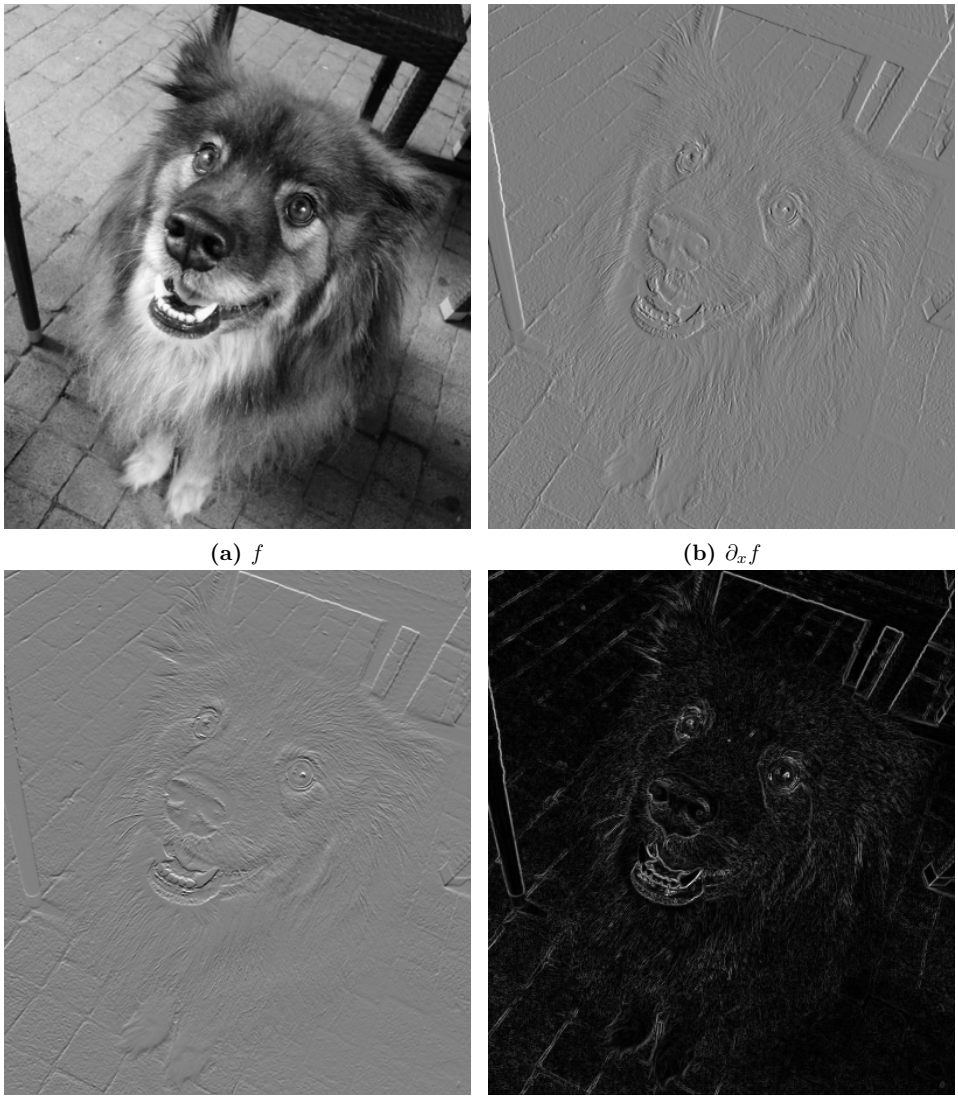
$$\partial_x f_{i,j,k} = \begin{cases} f_{i,j+1,k} - f_{i,j,k}, & j \in \{1, \dots, N_2 - 1\} \\ 0, & j = N_2, \end{cases} \tag{3.10}$$

$$\partial_y f_{i,j,k} = \begin{cases} f_{i+1,j,k} - f_{i,j,k}, & i \in \{1, \dots, N_1 - 1\} \\ 0, & i = N_1, \end{cases} \tag{3.11}$$

$$\partial_z f_{i,j,k} = \begin{cases} f_{i,j,k+1} - f_{i,j,k}, & k \in \{1, \dots, N_3 - 1\} \\ 0, & k = N_3, \end{cases} \tag{3.12}$$

the 3D discrete gradient is the vector-valued linear map $\nabla : \mathbb{R}^{N_1 \times N_2} \mapsto \left(\mathbb{R}^{N_1 \times N_2}\right)^3$, the gradient magnitude at $f_{i,j,k}$ is

$$||\nabla f_{i,j,k}||_2 = \sqrt{(\partial_x f_{i,j,k})^2 + (\partial_y f_{i,j,k})^2 + (\partial_z f_{i,j,k})^2}, \tag{3.13}$$



(a) f **(b)** $\partial_x f$

(c) $\partial_y f$ **(d)** $||\nabla f||_2$

Fig. 3.1. 2D image data, its discrete partial derivatives, and its gradient magnitude. In $\partial_x f$ and $\partial_y f$, gray pixels have a small gradient values, dark pixels have large negative values, and light pixels have large positive values. In $||\nabla f||_2$, black pixels are close to zero and light pixels have high values of the gradient magnitude. (Image credit: Outi Meaney.)

and the total variation of f is given by

$$TV(f) = \sum_{k}^{N_3} \sum_{i}^{N_1} \sum_{j}^{N_2} \sqrt{(\partial_x f_{i,j,k})^2 + (\partial_y f_{i,j,k})^2 + (\partial_z f_{i,j,k})^2}. \tag{3.14}$$

This definition for 3D total variation was used in Publication IV of this work. The notation in Eq. (3.9) holds for both 2D and 3D cases, as the definitions of ∇, f, and $||\cdot||_{2,1}$ will implicitly contain the information on the problem dimensions.

The definition above for computing gradients of discretized images and volumes is not the only possible one. Using finite forward differences will in fact result in shifting the data by half a pixel/voxel [89]. A more symmetric definition for discrete partial derivatives can be created using *central differences*, defined in a 2D setting as

$$\Delta_x f_{i,j} = (f_{i,j+1} - f_{i,j-1})/2 \tag{3.15}$$

and

$$\Delta_y f_{i,j} = (f_{i+1,j} - f_{i-1,j})/2. \tag{3.16}$$

The image edges can be handled by extending the image with its boundary values in each direction [137], resulting in the definitions

$$\partial_x f_{i,j} = \begin{cases} (f_{i,j+1} - f_{i,j})/2, & j = 1, \\ (f_{i,j+1} - f_{i,j-1})/2, & j \in \{2, \ldots, N_2 - 1\}, \\ (f_{i,j} - f_{i,j-1})/2, & j = N_2, \end{cases} \tag{3.17}$$

$$\partial_y f_{i,j} = \begin{cases} (f_{i+1,j} - f_{i,j})/2, & i = 1, \\ (f_{i+1,j} - f_{i-1,j})/2, & i \in \{2, \ldots, N_1 - 1\}, \\ (f_{i,j} - f_{i-1,j})/2, & i = N_1. \end{cases} \tag{3.18}$$

These definitions can easily be extended to 3D volumes. However, central differences miss thin structures, and are therefore not well suited for TV discretization in general [86].

In the field of digital image processing, there exist further definitions for $\partial_x f_{i,j}$ and $\partial_y f_{i,j}$, each involving a different kind of weighted sum of a local neighbourhood of $f_{i,j}$ [138]. These are rarely seen in the TV literature, as they can suffer from the same shortcomings as central differences, and they can also lead to computationally cumbersome expressions in the TV minimization problem.

Alternative approaches to gradient discretization for total variation regularization include upwind discretization [139], Condat's definition [89], and the Shannon Total Variation [140]. Further considerations relating to gradient discretization in the context of TV regularization can be found also in [86]. In practice, the discretization approach used in this work, based on finite forward differences with zero-valued dummy variables used where necessary, remains the most often seen approach for TV regularization in the scientific literature.

3.3 Variants

There exists an abundance of variants and extensions of total variation regularization, many of which appear in the works referenced at the beginning of this chapter. This section will introduce some variations of TV methods which have been used in the publications that form the body of this work.

3.3.1 Smoothed Total Variation

The gradient magnitude expressions in Eqs. (3.7) and (3.13) are not everywhere differentiable. This significantly limits choosing and implementing a numerical optimization method for the minimization in Eq. (3.1), because it rules out gradient-based methods such as gradient descent and the conjugate gradient method, which are widely used and relatively straighforward to implement.

The total variation scheme can be made everywhere differentiable by using *smoothed total variation* (STV). In the 2D case, the expression in Eq. (3.7) is replaced with

$$||\nabla f_{i,j}||_\beta = \sqrt{(\partial_x f_{i,j})^2 + (\partial_y f_{i,j})^2 + \beta^2}, \tag{3.19}$$

where the *smoothing parameter* $\beta > 0$ is small. The total variation functional now becomes

$$TV_\beta(f) = \sum_i^{N_1} \sum_j^{N_2} \sqrt{(\partial_x f_{i,j})^2 + (\partial_y f_{i,j})^2 + \beta^2}. \tag{3.20}$$

In the 3D setting, the corresponding equations are

$$||\nabla f_{i,j,k}||_\beta = \sqrt{(\partial_x f_{i,j,k})^2 + (\partial_y f_{i,j,k})^2 + (\partial_z f_{i,j,k})^2 + \beta^2}, \tag{3.21}$$

and

$$TV_\beta(f) = \sum_k^{N_3} \sum_i^{N_1} \sum_j^{N_2} \sqrt{(\partial_x f_{i,j,k})^2 + (\partial_y f_{i,j,k})^2 + (\partial_z f_{i,j,k})^2 + \beta^2}. \tag{3.22}$$

Smoothed total variation regularization has been extensively applied in computed tomography [42, 96, 98, 105, 141, 132, 133, 142]. The smoothed 2D TV scheme given by Eqs. (3.19) and (3.20) was also used in Publication I of this work.

3.3.2 Anisotropic Total Variation

Anisotropic total variation (ATV) directly uses the absolute values of the gradient coefficients. The regularizers for the 2D and 3D settings now become

$$ATV(f) = \sum_i^{N_1} \sum_j^{N_2} \left(|\partial_x f_{i,j}| + |\partial_y f_{i,j}|\right) \tag{3.23}$$

and

$$ATV(f) = \sum_k^{N_3} \sum_i^{N_1} \sum_j^{N_2} \left(|\partial_x f_{i,j,k}| + |\partial_y f_{i,j,k}| + |\partial_z f_{i,j,k}|\right), \tag{3.24}$$

respectively. ATV regularization is more straightforward to implement than the standard (isotropic) TV formulation, as it does not involve the square root computations and the $\ell^{2,1}$ mixed norm. Unfortunately, ATV tends to favour transitions that are aligned with the Cartesian coordinate axes, leading to unrealistic reconstructions, especially at high values of the regularization parameter α. There are however certain settings where this can be turned into an advantage, for example by using separate regularization parameters α_x and α_y for the gradient coefficients in the x-, and y-directions in cases where the data collection process itself has been anisotropic in nature, an idea that can also be further extended to favour arbitrary directions in the gradient [143].

In CT reconstruction, ATV as formulated in Eq. (3.23) has been used in, *e.g.* [144–146], and in the ATV-NN-PDFP algorithm detailed in Section 5.5.1.

3.3.3 Joint Total Variation

Joint total variation (JTV) is a useful regularization technique in multispectral data reconstruction as it favours non-zero gradients, *i.e.* edges, in parallel locations across the different data channels. JTV is also sometimes called *vectorial total variation*, and it was first developed for denoising colour television signals [147]. For multispectral data $f \in \left(\mathbb{R}^N\right)^Q$ consisting of Q channels, the JTV regularizer was originally defined as

$$JTV(f) = \sqrt{\sum_{q=1}^{Q} [(TV(f_q)]^2}. \tag{3.25}$$

Using JTV thus requires that each channel in the multispectral data has the same discretization. Joint total variation creates an implicit edge-preserving coupling between the channels, favouring sparsity of the joint gradient.

Joint total variation was used for 2D multispectral CT reconstruction in Publication I of this work. This work used the definition

$$JTV(f) = \sum_{i}^{N_1} \sum_{j}^{N_2} \sqrt{\sum_{q=1}^{Q} ||\nabla f_{i,j,q}||_2^2 + \beta^2}, \tag{3.26}$$

where a smoothing parameter was used to make the regularizer everywhere differentiable to allow for gradient-based optimization.

3.3.4 Directional Total Variation

Directional total variation (dTV) can be used to add *a priori* information to image reconstruction when some kind of *side information* f_s on the structure of the object is available [148]. The assumption in dTV is that the imaging data being reconstructed and the side information share both the location and the direction of edges, modulo their sign. The sign modulo is needed, as it is possible that in one image or volume the gradient is increasing, whereas in the other it is decreasing. The dTV scheme will only be considered for a 2D setting here, but it is possible to extend it to higher dimensions.

Let $\xi \in \left(\mathbb{R}^2\right)^{N_1 \times N_2}$ be a vector field such that $0 \leq ||\xi_{i,j}||_2 \leq 1$. ξ now represents some type of edge information on the side information f_s. Let $P_{\xi_{i,j}} \in \mathbb{R}^{2\times 2}$ be $P_{\xi_{i,j}} = I - \xi_{i,j} \otimes \xi_{i,j}$, where $\otimes$ is the outer product between two vectors. This yields $P_{\xi_{i,j}} u = u - \langle \xi_{i,j}, u \rangle \xi_{i,j}$ for $u \in \mathbb{R}^2$. The directional total variation of f is now a mapping $dTV : \mathbb{R}^{N_1 \times N_2} \mapsto \mathbb{R}$ such that

$$dTV(f) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} ||P_{\xi_{i,j}} \nabla f_{i,j}||_2. \tag{3.27}$$

Directional total variation was used for 2D multispectral CT reconstruction in Publication II of this work. In this work, the edge information f_s was a single prior image obtained by combining all the data from the different energy channels into a single reconstruction, and the vector field used in the dTV computations was

$$\xi_{i,j} = \eta \frac{\nabla f_{i,j}}{||\nabla f_{i,j}||_\beta}, \tag{3.28}$$

where η is an edge parameter related to the size of an edge.

3.4 Discrete Gradient Matrices

Every linear map between Euclidean spaces can be represented as a matrix. It can be beneficial to use an explicit matrix formulation of the discrete partial derivative operators in Eqs. (3.5)-(3.6) and Eqs. (3.10)-(3.12), for multiple reasons. Many programming languages and numerical computing environments are highly optimized for efficient matrix-vector product computations. Furthermore, many numerical algorithms require using the transpose ∇^T of the linear map representing the discrete gradient, and it is not always clear how this should be computed when an explicit matrix representation is not available, or what is the most efficient way of performing the computations in such cases. Finally, many numerical algorithms also require information on the operator norm of the linear map ∇, and having an explicit matrix representation provides a convenient way of studying this.

This section will discuss the construction and operator norm properties of *discrete gradient matrices*. Whenever this matrix formulation for discrete partial derivative operators is used, f must always be treated in flattened form, *i.e.* as a vector, as described in Section 2.5. Discrete gradient matrices were used in the algorithms detailed in Section 5.5 and in Publication IV of this work.

We begin with the 2D case. Let I_{N_1} be the $N_1 \times N_1$ identity matrix. The operator ∂_x as defined in Eq. (3.5) can be constructed as the $(N_1N_2) \times (N_1N_2)$ block bidiagonal matrix

$$\Delta_x = \begin{bmatrix} -I_{N_1} & I_{N_1} & 0 & \cdots & 0 \\ 0 & -I_{N_1} & I_{N_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -I_{N_1} & I_{N_1} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{3.29}$$

where 0 indicates the $N_1 \times N_1$ zero matrix. The matrix Δ_x consists of $N_2 \times N_2$ blocks. To avoid confusion, it is important to note that the symbol Δ is taken to signify finite forward differences as defined in Eqs. (3.3)-(3.4) when applied to an individual pixel $f_{i,j}$ or voxel $f_{i,j,k}$, and a matrix definition when it is applied to the full image or volume f. The difference is assumed to be implicitly clear from the context.

For ∂_y, we begin by defining the $N_1 \times N_1$ bidiagonal matrix

$$\delta_y = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{3.30}$$

The operator ∂_y as defined in Eq. (3.6) can be constructed as the $(N_1N_2) \times (N_1N_2)$ block diagonal matrix

$$\Delta_y = \begin{bmatrix} \delta_y & 0 & \cdots & 0 \\ 0 & \delta_y & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_y \end{bmatrix}, \tag{3.31}$$

where 0 indicates the $N_1 \times N_1$ zero matrix. Δ_y consists of $N_2 \times N_2$ blocks.

We now define the $(2N_1N_2) \times (N_1N_2)$ discrete gradient matrix

$$D = \begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} \tag{3.32}$$

for 2D images. After computing Df, its elements $1, \ldots, (N_1N_2)$ contain the discrete partial derivative $\partial_x f$ in flattened form, and elements $(N_1N_2 + 1), \ldots, (2N_1N_2)$ contain the discrete partial derivative $\partial_y f$ in flattened form. If one wishes to study $\partial_x f$ or $\partial_y f$ individually, they should be extracted and reshaped into $N_1 \times N_2$ images.

In the 3D case, $\partial_x f$ and $\partial_y f$ can be obtained by repeatedly applying Δ_x and Δ_y to each z-plane in the volume data, respectively. Thus, the operators ∂_x and ∂_y as defined in Eqs. (3.10) and (3.11) can be constructed as the $(N_1N_2N_3) \times (N_1N_2N_3)$ block diagonal matrices

$$D_x = \begin{bmatrix} \Delta_x & 0 & \cdots & 0 \\ 0 & \Delta_x & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ddots & \Delta_x \end{bmatrix}, \tag{3.33}$$

and

$$D_y = \begin{bmatrix} \Delta_y & 0 & \cdots & 0 \\ 0 & \Delta_y & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ddots & \Delta_y \end{bmatrix}, \tag{3.34}$$

where 0 indicates the $(N_1N_2) \times (N_1N_2)$ zero matrix. D_x and D_y both consist of $N_3 \times N_3$ blocks.

The operator ∂_z is also very straightforward to construct. Let $I_{N_1N_2}$ be the $(N_1N_2) \times (N_1N_2)$ identity matrix. The operator ∂_z as defined in Eq. (3.12) can be constructed as the $(N_1N_2N_3) \times (N_1N_2N_3)$ block bidiagonal matrix

$$D_z = \begin{bmatrix} -I_{N_1N_2} & I_{N_1N_2} & 0 & \cdots & 0 \\ 0 & -I_{N_1N_2} & I_{N_1N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -I_{N_1N_2} & I_{N_1N_2} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{3.35}$$

where 0 once again indicates the $(N_1N_2) \times (N_1N_2)$ zero matrix. D_z also consists of $N_3 \times N_3$ blocks.

We now define the $(3N_1N_2N_3) \times (N_1N_2N_3)$ discrete gradient matrix

$$D = \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} \tag{3.36}$$

for 3D volumes. In Df, elements $1, \ldots, (N_1N_2N_3)$ contain $\partial_x f$, elements $(N_1N_2N_3 + 1), \ldots, (2N_1N_2N_3)$ contain $\partial_y f$, and elements $(2N_1N_2N_3+1), \ldots, (3N_1N_2N_3)$ contain $\partial_z f$, each in flattened form. If one wishes to study these discrete partial derivatives individually, they should be extracted and reshaped into $N_1 \times N_2 \times N_3$ volumes.

It is important to note that discrete gradient matrices are extremely sparse: by definition, each row can contain only two non-zero elements. This sparsity can be used for efficient data storage in many computational environments. Nevertheless, for 3D volumes where the number of elements N is large, memory requirements for D can become significant.

3.4.1 Discrete Gradient Matrix Operator Norm

Many optimization algorithms designed to numerically solve variational regularization problems of the form Eq. (2.20) require knowledge on the operator norm of the linear transformation used in the regularization term. The ℓ^p operator norm for any real square matrix A is defined as

$$||A||_p = \sup_{f \neq 0} \frac{||Af||_p}{||f||_p}. \tag{3.37}$$

The interpretation of the ℓ^p operator norm is that it gives an upper limit to how much multiplication with A can stretch a vector with a given length in the ℓ^p norm.

The ℓ^2 operator norm $||A||_2$ is also known as the *spectral norm*. The spectral norm has the property [44]

$$||A||_2 = \sqrt{\lambda_{\max}(A^T A)} = \sigma_{\max}(A). \tag{3.38}$$

This definition of the ℓ^2 operator norm holds for non-square matrices also. The maximum of the absolute values of the eigenvalues of a square matrix is called the *spectral radius* of the matrix, not to be confused with the spectral norm.

We will now show that $\lambda_{\max}(D^T D) \leq 8$ for 2D problems and and $\lambda_{\max}(D^T D) \leq 12$ for 3D problems. This will be done using *Gershgorin's Disk Theorem*, also known as the *Gershgorin Circle Theorem* [44, 149].

Definition 3.1 *Let A be an $N \times N$ matrix with elements a_{ij}. Let r_i be the sum of the absolute values of the off-diagonal elements on row i of A:*

$$r_i = \sum_{i \neq j} |a_{ij}|.$$

Let $\overline{B}(a_{ii}, r_i)$ be a closed disk[1] *with center a_{ii} and radius r_i. Such a disk is called a Gershgorin disk.*

Theorem 3.1 ***Gershgorin's Disk Theorem*** *Let A be an $N \times N$ matrix. Every eigenvalue of A is contained within a Gershorin disk.*

Now, let D be a discrete gradient matrix, and let $d \in \{2, 3\}$ indicate the dimensionality of the underlying problem. We begin by noting that $D^T D$ is symmetric, and its eigenvalues are therefore real. Next, we note that $D^T D$ and DD^T have the same non-zero eigenvalues. In this particular case, the eigenvalues of the $(3N_1N_2N_3) \times (3N_1N_2N_3)$ matrix DD^T are easier to determine.

By definition, D does not contain any identical rows, each one of its rows contains either zero or two non-zero elements, and all rows containing non-zero elements will contain the values -1 and 1. Directly from the definitions in Eqs. (3.5)-(3.6) and Eqs. (3.10)-(3.12) it can be seen that a given element of f can be be involved in the discrete partial derivative computations a maximum of d times with a positive coefficient and d times with a negative coefficient. Each column of D corresponds to a specific element of f, and therefore each column can have a maximum of d positive coefficients and d negative coefficients.

The element (i, j) of the matrix DD^T is the dot product of rows i and j of D. By definition, D does not contain any identical rows, so therefore only the diagonal elements of DD^T can contain the value 2. Some diagonal elements will be zero-valued. The off-diagonal elements can only ever contain the values -1, 0, or 1. From the properties of D

[1] Standard notation for a disk uses the symbol D. To avoid ambiguous notation, this work uses the notation for a topological closed ball in $\mathbb{R}$ or $\mathbb{C}$ for the Gershgorin disk.

described above, it is possible to compile a comprehensive list of scenarios that result in a non-zero element at location (i, j) of the matrix DD^T when $i \neq j$:

- A value of -1 on row i of D can be multiplied with a value of -1 on row j of D in exactly $d - 1$ ways.
- A value of -1 on row i of D can be multiplied with a value of 1 on row j of D in exactly d ways.
- A value of 1 on row i of D can be multiplied with a value of 1 on row j of D in exactly $d - 1$ ways.
- A value of 1 on row i of D can be multiplied with a value of -1 on row j of D in exactly d ways.

Therefore, the maximum possible number of non-zero entries on row i of DD^T is

$$2(d-1) + 2d = 4d - 2.$$

Because the absolute value of all non-zero, off-diagonal elements of DD^T is 1, the largest possible radius of a Gershgorin disk is

$$r_i = 6 \quad \text{when } d = 2$$

and

$$r_i = 10 \quad \text{when } d = 3.$$

From Gershgorin's Disk Theorem it then follows that

$$||D^T D||_2^2 = \lambda_{\max}(D^T D) \leq 8 \quad \text{when } d = 2, \tag{3.39}$$

and

$$||D^T D||_2^2 = \lambda_{\max}(D^T D) \leq 12 \quad \text{when } d = 3. \tag{3.40}$$

4 Mathematical Optimization

4.1 Mathematical Optimization Problems

Mathematical optimization deals with problems of the form

$$\begin{aligned} &\text{minimize} && L(f) \\ &\text{subject to} && G_i(f) \leq 0, \quad i = 1, \ldots, P \\ &&& H_i(f) = 0, \quad i = 1, \ldots, S \end{aligned} \tag{4.1}$$

where $f \in \mathbb{R}^N$ is the optimization variable, $L : \mathbb{R}^N \mapsto \mathbb{R}$ is the *objective function* (or *loss function* or *cost function)*, $G_i : \mathbb{R}^N \mapsto \mathbb{R}$, $i = 1, \ldots, P$ are the *inequality constraints*, and $H_i : \mathbb{R}^N \mapsto \mathbb{R}$, $i = 1, \ldots, S$ are the *equality constraints* [150]. The constraint functions define the *domain* $\Omega \subset \mathbb{R}^N$ of the problem, and Eq. (4.1) can also be expressed as

$$\begin{aligned} &\text{minimize} && L(f) \\ &\text{subject to} && f \in \Omega \end{aligned} \tag{4.2}$$

or

$$\min_{f \in \Omega} L(f). \tag{4.3}$$

In the case where there are no constraint functions, the problem is an *unconstrained optimization* problem. A point $f \in \mathbb{R}^N$ is called *feasible* if it belongs to Ω, the collection of feasible points is known as the *feasible set*, and the problem is said to be feasible if at least one feasible point exists.

Mathematical optimization, also known as *mathematical programming*, is a vast and diverse field of applied mathematics, and an even cursory overview is beyond the scope of this work. In the field of computational inverse problems, optimization theory and techniques are used to solve minimization problems such as the variational regularization problem in Eq. (2.19). Identifying the properties of an optimization problem is crucial to choosing and implementing a suitable *solution method* for solving it. This chapter will briefly discuss some key concepts and techniques in optimization essential to this work. Basic definitions and results about optimization theory can be found in many standard textbooks, such as [150–152], and they are stated here without proofs.

4.2 Basic Concepts

Let $L : \Omega \mapsto \mathbb{R}$, where $\Omega \subset \mathbb{R}^N$. We say that the point $f^* \in \Omega$ is a *local minimizer* of $L(f)$ if there exists a neighbourhood U of f^* such that

$$L(f^*) \leq L(f) \quad \forall\, f \in U. \tag{4.4}$$

In the context of mathematical optimization, the concepts local minimizer and *local optimum* can be used interchangeably. The point $f^* \in \Omega$ is said to be a *global minimizer* or *global optimum* if

$$L(f^*) \leq L(f) \quad \forall\, f \in \Omega. \tag{4.5}$$

Solving an optimization problem means finding the global optimum, which requires that one is able to recognize both locally and globally optimal points. A point is a local minimizer if it fulfills certain *optimality conditions*, which depend on the nature of the optimization problem. Recognizing local optimality is easier than recognizing global optimality, which can range from straightforward to hope-crushingly difficult.

It is quite rare for a practical optimization problem to have a solution that can be found analytically. Therefore, most optimization problems need to be solved iteratively, whereby a starting point f^0 is chosen, and set of optimization *steps* are taken, such that the solutions f^ν gradually improve (*i.e.* the value of the cost function decreases) and converge towards a local or global optimum as the iterative process proceeds. The development of robust optimization algorithms with good convergence properties is a central part of the field of optimization research. Due to the nature of floating-point arithmetic it may be that the exact optimality conditions are never met, and therefore iterative optimization methods usually run for a set number of iterations, or until some *stopping criterion* is met, whereby the solution is deemed to be sufficiently good by some metric.

Optimization problems sometimes make use of the *extended real number line*, where the set of real numbers is extended with the infinity elements $-\infty$ and ∞. The notation used for the extended real number line is $\mathbb{R} \cup \{-\infty, \infty\}$. The negative infinity $-\infty$ rarely appears in optimization. However, the positive infinity ∞ can be used to define the value of the cost function as infinitely high in some region of $\mathbb{R}^N$, effectively constraining the domain of the problem outside that region.

There are multiple ways of classifying optimization problems and solution methods. In *linear optimization* the objective function and the constraint functions are linear. Linear optimization is well understood theoretically, and many highly efficient solution methods exist. *Nonlinear optimization* constitutes a large group of optimization problems where the objective or constraint functions are not linear. The difficulty of nonlinear optimization problems varies considerably.

When the objective function $L(f)$ is differentiable, the expression *smooth optimization* is often used. Solution methods in such cases can be categorized as *first order* methods, *second order* methods, etc., depending on the degree of derivatives that the method uses. A *zeroth-order* method is one that only uses evaluations of the objective function, but no differential information.

Yet another hugely important way of classifying optimization problems is whether they are *convex* of *non-convex*, which greatly impacts the identification of global optimizers.

4.3 Convexity

Convexity is a concept of central importance in mathematical optimization. Identifying a problem as convex or non-convex guides the choice a suitable solution method. The theory of convex optimization is well understood, and solving convex problems is substantially easier and more straightforward than solving non-convex problems. Indeed, R. Tyrrell Rockafellar, a pioneer of optimization theory and convex analysis, has stated the following: "*the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity*" [153].

In order to identify a problem being convex, the concepts of a *convex set* and a *convex function* are needed. Their definitions and some of their properties will now follow.

Definition 4.1 *Let $\Omega \subset \mathbb{R}^N$. The set Ω is convex if for all $f_1, f_2 \in \Omega$ and any λ such that $0 \leq \lambda \leq 1$, we have $\lambda f_1 + (1-\lambda) f_2 \in \Omega$.*

In other words, if every point in the line segment connecting f_1 and f_2 also belongs to Ω, then Ω is a convex set. Important cases of convex sets include the hyperplane

$$H = \left\{ f \in \mathbb{R}^N \mid a^T (f - f_0) = 0 \right\}, \tag{4.6}$$

where $a \in \mathbb{R}^N$, $a \neq 0$ is the normal vector to the hyperplane, and $f_0 \in \mathbb{R}^N$ is any point in the hyperplane. The hyperplane splits $\mathbb{R}^N$ into two, defining the *halfspace*

$$S = \{ f \in \mathbb{R}^N \mid a^T (f - f_0) \leq 0 \}, \tag{4.7}$$

where a is the outward-pointing normal vector of S

It is important to identify operations which preserve convexity of sets. Two such operations are presented in the following proposition.

Proposition 4.1 *Let Ω_1 and Ω_2 be convex sets in $\mathbb{R}^N$. Then the following properties hold:*

- *The intersection $\Omega = \Omega_1 \cap \Omega_2$ is also a convex set.*
- *Let $A \in \mathbb{R}^{M \times N}$ and $b \in \mathbb{R}^M$. The image of Ω_1 under the affine transformation $\Omega = \{Af + b \,|\, f \in \Omega_1\}$ is also a convex set.*

Important to this work, the *nonnegative orthant*

$$\mathbb{R}^N_+ = \{ f \in \mathbb{R}^N \mid f_i \geq 0, i = 1, \ldots, N \} \tag{4.8}$$

is a convex set, as it can be constructed as an intersection of halfspaces.

We proceed to convex functions and some of their key properties.

Definition 4.2 *Let $\Omega \subset \mathbb{R}^N$ be a convex set. The function $L : \Omega \mapsto \mathbb{R}$ is convex if for all $f_1, f_2 \in \Omega$ and any λ such that $0 \leq \lambda \leq 1$, we have*

$$L(\lambda f_1 + (1-\lambda) f_2) \leq \lambda L(f_1) + (1-\lambda) L(f_2).$$

If strict inequality holds in Definition 4.2 whenever $f_1 \neq f_2$ and $0 < \lambda < 1$, then f is said to be *strictly convex*.

The following propositions present important information on the convexity of certain types of functions and on convexity-preserving operations used in this work.

Proposition 4.2 *The following functions are known to be convex:*

1. *$L : \mathbb{R}_+ \mapsto \mathbb{R}$, $L(f) = f^p$ on when $p \geq 1$ or $p \leq 0$.*
2. *$L : \mathbb{R} \mapsto \mathbb{R}$, $L(f) = |f|^p$ when $p \geq 1$.*
3. *Every norm on $\mathbb{R}^N$ is convex.*

Proposition 4.3 *The following operations are known to preserve convexity:*

1. *Let* $L_1, \ldots, L_K : \mathbb{R}^N \mapsto \mathbb{R}$ *be convex and* $w_i \geq 0 \; \forall \; i \in \{1, \ldots, K\}$. *The nonnegative weighted sum* $L(f) = \sum_{i=1}^{K} w_i L_i$ *is convex.*

2. *Let* $L : \mathbb{R}^N \mapsto \mathbb{R}$ *be convex,* $A \in \mathbb{R}^{M \times N}$, $b \in \mathbb{R}^M$, *and* $f \in \mathbb{R}^N$. *The function* $G : \mathbb{R}^M \mapsto \mathbb{R}$, $G(f) = L(Af + b)$ *is convex.*

3. *Let* $H : \mathbb{R}^N \mapsto \mathbb{R}$ *be convex and* $G : \mathbb{R} \mapsto \mathbb{R}$ *be convex and nondecreasing. Then* $L : \mathbb{R}^N \mapsto \mathbb{R}$, $L(f) = G(H(f))$ *is convex.*

If the domain Ω and cost function $L(f)$ in the problem defined in Eq. (4.2) are both convex, then the optimization problem is said to be convex.

The reason that convex optimization problems are of such high interest can be inferred from the following theorem.

Theorem 4.1 *Let* Ω *be a nonempty convex set and let* $L : \Omega \mapsto \mathbb{R}$ *convex on* Ω. *Let* f^* *be a local optimizer in the problem*

$$\min_{f \in \Omega} L(f).$$

Then f^* *is also the global optimizer.*

Convexity thus allows knowledge on global properties based on only local knowledge. Solving a convex optimization problem requires only that the solution method is able to find local solutions, and knowledge on the convexity of the problem guarantees that the solution is the global optimizer.

As mentioned in Section 4.2, optimization problems sometimes use the extended real number line $\mathbb{R} \cup \{-\infty, \infty\}$ as the codomain of the cost function $L(f)$. This can lead to problems if it is not applied judiciously. A function is said to *proper* if $L(f) < \infty$ for at least one $f \in \Omega$ and $L(f) > -\infty$ for all $f \in \Omega$. If a function is not proper, it is *improper*. Convex optimization rules out using improper functions as the objective function.

The necessary tools are now in place to identify many computational inverse problems as convex optimization problems. Take the regularized variational problem that uses the loss function in Eq. (2.20), *i.e.* we want to find f^* that solves the problem

$$\min_{f \in \mathbb{R}^N} \frac{1}{2} ||Af - m||_2^2 + \alpha ||Bf||_p^p. \tag{4.9}$$

The term $||Af - m||_2$ is the composite of an affine map and a norm, and is therefore a convex function. Norms by definition map to the nonnegative numbers $\mathbb{R}_+$ and therefore the term $||Af - m||_2^2$ is also a convex function. Likewise, $||Bf||_p^p$ is the composite of a linear map and an ℓ^p norm raised to the power $p \geq 1$, and it is thus convex. The regularization parameter α is defined to be greater than zero, and the cost function is a nonnegative weighted sum of two convex functions, and it is therefore convex, allowing convex optimization methods to be applied to the problem.

Unfortunately, identifying a problem as convex is not yet an exhaustive description and doesn't yield all the information needed to solve it. Namely, it gives no information on the smoothness of the cost function, or how challenging the constraints of Ω are to handle. These factors further dictate the choice of the solution method.

Convex optimization algorithms play an important role in advanced medical image reconstruction methods. They are used in both model-based iterative reconstruction, as seen in this work, and also in modern hybrid methods which combine iterative reconstruction with techniques based on deep learning (DL) and artificial intelligence (AI) [154].

Even as DL and AI continue to take medical image reconstruction into new directions beyond conventional model-based regularized inversion [32, 155], it seems highly probable that convex optimization will retain its importance in the field.

4.4 Gradient-Based Algorithms

Consider the unconstrained optimization problem

$$\min_{f \in \mathbb{R}^N} L(f), \tag{4.10}$$

where $L : \mathbb{R}^N \mapsto \mathbb{R}$ is a proper convex continuously differentiable function. Because $L(f)$ is continuously differentiable everywhere in $\mathbb{R}^N$, the problem is also categorized as a smooth optimization problem. The gradient contains information on the direction and rate of change of a function, so it is natural to use this information to iteratively minimize $L(f)$. Such optimization methods are called *gradient-based algorithms*.

All gradient-based algorithms are, in principle, based on similar procedures taken at each iteration $\nu \in \mathbb{N}$:

1. Based on information about $\nabla L(f^{\nu-1})$, choose a *descent direction* S^ν.
2. Choose a *step length* $\gamma^\nu \geq 0$.
3. Obtain the new iterate with $f^\nu = f^{\nu-1} + \gamma^\nu S^\nu$.

Smooth unconstrained convex optimization problems are often highly desirable, as they also have very simple optimality conditions:

Theorem 4.2 *Let $L : \mathbb{R}^N \mapsto \mathbb{R}$ be proper convex continuously differentiable. f^* is the global minimizer of $L(f)$ if and only if $\nabla L(f^*) = 0$.*

In computational settings, the problem is treated using floating-point arithmetic on a computer, and $\nabla L(f)$ will practically never be exactly zero. Therefore a suitable stopping criterion must be defined to stop the iterations. One approach is to simply let the iterations run for a maximum number of iterations $\nu_{\max}$. Additional criteria can be established by defining some quantity that is compared to a tolerance ε at each iteration, and if the quantity is below the tolerance, the iterations are interrupted. Possible stopping criteria include the gradient magnitude $||\nabla L(f^\nu)||_2$, the absolute difference between consecutive iterations $||f^\nu - f^{\nu-1}||_2$, and the relative difference between consecutive iterations $||f^\nu - f^{\nu-1}||_2 / ||f^\nu||_2$.

The straightforward nature of the solution method and the optimality conditions have made gradient-based optimization popular in medical image reconstruction [37, 156]. For instance, the purpose of using a smoothing parameter in total variation regularization, as discussed in Section 3.3, is to render the problem solvable via smooth unconstrained optimization.

Gradient-based optimization methods are categorized as first-order methods, because they only use information on the first (partial) derivatives of $L(f)$. The following section will discuss two gradient-based algorithms that are widely applied in computational inverse problems, including this work: *gradient descent* and *conjugate gradient*.

4.4.1 Gradient Descent and Steepest Descent

Gradient descent (GD) is a first-order iterative optimization method that is widely used in engineering and machine learning. In GD we set

$$S^{\nu} = -\nabla L(f^{\nu-1}). \tag{4.11}$$

There are many different ways of choosing the step length $\gamma^{\nu} \geq 0$. If it chosen so that $L(f)$ is minimized along $f^{\nu-1}+\gamma^{\nu}S^{\nu}$, the method is often called *steepest descent* (SD), although the terms gradient descent and steepest descent are not always used in a consistent manner in the literature. Minimization-within-minimization tasks such as the step length choice in SD are frequently encountered in mathematical optimization. In steepest descent, it is usually necessary to use a *line search* algorithm to solve the univariate minimization problem

$$\gamma^{\nu} = \underset{\gamma \in \mathbb{R}_+}{\arg\min} \left\{ L\left(f^{\nu-1} + \gamma S^{\nu}\right) \right\}. \tag{4.12}$$

One widely applied line search method is the zeroth-order algorithm *golden section line search* [157, 158]. It has robust convergence properties, and it is widely applicable and easy to implement, as no derivatives are needed. The pseudocode for golden section line search is given in Algorithm 4.1. It is important to emphasize that although line search algorithms are initially presented here in the context of steepest descent optimization, they are very widely and generally applicable in mathematical optimization. The pseudocode for steepest descent is given in Algorithm 4.2.

Algorithm 4.1 Golden Section Line Search

Input: function handle $L(f)$, starting point $f^{\nu-1}$, search direction S^{ν}, search interval $[a, b]$, tolerance ε
Initialize: $\varphi \leftarrow \frac{1+\sqrt{5}}{2}$
while $|b - a| > \varepsilon$ **do**
  $c = b - \frac{b-a}{\varphi}$
  $d = a + \frac{b-a}{\varphi}$;
  if $L(f^{\nu-1} + cS^{\nu}) < L(f^{\nu-1} + dS^{\nu})$ **then**
    $b \leftarrow d$
  else
    $a \leftarrow c$
  end if
end while
return $\frac{a+b}{2}$

Despite its simplicity and wide applicability, steepest descent is somewhat handicapped by slow convergence properties. This can be demonstrated heuristically via a simple toy example. In order to experimentally study the convergence properties of gradient-based algorithms, consider the problem of minimizing the least squares functional

$$L(f) = ||Af - m||_2^2, \tag{4.13}$$

where the example problem is now made two-dimensional with the values

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}, \quad m = \begin{bmatrix} 2 \\ -1 \end{bmatrix}. \tag{4.14}$$

Algorithm 4.2 Steepest Descent

Input: function handles $L(f)$ and $\nabla L(f)$, starting point f^0, maximum number of iterations $\nu_{\max}$, tolerance ε
for $\nu \leftarrow 1, \ldots, \nu_{\max}$ **do**
  $S^\nu \leftarrow -\nabla L(f^{\nu-1})$
  $\gamma^\nu \leftarrow \underset{\gamma \in \mathbb{R}_+}{\arg\min}\, L(f^{\nu-1} + \gamma S^\nu)$
  $f^\nu = f^{\nu-1} + \gamma^\nu S^\nu$
  if stopping criterion quantity $< \varepsilon$ **then**
    return f^ν
  end if
end for
return f^ν

Figure 4.1 shows the steepest descent steps, the univariate line search function $h(\gamma) = L(f^{\nu-1} - \gamma \nabla L(f^{\nu-1}))$ for iteration $\nu = 1$, and the convergence behaviour of $L(f^\nu)$ and $||\nabla L(f^\nu)||_2$ for problem (4.13) - (4.14) with the initial guess $f^0 = [-4, 4]^T$. The SD steps demonstrate a very clear zig-zag behaviour, and despite the simplicity of the problem setup, it took 39 iterations on a computer test to reach a stopping criterion of $||\nabla L(f^\nu)||_2 < 10^{-3}$. The cost function $L(f^\nu)$ decreases monotonously with the iterates.

In order to speed up convergence in gradient descent, alternative methods are usually chosen for determining the step length. These include backtracking line search, the Barzilai-Borwein method, or using a fixed step size [159–163]. Other approaches include using variations of gradient descent such as stochastic gradient descent or the momentum method [164].

4.4.2 Conjugate Gradient Methods

The conjugate gradient (CG) method is a popular gradient-based optimization algorithm designed for solving problems of the form

$$Af = m \tag{4.15}$$

for cases when A is symmetric positive-definite [165]. It can be shown that this is equivalent to finding the unique minimizer of the quadratic function

$$L(f) = \frac{1}{2} f^T A f - m^T f. \tag{4.16}$$

Let $A \in \mathbb{R}^{N \times N}$ be symmetric positive-definite. The vectors $d^i, d^j \in \mathbb{R}^N$ are called A-orthogonal if

$$(d^i)^T A d^j = 0. \tag{4.17}$$

The CG method is based on constructing a set of A-orthogonal search directions $\{d^0, d^1, \ldots, d^{N-1}\}$ such that the iteration scheme to solve (4.15) will theoretically converge in exactly N steps using the iterations[1]

$$f^{\nu+1} = f^\nu + \gamma^\nu d^\nu. \tag{4.18}$$

The construction of the CG iteration scheme is explained in a highly understandable manner in [166], and the algorithm implementation is summarized in Algorithm 4.3.

[1]By convention, the symbol α is commonly used for the step length in CG. However, to avoid confusion with the regularization parameter, the symbol γ is used in this work for the CG step length.

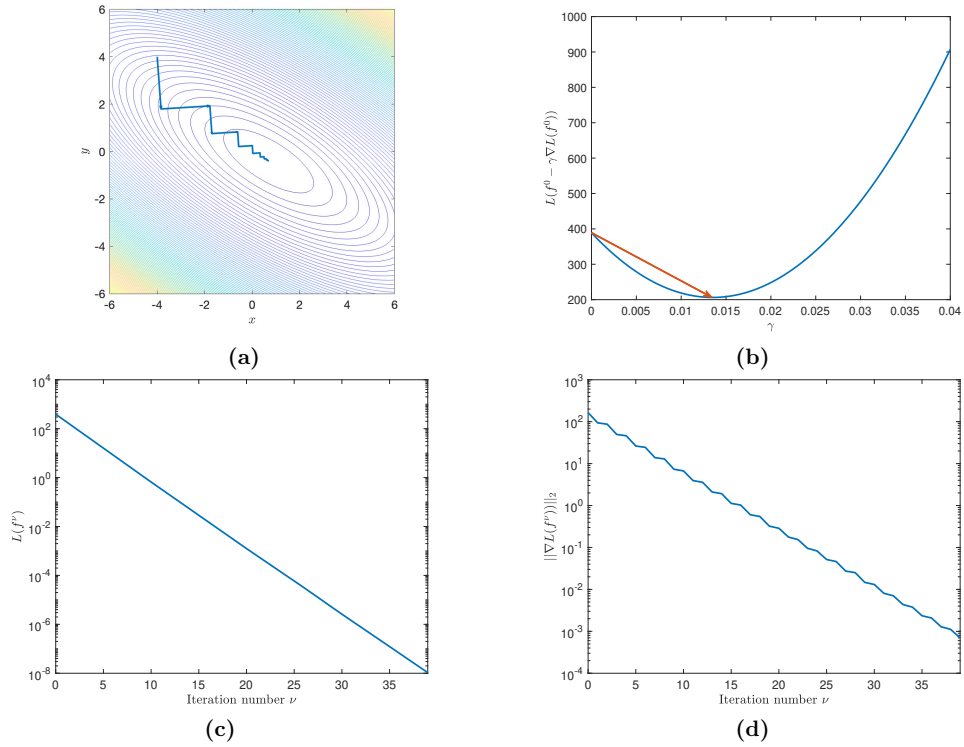


Fig. 4.1. Convergence of the example problem (4.13) - (4.14) using the steepest descent algorithm. (a) SD steps. (b) Univariate line search function for iteration $\nu = 1$. (c) Values of the cost function $L(f^\nu)$ at each iteration. (d) Values of the cost function gradient magnitude $||\nabla L(f^\nu)||_2$ at each iteration.

Algorithm 4.3 Conjugate Gradient

Input: linear map A, data m, starting point f^0, maximum number of iterations $\nu_{\max}$, tolerance ε
Initialize: $r^0 = m - Af^0$, $d^0 = r^0$
for $\nu \leftarrow 1, \ldots, \nu_{\max}$ **do**
  $\gamma^{\nu-1} = \frac{(r^{\nu-1})^T r^{\nu-1}}{(d^{\nu-1})^T A d^{\nu-1}}$
  $f^\nu = f^{\nu-1} + \gamma^{\nu-1} d^{\nu-1}$
  $r^\nu = r^{\nu-1} - \gamma^{\nu-1} A d^{\nu-1}$
  $\beta^\nu = \frac{(r^\nu)^T r^\nu}{(r^{\nu-1})^T r^{\nu-1}}$
  $d^\nu = r^\nu + \beta^\nu d^{\nu-1}$
  if stopping criterion quantity $< \varepsilon$ **then**
    return f^ν
  end if
end for
return f^ν

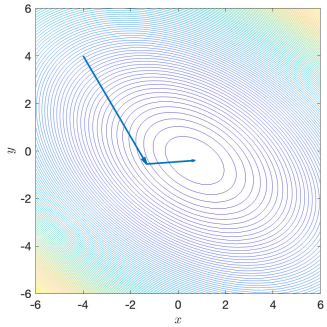


Fig. 4.2. CG steps for the example problem using the values in Eq. (4.14).

Figure 4.2 shows the CG steps using the values in Eq. (4.14). After 2 iterations, the algorithm has converged to very high numerical accuracy, with $||Af^2 - m||_2 \approx 2 \cdot 10^{-15}$ when run on a computer test. It is important to note that although $L(f)$ is different in the SD and CG examples, they share the same minimizer. Clearly, using conjugate gradient yields great advantages in convergence speed when compared to steepest descent. The severity of the zig-zag behaviour is affected both by the condition number of A and by how lucky one is with the starting point f_0.

Minimizing Eq. (4.13) is equivalent to solving the *normal equations*

$$A^T A f = A^T m. \tag{4.19}$$

However, for most ill-posed inverse problems, $A^T A$ may not be symmetric positive-definite, ruling out CG, and even if it is, it can suffer from a very high condition number. In such cases, iterative solvers can converge towards meaningless noise. If a quadratic regularization term is added, *i.e.* we use the form in Eq. (2.20) with $p = 2$, the problem may be formulated in so-called *stacked form*, which allows using CG to solve the regularized problem [42]. Unfortunately, this still rules out non-quadratic regularization strategies, such as the (smoothed) total variation scheme introduced in Chapter 3.

Nonlinear conjugate gradient methods generalize the CG method to unconstrained nonlinear optimization problems of the form in Eq. (4.10) when $L : \mathbb{R}^N \mapsto \mathbb{R}$ is a continuously differentiable function bounded from below [151, 166, 167]. Nonlinear CG uses the iterations in Eq. (4.18), but the step length γ^ν needs to be determined using a line search algorithm, and the step directions are generated using

$$d^\nu = r^\nu + \beta^\nu d^{\nu-1}, \tag{4.20}$$

where

$$r^\nu = -\nabla L(f^\nu), \tag{4.21}$$

and various methods exist for determining the CG update parameter β^ν. Initialization is made using $d^0 = r^0 = -\nabla L(f^0)$. In Publication I of this work, 2D image reconstruction in multienergy CT comparing multiple different smooth regularizers was implemented using the nonlinear conjugate gradient method with golden section line search for choosing the step lengths. For choosing β^ν, this work used the Polak-Ribière formula

$$\beta^\nu_{\mathrm{PR}} = \max\left\{ \frac{(r^\nu)^T(r^\nu - r^{\nu-1})}{(r^{\nu-1})^T r^{\nu-1}}, 0 \right\}, \tag{4.22}$$

which is also known as the Polak-Ribière-Polyak formula B^ν_{PRP} [168, 169]. When the parameter has a value of zero, the iteration becomes a steepest descent step. Pseudocode for this optimization scheme is shown in Algorithm 4.4.

Algorithm 4.4 Polak-Ribière-Polyak Nonlinear Conjugate Gradient

Input: function handles $L(f)$ and $\nabla L(f)$, starting point f^0, maximum number of iterations $\nu_{\max}$, tolerance ε
Initialize: $r^0 = -\nabla L(f^0)$, $d^0 = r^0$
for $\nu \leftarrow 1, \ldots, \nu_{\max}$ **do**
  $\gamma^{\nu-1} \leftarrow \underset{\gamma \in \mathbb{R}_+}{\arg\min}\, L(f^{\nu-1} + \gamma d^{\nu-1})$
  $f^\nu = f^{\nu-1} + \gamma^{\nu-1} d^{\nu-1}$
  $r^\nu = -\nabla L(f^\nu)$
  $\beta^\nu_{\mathrm{PR}} = \max\left\{ \frac{(r^\nu)^T(r^\nu - r^{\nu-1})}{(r^{\nu-1})^T r^{\nu-1}}, 0 \right\}$
  $d^\nu = r^\nu + \beta^\nu d^{\nu-1}$
  if stopping criterion quantity $< \varepsilon$ **then**
    return f^ν
  end if
end for
return f^ν

4.5 Proximal Algorithms

Proximal algorithms are a family of convex optimization algorithms suitable for solving constrained optimization problems where the objective function $L(f)$ is non-differentiable [170–172]. Proximal algorithms operate at a higher level of abstraction than classical methods such as the gradient-based methods discussed in Section 4.4, and the mathematical machinery involved is considerably more complicated. On the other hand, proximal algorithms allow for a much larger range of problems, and their performance can be highly

efficient. Proximal algorithms are called such because the *proximal operator*, also known as the *proximity operator*, plays a pivotal role.

In order to define the proximal operator, the definition of a *closed proper convex function* is first needed.

Definition 4.3 *Let $L : \mathbb{R}^N \mapsto \mathbb{R} \cup \{\infty\}$. $L(f)$ is said to be a closed proper convex function if its epigraph*

$$\operatorname{epi} L = \left\{(f, t) \in \mathbb{R}^N \times \mathbb{R} \mid L(f) \leq t\right\}$$

is a nonempty closed convex set.

A closed proper convex function is always lower semi-continuous. We now proceed to the proximal operator.

Definition 4.4 *Let $L : \mathbb{R}^N \mapsto \mathbb{R} \cup \{\infty\}$ be a closed proper convex function. The proximal operator $\operatorname{prox}_L : \mathbb{R}^N \mapsto \mathbb{R}^N$ of L is defined as*

$$\operatorname{prox}_L(u) = \underset{f \in \mathbb{R}^N}{\arg\min} \left(L(f) + \frac{1}{2}||f - u||_2^2 \right).$$

The proximal operator can be seen as a generalization of projection onto a convex set. The optimization problem in Definition 4.4 has closed form solutions for many functions of interest that appear in numerical optimization, making the proximal operator straightforward to implement in such cases.

In many convex optimization problems the objective function can be expressed as

$$L(f) = L_1(f) + L_2(f) + \ldots + L_K(f), \tag{4.23}$$

consisting of K individual convex cost functions $L_i(f)$. Proximal algorithms operate by *splitting* $L(f)$ into parts, so that each individual term $L_i(f)$ is handled separately, and each non-differentiable (nonsmooth) term is involved via its proximal operator. Thus, a complex objective function is made workable by splitting it into smaller, manageable parts. This type of divide-and-conquer approach is often used in the computational sciences. The proximal operator of $L_i(f)$ can also be interpreted as a kind of gradient step, implying a connection to gradient-based optimization.

4.5.1 Primal-Dual Fixed Point Algorithm

The primal-dual fixed point (PDFP) algorithm by Chen *et al.* is a proximal algorithm well suited for solving convex, constrained optimization problems using variational regularization [173]. It is designed for solving the minimization problem

$$\min_{f \in \mathbb{R}^N} L_1(f) + (L_2 \circ B)(f) + L_3(f), \tag{4.24}$$

where L_1, L_2, and L_3 are three proper lower semi-continuous convex functions, L_1 is differentiable on $\mathbb{R}^N$ with a $1/\beta$-Lipschitz continuous gradient for some $\beta \in (0, \infty]$, and B is a linear transformation from $\mathbb{R}^N$ to a Euclidean space. The PDFP algorithm can be used for mathematically robust total variation regularized image reconstruction with nonnegativity constraints. The mathematical steps of adapting the PDFP algorithm for this purpose are detailed here, and specific applications to computed tomography with pseudocode and results are shown in Section 5.5 and in Publication IV.

Now, let $L_1(f)$ be the data fidelity term

$$L_1(f) = \frac{1}{2}||Af - m||_2^2, \tag{4.25}$$

let B be the discrete gradient matrix D as defined in Section 3.4, and let

$$(L_2 \circ B)(f) = \alpha||Df||_p, \tag{4.26}$$

where $\alpha \in \mathbb{R}_+$ is the regularization parameter, and

$$||\cdot||_p = ||\cdot||_1 \tag{4.27}$$

for anisotropic total variation, and

$$||\cdot||_p = ||\cdot||_{2,1} \tag{4.28}$$

for isotropic total variation. Finally, let $L_3(f)$ be the indicator function

$$\chi_{\mathbb{R}_+^N}(f) = \begin{cases} 0, & f \in \mathbb{R}_+^N, \\ \infty, & f \notin \mathbb{R}_+^N, \end{cases} \tag{4.29}$$

which constrains the possible solutions to the nonnegative orthant of $\mathbb{R}^N$. The optimization problem in Eq. (4.24) now becomes

$$\min_{f \in \mathbb{R}_+^N} \frac{1}{2}||Af - m||_2^2 + \alpha||Df||_p, \tag{4.30}$$

analogous to Eq. (2.20).

The PDFP algorithm which solves Eq. (4.24) has the formulation:

$$\text{(PDFP)} \begin{cases} g^{\nu+1} &= \text{prox}_{\gamma L_3}(f^\nu - \gamma \nabla L_1(f^\nu) - \lambda B^T v^\nu), \\ v^{\nu+1} &= (I - \text{prox}_{\frac{\gamma}{\lambda} L_2})(Bg^{\nu+1} + v^\nu), \\ f^{\nu+1} &= \text{prox}_{\gamma L_3}(f^\nu - \gamma \nabla L_1(f^\nu) - \lambda B^T v^{\nu+1}), \end{cases} \tag{4.31}$$

where $0 < \lambda < 1/\lambda_{\max}(BB^T)$ and $0 < \gamma < 2\beta$.

The gradient of the data fidelity term is $\nabla L_1(f) = A^T(Af - m)$. When $L_3(f) = \chi_{\mathbb{R}_+^N}(f)$, then $\text{prox}_{\gamma L_3} = \text{proj}_{\mathbb{R}_+^N}$, which is the component-wise projection of f onto the nonnegative orthant of $\mathbb{R}^N$:

$$\text{proj}_{\mathbb{R}_+^N}(f) = \begin{cases} f_l = f_l, & f_l \geq 0, \\ f_l = 0, & f_l < 0, \end{cases} \quad \forall\, f_l \in f. \tag{4.32}$$

The algorithm now becomes

$$\text{(PDFP)} \begin{cases} g^{\nu+1} &= \text{proj}_{\mathbb{R}_+^N}(f^\nu - \gamma A^T(Af^\nu - m) - \lambda D^T v^\nu), \\ v^{\nu+1} &= (I - \text{prox}_{\frac{\gamma}{\lambda} L_2})(Dg^{\nu+1} + v^\nu), \\ f^{\nu+1} &= \text{proj}_{\mathbb{R}_+^N}(f^\nu - \gamma A^T(Af^\nu - m) - \lambda D^T v^{\nu+1}). \end{cases} \tag{4.33}$$

For anisotropic total variation, where $L_2(f) = \alpha||f||_1$, the proximal operator of $\frac{\gamma}{\lambda} L_2(f)$ becomes the *soft thresholding operator*:

$$\text{prox}_{\frac{\gamma}{\lambda}\alpha||\cdot||_1}(f) = S_{\frac{\gamma}{\lambda}\alpha}(f) \tag{4.34}$$

$$= \text{sgn}(f) \cdot \max\left\{|f| - \frac{\gamma}{\lambda}\alpha,\, 0\right\}, \tag{4.35}$$

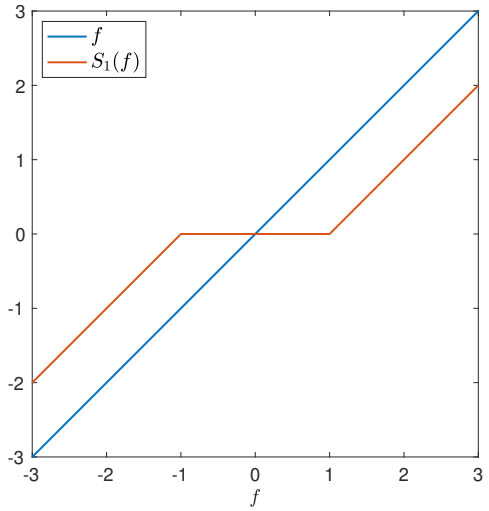


Fig. 4.3. The soft thresholding function S_τ, $\tau \geq 0$ for $\tau = 1$.

where the absolute value, the sign function, and the max operator are understood componentwise. The nature of the soft thresholding operator is demonstrated in Figure 4.3: values that are closer to the origin than the cutoff level τ are set to zero, and values above that have their magnitude reduced by the fixed amount τ.

It is the proximal operator $\text{prox}_{\frac{\gamma}{\lambda}\alpha||\cdot||_1}$ that induces gradient sparsity in the solution, as on each iteration ν a number of coefficients are set to zero. In anistropic TV, this thresholding is applied directly to the gradient coefficients. In many locations of f, only the largest gradient coefficient will remain nonzero after the thresholding, resulting in reconstructions that favor changes which are aligned with the main coordinate axes of the Cartesian system. This mostly unfavourable effect increases when α is increased and the regularization becomes stronger.

The proximal operator for isotropic total variation, where $L_2(f) = \alpha||f||_{2,1}$, is somewhat more complicated, as it is necessary to track both the individual gradient components as well as the gradient magnitudes. A discretization f with N elements will also have exactly N gradient magnitudes associated with it, one for each element, and the number of gradient components for f will be nN, where n is the dimensionality of the underlying problem (2 or 3 in the context of this work). Let $l \in 1, \ldots, nN$, and let $(Df)_l$ be the lth element in the discrete gradient data. Every such element will be associated with some element in f, and every such element will also have a gradient magnitude $||(Df)_l||_2$ associated with it, computed with the standard Euclidean distance (ℓ^2 norm) using all of the individual gradient components associated with the corresponding element of f. The proximal operator for the $L_2(f)$ term can then be written as

$$\text{prox}_{\frac{\gamma}{\lambda}\alpha||\cdot||_{2,1}}(f) = \frac{f}{||f||_2} \cdot \max\left\{||f||_2 - \frac{\gamma}{\lambda}\alpha,\, 0\right\}, \tag{4.36}$$

where all operations should be understood component-wise, keeping track of the how the gradient components in Df and gradient magnitudes $||Df||_2$ are associated with each other, and where the convention $\frac{0}{0} \cdot 0 = 0$ is used [174–176].

In isotropic TV, any element of f where the associated gradient magnitude is below the threshold level $\frac{\gamma}{\lambda}\alpha$ has *all* of its gradient components set to zero, and when the gradient

magnitude is above the threshold, each normalized gradient component is scaled down by the same factor. This property enforces reconstructions where the gradient for many elements of f will be zero, yet all spatial directions are favoured equally.

5 Image Reconstruction in Computed Tomography

5.1 Projections and X-ray Measurements

The purpose of computed tomography is to recover the internal structure of an object from its X-ray projections, facilitating its detailed study. The very expression X-ray projection implies that there is a deep connection between X-ray images and mathematical projection operations. Developing image reconstruction methods for computed tomography therefore begins with formulating a mathematical and physical model for such X-ray measurements. It needs to be emphasized that the expression *image reconstruction* is used for both 2D and 3D CT reconstructions, and the actual dimensionality of the reconstruction task can be inferred from the context. In the same vein, the word *image* is used in such contexts as image quality, image noise, etc. for both 2D and 3D computed tomography.

5.1.1 The Radon Transform

The Radon transform is an integral transform that forms the mathematical basis of CT imaging [177, 178]. It connects a function $\mathcal{F} : \mathbb{R}^2 \mapsto \mathbb{R}$ to a set of *line integrals* taken along straight lines across the function.

Let $\mathcal{F} : \mathbb{R}^2 \mapsto \mathbb{R}$ be a square-integrable function with compact support. The integration path along a straight line can be specified with the parameters (θ, t), so that another straight line passing through the origin and forming the angle θ with the x-axis intersects in a right angle with the integration line at the point t. Such an integration path obeys the equation

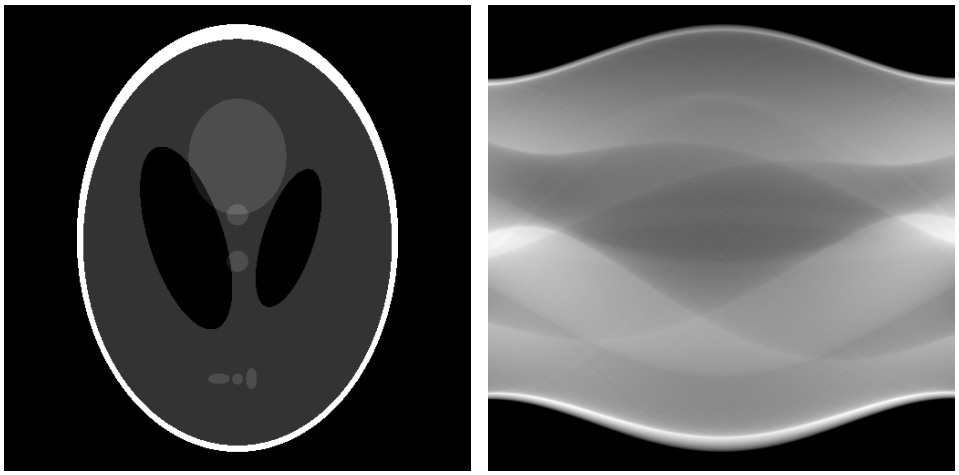
$$x \cos\theta + y \sin\theta = t. \tag{5.1}$$

The set of all line integrals over $\mathcal{F}(x, y)$ at the angle θ is now given by

$$\mathcal{P}_\theta(t) = \int\limits_{(\theta,t)\text{ line}} \mathcal{F}(x, y)ds. \tag{5.2}$$

The full collection of line integrals as defined in Eq. (5.2) and collected across $\theta \in [0, \pi)$ is known as the *Radon transform* of $\mathcal{F}(x, y)$. This type of projection geometry is known as *parallel beam geometry*, as all the integration paths for a fixed value of θ are parallel lines.

The Radon transform data is also commonly called a *sinogram*, as the projections of a single point-like structure in $\mathcal{F}(x, y)$ form a sinusoidal curve in the Radon transform. Figure 5.1(a) shows a test image f known as the Shepp-Logan phantom [179], widely used in testing CT image reconstruction algorithms, and Figure 5.1(b) shows its sinogram. It should be emphasized that the images have been created computationally, and therefore the continuum model $\mathcal{F}(x, y)$ has been discretized as f, here using a 512×512 pixel grid,



(a) **(b)**

Fig. 5.1. The Shepp-Logan phantom and its Radon transform.

and the continuous interval $[0, \pi)$ has been discretized into discrete angles, in this case 512 evenly spaced angles.

5.1.2 X-ray Projections

X-rays are part of the electromagnetic spectrum. They consist of massless elementary particles called photons, with energies ranging from 100 eV to hundreds of kiloelectron volts (keV). The energy range used in CT imaging is approximately 20-150 keV. As X-ray quanta travel through matter, photons are absorbed or scattered away from the ray path. The most important photon-matter interaction processes in the imaging energy range are the *photoelectric effect* and inelastic *Compton scattering*.

The intensity loss of a narrow X-ray beam as it passes through matter is given by the *Beer-Lambert law*. Consider such a beam of monoenergetic X-ray photons with initial intensity I_0 passing through a homogenous slab of some material of thickness x with the *linear attenuation coefficient* μ. The linear attenuation coefficient contains the sum effect of all the photon-matter interaction processes in the matter. The X-ray intensity leaving the slab will then be

$$I = I_0 e^{-\mu x}. \tag{5.3}$$

Homogenous slabs rarely make interesting objects for imaging. For a cross-sectional xy-slice in a non-homogenous medium the attenuated intensity will be

$$I = I_0 e^{-\int_{\text{ray}} \mu(x,y) ds}. \tag{5.4}$$

This can be rearranged into

$$-\ln \frac{I}{I_0} = \int_{\text{ray}} \mu(x, y) ds. \tag{5.5}$$

Eq. (5.5) is known as the *log-transform* of the X-ray intensity measurement I. Comparison of Eqs. (5.2) and (5.5) reveals that the 2D distribution of linear attenuation coefficients $\mu(x, y)$ now becomes the quantity of interest $\mathcal{F}(x, y)$ in CT imaging, and the log-transform

of X-ray measurements provides a way to physically measure line integrals as defined in the Radon transform.

Now, consider a one-dimensional X-ray detector strip consisting of P detector elements, *i.e.* detector pixels. We now define the *CT measurement* $m_{\theta,p}$ at detector pixel $p \in \{1, \ldots, P\}$ taken at projection angle θ as

$$m_{\theta,p} = -\ln \frac{I_{\theta,p}}{I_{\theta,p,0}}, \tag{5.6}$$

where $I_{\theta,p,0}$ is the initial intensity of the X-ray beam and $I_{\theta,p}$ is the X-ray intensity measured by the detector element. Let $\#\theta$ be the number of projection angles in the X-ray measurements. The full set of CT measurements m can then be arranged into a $P \times \#\theta$ grid, which is the sinogram.

The imaging model introduced here contains a number of idealizations. Firstly, all interactions involving X-ray quanta are stochastic processes governed by quantum mechanics. The generation of X-ray photons in an X-ray tube, and photon-matter interactions in both the imaged object and in the X-ray detector all obey Poisson statistics. Therefore the values I and I_0 should be interpreted only as *expectation values*, and the actual measurements will invariably contain fluctuations (*i.e.* noise) arising from photon statistics.

Secondly, the model presented above ignores the effects of photon scatter. A photon that has been scattered away from its original path may end up being detected by another detector element, resulting in degradation of the measurement quality.

Thirdly, the response of X-ray detectors to radiation is always non-ideal, and they need to be calibrated so that the best possible match is made to the mathematical model in Eq. (5.5). A standard approach to X-ray detector calibration is *dark-frame subtraction* combined with *flat-field correction* [180, 181]. Before the log-transform, the intensity value of each pixel is corrected with

$$I_{\text{calibrated}} = \frac{I - I_{\text{dark}}}{I_0 - I_{\text{dark}}} \times G, \tag{5.7}$$

$I_{\text{calibrated}}$ is the calibrated output to be used in the log-transform, I is the raw intensity measurement, I_{dark} is the level of electronic background noise in the pixel, I_0 is the intensity registered when there is no object between the X-ray source and the detector, and G is a scalar determined by the average value of the all the detector outputs.

Finally, the linear attenuation coefficient is actually strongly and nonlinearly dependent on photon energy E, and can be written $\mu(E)$. The X-ray spectra produced by most laboratory X-ray sources is polyenergetic, and therefore Beer-Lambert's law should be written as

$$I = \int_0^{E_{\max}} \left\{ e^{-\int_{\text{ray}} \mu(x,y,E)ds} \right\} dE. \tag{5.8}$$

This can lead to significant nonlinearities in the measurement model, and to prominent *image artifacts*, *i.e.* image distortions, in CT reconstructions, due to an effect known as *beam hardening* [182]. In fact, if we are also measuring energy integrals over the linear attenuation coefficients and not just integrals over their spatial distribution, one might ask the question: what are we even trying to reconstruct? Most of the time such questions are swept under the rug, the measurements are assumed to be made at some *effective energy* E_{eff} determined by the X-ray spectrum, a linear measurement model is approximated, and possible beam hardening artifacts caused by polyenergetic X-rays are corrected using various *ad hoc* means [183–185].

5.1.3 Multispectral X-ray Imaging

Spectral computed tomography makes explicit use of the fact that X-ray attenuation coefficients are strongly dependent on photon energy. In many practical cases, it may occur that two or more materials of interest end up having the same grayscale values in a conventional CT reconstruction. In such cases, the object can be scanned using two or more X-ray energies[1] or spectra. Combining the multispectral measurements with *a priori* knowledge on the energy-dependent attenuation properties of the materials present in the scanned object make it possible to compute a *material decomposition* into different material bases [80–84]. For example, in a clinical setting the basis materials could be soft tissue, bone, and iodine contrast agent.

Many technical approaches to measuring multispectral CT data exist. These include sequential scans made using different X-ray spectra, fast alternation of the X-ray spectra during one scan, achieved by modulating the voltage in the X-ray tube producing the radiation, and multi-layered X-ray detectors.

In recent years, the main focus of spectral CT has been on *photon-counting detector* computed tomography (PCD-CT) [22–25]. Photon-counting detectors use a process called *direct conversion* to make energy-sensitive measurements of individual X-ray photons. This is in contrast to *energy-integrating detectors* (EIDs) used in conventional CT, where the X-ray photon flux is first converted into visible light which is then measured, losing all energy information on the photons. PCDs use a semiconductor conversion material to convert each X-ray photon into a cloud of charge carriers [186]. The most commonly used conversion materials are cadmium telluride (CdTe), cadmium zinc telluride (CZT), and silicon (Si). A bias voltage is applied to the semiconductor, and the charge cloud is collected and analyzed using an application specific integrated circuit (ASIC). The resulting electronic pulses caused by individual detected X-rays can be categorized using pulse-height analysis, making it possible to divide the measurements across the X-ray spectrum into multiple energy windows called bins. In principle, the number of PCD energy bins can be arbitrarily high. In practice, this is limited by semiconductor physics considerations and by the processing and readout speed of the detector electronics. Another consideration is that as the number of energy bins is increased, the photon noise statistics in each bin become worse as fewer photons are counted. The first medical PCD-CT scanners have recently been taken into clinical use [187], and dedicated PCD-CT reconstruction algorithms are an active research topic [131, 134, 135, 188–192].

Regardless of how the spectral CT measurements are made, the multispectral data is divided into Q energy channels, and the definition of the CT measurement needs to be modified as

$$m_{\theta,p,q} = -\ln \frac{I_{\theta,p,q}}{I_{\theta,p,q,0}}, \tag{5.9}$$

where $q \in \{1, \ldots, Q\}$ indicates the energy channel, as discussed in Section 2.5.3. It should be mentioned that Eqs. (5.6) and (5.9) are not the only possible ways to define the CT measurement data points. For example, some approaches use experimentally defined calibration coefficients so that the measured intensity data I is directly converted into effective thickness in some material basis [193, 194].

Publications I and II of this work focus on image reconstruction in 2D multispectral computed tomography.

[1] Monochromatic imaging using a single X-ray energy is practically achievable only by synchrotron.

5.2 Analytical Reconstruction

Analytical reconstruction methods are called such because they are based on applying integration and filtering operations on the continuous Radon transform presented in Section 5.1.1. As such, they entirely sidestep inverting a discretized linear algebra model such as presented in Eq. (2.2), although the necessary operations still need to be discretized appropriately to work with discrete data points in a computational setting.

Analytical algorithms have been the standard reconstruction method in CT for decades, and there is no reason to assume that their use will end in the future. They are briefly discussed here to provide a motivation and a reference point for the development of further refined reconstruction algorithms based on the mathematics of inverse problems. Detailed descriptions of analytical reconstruction methods can be found in standard works such as [178, 177], and the 2D algorithms presented here were derived in detail in the previous work [195].

5.2.1 Filtered Backprojection Reconstruction

Filtered backprojection (FBP) is the analytical 2D reconstruction method that inverts the Radon transform. Both the transform and its inversion were introduced by Johann Radon in his seminal 1917 article [1, 2].

FBP is based on the *Fourier slice theorem*, also known as the *projection-slice theorem* and the *central slice theorem*.

Theorem 5.1 (Fourier Slice Theorem) *Let $\widehat{\mathcal{P}}_\theta(w)$ be the one-dimensional Fourier transform of the projection $\mathcal{P}_\theta(t)$, and let $\widehat{\mathcal{F}}(u,v)$ be the two-dimensional Fourier transform of $\mathcal{F}(x,y)$. Then*

$$\widehat{\mathcal{P}}_\theta(w) = \widehat{\mathcal{F}}(w\cos\theta, w\sin\theta).$$

In other words, the Fourier transform of a projection $\mathcal{P}_\theta(t)$ is equal to the values of the two-dimensional Fourier transform of $\mathcal{F}(x,y)$, evaluated along a radial line (*i.e. slice*) forming the angle θ with the horizontal axis. The Fourier slice theorem is quite remarkable in that it yields information on the two-dimensional function $\widehat{\mathcal{F}}(u,v)$ based on purely one-dimensional projection data.

The FBP reconstruction method for parallel beam geometry is relatively straightforward to derive in the continuum setting by rewriting the inverse Fourier transform of $\widehat{\mathcal{F}}(u,v)$ in polar coordinates and rearranging the integrals. It is equally easy to understand FBP heuristically. The Fourier slice theorem states that the line integral projections $\mathcal{P}_\theta(t)$ give access to radial slices of $\widehat{\mathcal{F}}(u,v)$. The density of these radial slices increases towards the origin in the Fourier domain. The slices are therefore multiplied with a weighting function $|w|$ that balances out this uneven density. This can be understood as a *filter* operation in the Fourier domain. The filtered projections are then projected back along the original integration line paths. If this backprojection operation is done without the filtering, the result is something that loosely resembles the original object f but is extremely blurred. This is known as *unfiltered backprojection* or *simple backprojection*.

Mathematically, we define the filtered projection as

$$\mathcal{Q}_\theta(t) = \int_{-\infty}^{\infty} \widehat{\mathcal{P}}_\theta(w)|w|e^{i2\pi wt}dw, \tag{5.10}$$

where i is the imaginary unit. The FBP reconstruction is now given by

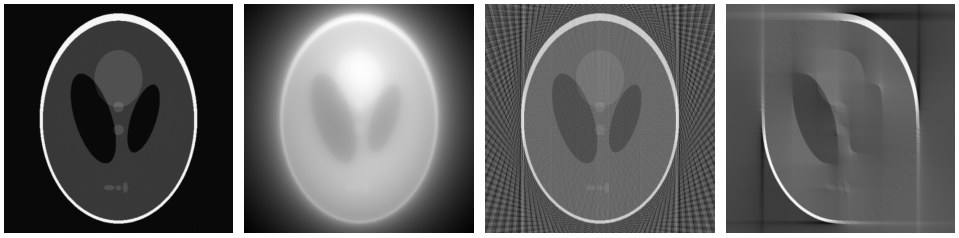
$$\mathcal{F}(x,y) = \int_0^\pi \mathcal{Q}_\theta(x\cos\theta + y\sin\theta)d\theta. \tag{5.11}$$

In a computational setting, Eqs. (5.10) and (5.11) need to be discretized, and the integrals become sums over finite sets. In some practical cases, the ramp filter $|w|$ can be replaced with a different function. Although this does not lead to a mathematically exact reconstruction formulation, the outcome may be more beneficial in term of noise or spatial resolution in an applied setting.

FBP has many strengths as reconstruction algorithm. It is very well understood, fast to compute, and parallelizable. Modern graphics processing units (GPUs) are extremely well suited to computing projection and backprojection operations. It also has two major shortcomings. The first one is that due to the linearity of FBP, any noise in the measurement data will be directly transferred to the reconstruction. The second one is that the integral over $\theta \in [0, \pi)$ needs to be approximated with discrete projections obtained with dense angular sampling. If the projection angle intervals are too large (*sparse angle tomography)* or they do not cover the full angular range $\theta \in [0, \pi)$ (*limited angle tomography*) [196], FBP results in subpar reconstruction quality.

Figure 5.2 shows different types of analytical reconstructions based on the sinogram in Figure 5.1. The full angle FBP reconstruction results in near perfect recovery of the computational phantom. The full angle unfiltered backprojection results in a blurred reconstruction. When the sinogram is subsampled to produce sparse angle or limited angle data, the reconstruction quality is greatly impaired. As both sparse angle tomography and limited angle tomography are of interest in various imaging setups [197–208], further refined reconstruction techniques are clearly needed.

Publications I and II of this work investigated 2D CT reconstruction using sparse angular sampling, and limited angle tomography was the focus of Publication III.



(a) **(b)** **(c)** **(d)**

Fig. 5.2. Different reconstructions based on the sinogram in Figure 5.1. (a) Full angle FBP. (b) Full angle unfiltered backprojection. (c) Sparse angle FBP using every tenth projection. (d) Limited angle FBP using the angular range $\theta \in [0, \pi/2)$. Each image was linearly windowed between its minimum and maximum values.

5.2.2 Reconstruction in Fan-beam Geometry

Parallel beam geometry is challenging to achieve in practical settings, and most laboratory and clinical scanners use divergent X-rays that are emitted from a small focal spot. In a 2D setting this is known as *fan beam* geometry. There are two types of fan beam geometry.

Fan beam geometry with *equiangular rays* uses curved detectors, and the line integrals across $\mathcal{F}(x, y)$ are sampled with a constant angular interval between the rays. When using *equispaced intervals* on a flat detector strip the the angular between the line integrals is not constant. Publications I, II, and III of this work use latter geometry for real X-ray data.

The filtered back-projection reconstruction method presented in Section 5.2.1 can be adapted to both types of fan beam geometry. The adaptation is based on the fact that there is a one-to-one correspondence between each integration path in the fan beam geometry and a given (θ, t) integration path in parallel beam geometry. In both cases, the projection direction is defined using the *central ray* of the fan beam. In fan beam reconstruction, the angular range $[0, \theta)$ for projection directions does not provide a full set of line integrals over $\mathcal{F}(x, y)$, and the fan beam FBP reconstruction methods are defined so that the angular interval $[0, 2\pi)$ is used for the projection directions.

For a thorough mathematical treatment of FBP reconstruction using flat detectors with equispaced detectors elements, the previous work [195] is referred to.

5.2.3 Cone Beam Reconstruction

Cone beam computed tomography (CBCT) is a natural extension of fan beam geometry into 3D. In CBCT, the X-ray projection data is measured using a cone-shaped X-ray beam and a 2D X-ray detector panel moving a single rotation in a plane, forming a circular orbit around the imaged object. CBCT scanners are cheaper and less cumbersome than helical CT scanners, in which the X-ray source and detector rotate along a helical trajectory around the the imaged object. CBCT imaging has been widely applied in dentistry, imaging of the head and extremities, image-guided radiotherapy, and microtomography [18, 19, 209, 210]. The proximal reconstruction algorithms discussed in Section 5.5 and in Publication IV focus on image reconstruction in CBCT.

In practical solutions, either the source-detector pair or the sample can rotate. Generally, the former setup is used in clinical imaging, whereas the latter is used microtomography and other laboratory setups.

In CBCT, as in any 3D CT modality, we are interested in recovering the full three-dimensional spatial distribution of attenuation coefficients $\mu(x, y, z)$ in the imaged object. In building the CBCT measurement model, Eq. (5.5) needs to be modified. The log-transformed X-ray intensity measurement now becomes

$$-\ln \frac{I}{I_0} = \int_{\text{ray}} \mu(x, y, z) ds. \tag{5.12}$$

The X-ray detector is now a panel consisting of P_1 rows and P_2 columns of pixels, and the total number of detector pixels is $P = P_1 P_2$. In order to maintain consistent notation with Section 5.1.2, the pixels are labeled with a one-dimensional columnwise indexing $p \in \{1, \ldots, P\}$, analogous to Figure 2.3. The CBCT measurement points $m_{\theta,p}$ can then also be defined using Eq. (5.6).

The analytical CT reconstruction algorithm developed in 1984 by Feldkamp, Davis, and Kress, known as the FDK algorithm, has been the standard CBCT reconstruction algorithm for decades [211]. FDK is an approximate 3D extension of the FBP reconstruction method. In the central plane of the cone beam, *i.e.* the plane perpendicular to the axis of rotation, the FDK algorithm reduces to fan beam reconstruction, and it is exact. In the other reconstruction planes, the standard CBCT data acquisition scheme does not provide the full Radon transform data. This is known as failure to meet the *Tuy-Smith* sufficiency condition [212, 213], which will result in *cone beam artifacts* in the final 3D

reconstruction [18, 214]. The severity of these artifacts is strongly influenced by the cone angle in the z-direction, and they increase as one moves further away from the central plane. Nevertheless, the FDK algorithm shares many of the attractive properties of FBP, such as speed, parallelizability, and also high accuracy, if the cone effect is kept sufficienty small.

5.3 Iterative Reconstruction

Iterative reconstruction (IR) is an image reconstruction paradigm whereby the final CT image is obtained via a cyclic process. An initial estimate of the CT reconstruction is generated, synthetic X-ray projections are produced from this estimate and compared to the measured X-ray data, and the reconstruction is updated based on the data discrepancy [30, 31]. This process is repeated until a satisfactory result has been obtained.

The earliest CT reconstructions in the 1970s were computed using iterative methods. Due to computational considerations, they fell out of vogue for decades until a resurgence in the 2010s, and many CT scanners now use iterative reconstruction as a standard part of clinical imaging. Iterative methods are appealing in many settings, as they have been shown to reduce image noise and artifacts in CT reconstructions.

5.3.1 Discretized CT Imaging Model

Iterative methods take into account the discretized nature of the computational CT reconstruction task in the fundamental definition of the reconstruction problem. This is in contrast to analytical reconstruction, formulated in terms of continuous variables and functions which must be discretized for practical implementation *post hoc*.

Starting with the 2D case, let $f \in \mathbb{R}^N$ represent the CT image, using the flattened indexing defined in Section 2.5.1. The total number of measurement points in a CT scan is

$$M = \#\theta \cdot P, \tag{5.13}$$

where $\#\theta$ is the number of projections and P is the number of pixels in the X-ray detector. We define the index variable $r \in \{1, \dots, M\}$ such that there is a one-to-one correspondence between the (θ, p)-indexing of the measurement data as defined in Section 5.1.2 and the r-indexing, so that each CT measurement point $m_{\theta,p}$ can be represented as the element m_r in the measurement data vector $m \in \mathbb{R}^M$.

In the context of iterative reconstruction, an X-ray is defined as the radiation travelling within a strip from the X-ray source to detector pixel p at projection angle θ. This X-ray can be given the same index r as for the measurement point m_r, based on the (θ, p)-to-r index mapping. Next, we define the discrete variable

$$a_{r,l} = \frac{\text{ovelapping area of } r\text{th X-ray and } f_l}{\text{area of } f_l}. \tag{5.14}$$

The coefficient $a_{r,l} \in [0, 1]$ should be understood as a weight for the amount of attenuation accumulated as the path of the X-ray overlaps with pixel f_l. The line integral across $\mu(x, y)$, as given in Eq. (5.5), can now be approximated using a *ray-sum* over f, so that the CT measurement m_r becomes

$$m_r \approx \sum_{l=1}^{N} a_{r,l} f_l. \tag{5.15}$$

m_r is computed from the X-ray detector measurements using

$$m_r = -\ln \frac{I_r}{I_{r,0}}, \tag{5.16}$$

where I_r is the X-ray intensity measured by the detector element and $I_{r,0}$ is the unattenuated intensity of the X-ray. This principle of modelling CT measurements via ray-sums is visualized in Figure 5.3.

Stacking the expressions for m_r for all $r \in 1, \ldots, M$ leads to the matrix model

$$m = Af, \tag{5.17}$$

where $A \in \mathbb{R}^{M \times N}$. This is equivalent to Eq. (2.2), the fundamental model of linear computational inverse problems, minus the noise and modelling error term ε. For any single X-ray, most of the coefficients $a_{r,l}$ are zero, as only relatively few f_l contribute to the ray-sum, and A therefore becomes a sparse matrix. It is well known that the CT imaging model A is also ill-posed [42], and obtaining a CT reconstruction, *i.e.* solving Eq. (5.17), requires using inverse problems methodology.

The discretized CT imaging model presented above has been defined using flattened index variables, and with the exception of the discrete variable $a_{r,l}$, all the same expressions hold also for the 3D reconstruction case. As for the ray-sums, an X-ray is now defined as the radiation travelling within the volume from the X-ray source to the detector pixel p, using flattened 2D detector pixel indexing as discussed in Section 5.2.3. The ray-sum coefficients are now defined as

$$a_{r,l} = \frac{\text{ovelapping volume of } r\text{th X-ray and } f_l}{\text{volume of } f_l}. \tag{5.18}$$

5.3.2 Model-Based Iterative Reconstruction

The expression *iterative reconstruction* is something of an umbrella term, referring to any iterative process used to obtain the CT reconstruction. The early IR methods algebraic reconstruction technique (ART), simultaneous iterative reconstruction technique (SIRT), and simultaneous algebraic reconstruction technique (SART) directly tackle the system of linear equations $m = Af$ [215–217]. Such techniques are sometimes referred to as *algebraic reconstruction algorithms* in the literature [178].

Contemporary iterative algorithms attempt to model the CT system and the data acquisition process as accurately as possible, and they also include regularization into the reconstruction task to add *a priori* information, leading to the expression *model-based iterative reconstruction* (MBIR). Using the discretized model for CT measurements described in Section 5.3.1 allows the reconstruction problem to be formulated identically to the discretized inverse problem model in Eq. 2.2. This immediately brings into use the machinery for image reconstruction discussed throughout Chapters 2-4. The selection of possible regularizers and optimization algorithms is large, and a great number of regularization-based CT reconstruction algorithms have been introduced in scientific publications. In the context of this work, MBIR algorithms and regularized inversion are treated somewhat synonymously. However, it is important to note that the nomenclature used in the literature varies considerably, depending on the backgrounds of the algorithm designers and authors.

Regardless of the specifics of the algorithm, all regularization-based IR algorithms share some attractive properties. Using the discretized CT imaging model eliminates the

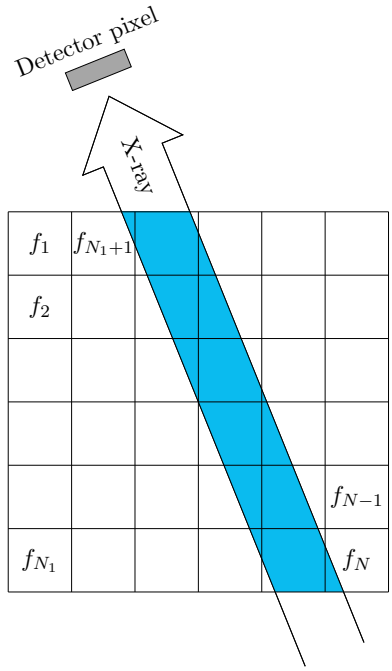


Fig. 5.3. The discretized CT imaging model for one X-ray. The intersecting areas of f and the X-ray are highlighted in cyan. In these regions, the ray-sum coefficients $a_{r,l}$ are nonzero.

need to specifically design different reconstruction algorithms for different imaging geometries, a distinct advantage over analytical reconstruction methods. Indeed, many iterative algorithms operate in a plug-and-play manner, where the algorithm itself is oblivious to the specifics of the forward model A and the data m, facilitating their flexible application to varying imaging settings. Through regularization, IR algorithms can suppress noise and reduce image artifacts arising from the many non-idealities of real CT measurements. This has allowed substantial reduction in patient dose in CT scans, as sufficiently good reconstructions can be obtained from noisier data acquired using lowered X-ray photon fluxes.

If MBIR has potential for improved image quality and reduced radiation dose over FBP, why should one ever wish to use FBP? Firstly, it should be emphasized that in many cases, FBP produces perfectly satisfactory CT images. Furthermore, FBP is in principle a parameter free model, whereas MBIR algorithms require tweaking of parameters which will affect the outcome. Even if the parameters are judiciously chosen, some kinds of assumptions are always made in their choice, be it implicitly or explicitly. FBP is linear, predictable, and well understood, whereas MBIR methods are non-linear, and interpretability of their results may be less intuitive than for FBP.

Finally, FBP-based reconstruction can be made extraordinarily fast compared to MBIR, especially with modern GPUs and parallel computing. To put this into perspective, consider a CBCT scanner with a 500×700 pixel X-ray detector, a data acquisition protocol with 720 projection directions, and a 3D reconstruction volume of $512 \times 512 \times 384$ voxels. Although this is a fictitious example, such numbers are entirely realistic for routine

CBCT scanning. The measurement vector m now contains 252 million data points, and f contains over 100 million voxel values that need to be solved. Compared to the parallelizable filtering + backprojection approach of FBP, using an optimization-based approach to solving the regularized image reconstruction problem will be many times slower, and it will also have higher memory requirements. The choice between using FBP or MBIR should therefore be made judiciously depending on the requirements and limitations of the imaging situation.

5.4 Image Quality, Noise, and Representation

Image quality (IQ) assessment in medical imaging, both in the general case and in the context of CT imaging, is an important research field in its own right, beyond the scope of this work. In general, desirable properties in CT images are accuracy, good contrast, good spatial resolution, few artifacts, and low noise. However, it should be emphasized that what counts as sufficiently good image quality is highly task-specific. For example, differentiating soft tissues with small contrast differences sets very different requirements on the image compared to identifying bony structures.

Many quantitative metrics are used to perform objective image quality analysis. These include signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), mean square error (MSE), and structural similarity index (SSIM), among others [218–220]. However, such metrics do not always fully correlate with what is perceived as good or useful image quality by the human observer interpreting and analyzing images, and *subjective image quality assessment* by experienced experts is often used as a benchmark [221–224]. Subjective IQ assessment is time-consuming and challenging to standardize [225], and mathematical *model observers* which attempt to mimic the human detection of features have been developed for objective IQ analysis [226, 227].

The purpose of this overview is to present in sufficient detail the theoretical and computational framework of CT image reconstruction based on inverse problems methodology. IQ analysis lies beyond this, though it is certainly vital that the image quality improvements (if any) of new CT algorithms are thoroughly scrutinized if their adoption into practical use is to be considered. Therefore, any IQ discussion seen here is subjective analysis by the algorithm designer, whereas in the publications the IQ metrics vary on a case-by-case basis.

Regardless of how image quality is assessed, noise is one of the greatest obstacles of good CT reconstruction quality. The most important noise source is the Poisson-distributed statistical fluctuation of photons, and increasing the number of photons used in a scan in will result in better noise statistics in the measurements. The radiation dose D received by the patient is directly proportional to number of photons used, whereas the signal-to-noise ratio in FBP will obey the proportionality [177]

$$\mathrm{SNR} \propto \sqrt{D}. \tag{5.19}$$

As patient dose is increased, there are diminishing returns in SNR, requiring careful balancing between dose and sufficient image quality. CT is a high dose modality, and the desire to reduce the population radiation dose from scans has been one of the driving forces in developing new CT technology [228]. Nonlinear reconstruction methods such as MBIR and AI-based techniques can escape the SNR-to-dose relationship of Eq. (5.19), which has been one of the motivations for developing new reconstruction algorithms, such as seen in the publications forming the main body of this work.

In clinical CT scanning, the grayscale values in the reconstruction are usually scaled into *CT numbers*. The CT number of a material with the linear attenuation coefficient μ is

$$\text{CT number} = \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}}} 1000, \tag{5.20}$$

and their unit is the *Hounsfield unit* (HU) [177]. The CT number of water is 0 and the CT number of air is -1000. In this work, the grayscale values of reconstructions obtained using laboratory scanners are unscaled, and reconstructions from clinical scanner data have been scaled to HUs.

5.5 Proximal Algorithms for CBCT Reconstruction

This section will detail some of the CBCT reconstruction algorithms developed during this work, specifically those based on the primal-dual fixed point (PDFP) proximal optimization algorithm introduced in Section 4.5.1, and which form the basis of the TV-CGS algorithm presented in Publication IV.

In CT imaging, the tissues or materials of interest are often assumed to be approximately piecewise constant in their attenuation properties. Furthermore, according Lambert-Beer's law, X-rays can only ever be attenuated, never amplified, when passing through matter. Therefore the image reconstruction scheme detailed in Section 4.5.1, which uses the PDFP optimization algorithm with total variation regularization and nonnegativity constraints, is well suited to image reconstruction in computed tomography. In this setting, A and m in Eq. (4.30) become the CBCT forward model and the sinogram data, respectively.

An iterative reconstruction algorithm that uses variational regularization will involve numerous forward projections (multiplying with A) and backprojections (multiplying with A^T). In order to preserve the total intensity of the grayscale values of f in such operations, we replace A with the normalized forward model $\widetilde{A} = A/||A||_2$ and the measurement data m with the normalized data $\widetilde{m} = m/||A||_2$ in the solver. Otherwise, the final values of f will not correspond (on average) with those obtained from an FDK reconstruction, and at worst they can spiral toward infinity or toward zero. The ℓ^2 operator norm of A can be estimated, for example, by using the power method, which was done in this work [44].

In TV regularization, the linear transformation B in Eq. (4.24) is assigned as the discrete gradient matrix D, as defined in Section 3.4. CBCT is a 3D imaging modality, and therefore we have the upper limit $\lambda_{\max}(DD^T) \leq 12$, as shown in Section 3.4.1. This establishes the allowed range for the PDFP parameter $0 < \lambda < 1/12 \leq 1/\lambda_{\max}(DD^T)$.

An upper limit for the Lipschitz constant K for the gradient of the (normalized) data fidelity term

$$\nabla \frac{1}{2} ||\widetilde{A}f - \widetilde{m}||_2^2 = \widetilde{A}^T(\widetilde{A}f - \widetilde{m})$$

can be estimated as follows. Let $f_1, f_2 \in \mathbb{R}^N$ and $\widetilde{A} = 1$. Now:

$$\begin{aligned}
||\widetilde{A}^T(\widetilde{A}f_1 - \widetilde{m}) - \widetilde{A}^T(\widetilde{A}f_2 - \widetilde{m})||_2 &= ||\widetilde{A}^T\widetilde{A}f_1 - \widetilde{A}^T\widetilde{m} - \widetilde{A}^T\widetilde{A}f_2 + \widetilde{A}^T\widetilde{m}||_2 \\
&= ||\widetilde{A}^T\widetilde{A}f_1 - \widetilde{A}^T\widetilde{A}f_2||_2 \\
&= ||\widetilde{A}^T\widetilde{A}(f_1 - f_2)||_2 \\
&\leq ||\widetilde{A}^T||_2||\widetilde{A}(f_1 - f_2)||_2 \\
&= ||\widetilde{A}(f_1 - f_2)||_2 \\
&\leq ||\widetilde{A}||_2||(f_1 - f_2)||_2 \\
&= ||(f_1 - f_2)||_2.
\end{aligned}$$

Therefore $K = 1/\beta \leq 1 \Rightarrow \beta \geq 1$. This establishes that $(0, 2)$ is an allowed range for the PDFP parameter $0 < \gamma < 2\beta$. In practical computations, it will only be true that $||\widetilde{A}||_2 \approx 1$, but if γ is chosen so that is not very close to 2, convergence issues can be avoided.

The chosen convergence metric for the reconstruction algorithms is the relative distance between iterations:

$$s^\nu = \frac{||f^\nu - f^{\nu-1}||_2}{||f^\nu||_2}, \tag{5.21}$$

and the termination criterion is a minimum distance $s_{\min}$. This metric is very general, as it does not rely on any external reference outside the iterative reconstruction procedure.

5.5.1 ATV-NN-PDFP

The algorithm that uses anisotropic total variation and nonnegativity constraints for CBCT reconstruction and is implemented using the PDFP optimization algorithm is titled the ATV-NN-PDFP algorithm. The pseudocode for ATV-NN-PDFP is shown in Algorithm 5.1.

Algorithm 5.1 ATV-NN-PDFP

Input: data $\widetilde{m}$, model $\widetilde{A}$, parameters α, $\nu_{\max}$, $s_{\min}$, starting point f^0
Initialize: $\gamma \leftarrow 1$, $\lambda \leftarrow \frac{1}{13}$, $v^0 = Df^0$
for $\nu \leftarrow 1, \ldots, \nu_{\max}$ **do**
  $g^\nu \leftarrow \text{proj}_{\mathbb{R}^N_+}(f^{\nu-1} - \gamma\widetilde{A}^T(\widetilde{A}f^{\nu-1} - \widetilde{m}) - \lambda D^T v^{\nu-1})$
  $v^\nu \leftarrow (I - S_{\frac{\gamma}{\lambda}\alpha})(Dg^\nu + v^{\nu-1})$
  $f^\nu \leftarrow \text{proj}_{\mathbb{R}^N_+}(f^{\nu-1} - \gamma\widetilde{A}^T(\widetilde{A}f^{\nu-1} - \widetilde{m}) - \lambda D^T v^\nu)$
  $s^\nu \leftarrow \frac{||f^\nu - f^{\nu-1}||_2}{||f^\nu||_2}$
  if $s^\nu < s_{\min}$ **then**
    break
  end if
end for
return f^ν

5.5.2 TV-NN-PDFP

The algorithm that uses (isotropic) total variation and nonnegativity constraints for CBCT reconstruction and is implemented using the PDFP optimization algorithm is titled the

TV-NN-PDFP algorithm. The pseudocode for TV-NN-PDFP is shown in Algorithm 5.2. It is worth emphasizing that the ATV-NN-PDFP and TV-NN-PDFP algorithms only differ in the proximal operator that is used in determining v^ν.

Algorithm 5.2 TV-NN-PDFP

Input: data $\widetilde{m}$, model $\widetilde{A}$, parameters α, $\nu_{\max}$, $s_{\min}$, starting point f^0
Initialize: $\gamma \leftarrow 1$, $\lambda \leftarrow \frac{1}{13}$, $v^0 = Df^0$
for $\nu \leftarrow 1, \ldots, \nu_{\max}$ **do**
  $g^\nu \leftarrow \text{proj}_{\mathbb{R}^N_+}(f^{\nu-1} - \gamma\widetilde{A}^T(\widetilde{A}f^{\nu-1} - \widetilde{m}) - \lambda D^T v^{\nu-1})$
  $v^\nu \leftarrow (I - \text{prox}_{\frac{\gamma}{\lambda}\alpha||\cdot||_{2,1}})(Dg^\nu + v^{\nu-1})$
  $f^\nu \leftarrow \text{proj}_{\mathbb{R}^N_+}(f^{\nu-1} - \gamma\widetilde{A}^T(\widetilde{A}f^{\nu-1} - \widetilde{m}) - \lambda D^T v^\nu)$
  $s^\nu \leftarrow \frac{||f^\nu - f^{\nu-1}||_2}{||f^\nu||_2}$
  if $s^\nu < s_{\min}$ **then**
    break
  end if
end for
return f^ν

5.5.3 Experimental Results for ATV-NN-PDFP and TV-NN-PDFP

In this thesis work, the algorithms were tested using two sets of real CBCT data.

The first test dataset was a series of scans of a walnut at varying dose levels, imaged using the in-house designed and constructed CBCT scanner in the Industrial Mathematics Computed Tomography Laboratory, detailed in Section 6.3. Projections over the full 360° angular range were acquired using a step-and-shoot method. This dataset could reflect imaging of biological samples in non-medical or pre-clinical CBCT and microtomography, and it is freely available at [229]. A high dose scan with good signal-to-noise properties was used as the reference scan (labeled "Dose 100%"). The reference scan was reconstructed using FDK. A low dose scan with a 1/10 dose level compared to the reference scan was reconstructed using FDK, ATV-NN-PDFP, and TV-NN-PDFP. The goal was to study the potential of the PDFP-based algorithms in improving image quality in ultra low dose CBCT scans. It should be stated that the dose levels were chosen somewhat arbitrarily, in order to compare two quite extreme imaging scenarios. In a practical imaging setting, the dose level is dictated by factors such as available scan time and the vulnerability of the sample to ionizing radiation.

The second dataset was a series of scans of the cervical vertebrae of an Erler-Zimmer X-ray phantom head [230] at varying dose levels. The Erler-Zimmer phantom contains an actual human skull safely embedded in plastic, resulting in realistic anatomy in the bony structures. The scans were made using a Planmeca Viso G7, a clinical CBCT scanner designed for head and neck imaging [231]. The scanner made all the measurements using short-scans and continuous acquisition where the scanner was in continuous motion as the projections were acquired. The dose level suggested by the scanner was used as the reference scan (labeled "Clinical dose"). The reference scan was reconstructed using FDK. The tube current - exposure time product (mAs) of the scanner was adjusted to produce an ultra low dose scan, reconstructed using FDK, ATV-NN-PDFP, and TV-NN-PDFP. As before, the low dose level was chosen somewhat arbitrarily, and creating an actual clinical scan protocol would require careful balancing of dose and image quality metrics.

Table 5.1: Settings for the CBCT scans of imaging phantoms.

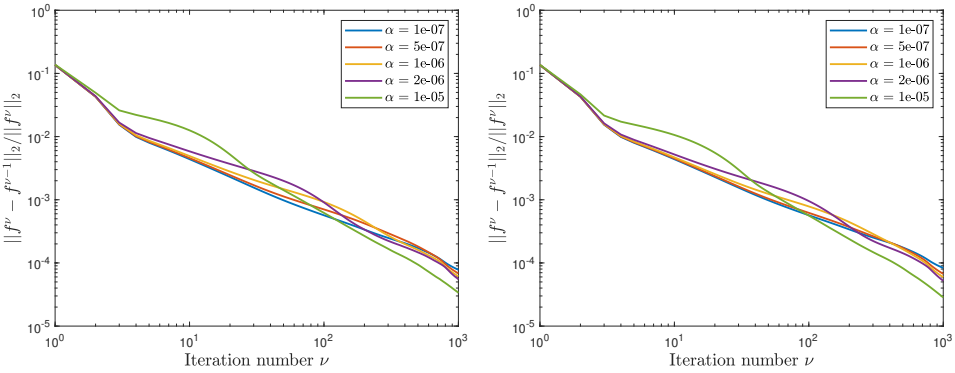
Phantom	Walnut	Walnut	Cervical vertebrae	Cervical vertebrae
Dose level	100%	10%	Clinical dose	Ultra low dose
Scanner	In-house	In-house	Viso G7	Viso G7
Target	Mo	Mo	W	W
X-ray filter	0.5 mm Al	0.5 mm Al	N/A	N/A
U (kV)	40	40	100	100
I (mA)	1.0	0.2	9.0	2.0
mAs	N/A	N/A	36.0	5.0
Exposure time (ms)	2000	1000	N/A	N/A
Angular range	360°	360°	N/A	N/A
# projections	360	360	500	500
Voxel size (mm)	0.0725	0.0725	0.450	0.450
x-size	420	420	468	468
y-size	420	420	468	468
z-size	420	420	334	334
α	N/A	$1 \cdot 10^{-7}$	N/A	$1 \cdot 10^{-7}$
		$5 \cdot 10^{-7}$		
		$1 \cdot 10^{-6}$		$1 \cdot 10^{-6}$
		$2 \cdot 10^{-6}$		$2 \cdot 10^{-6}$
		$1 \cdot 10^{-5}$		$1 \cdot 10^{-5}$
$\nu_{\max}$	N/A	1000	N/A	500
$s_{\min}$	N/A	10^{-6}	N/A	10^{-6}

All algorithms were implemented with the MATLAB programming language. To facilitate comparison of reconstruction times, all computations were made using MATLAB R2023b [232] on a Lenovo ThinkPad P16 Gen 2 PC workstation equipped with 32 GB of RAM and a 16 GB NVIDIA GeForce RTX 4090 graphics processing unit (GPU). GPUs are extremely well suited for CT computations, as they are specifically designed for projection operations and parallel computing. HelTomo, a MATLAB workflow toolbox developed in-house to facilitate efficient and smooth use of CBCT data, was used in computing the reconstructions [233]. As a backend, HelTomo used the ASTRA Toolbox v. 2.1.0 for modelling the CBCT imaging system and for FDK reconstructions [234, 235]. The forward models A were implemented as GPU-based matrix-free linear operators using ASTRA and the Spot linear-operator toolbox [236], and FDK reconstructions were also computed using the GPU. The reconstructions acquired using the clinical CBCT scanner data were converted into Hounsfield units (HU) based on attenuation coefficient data from the scanner logs.

The scan settings and reconstruction settings for both sets of scans are shown in Table 5.1. Exposure time in this case refers to the exposure time of a single X-ray projection.

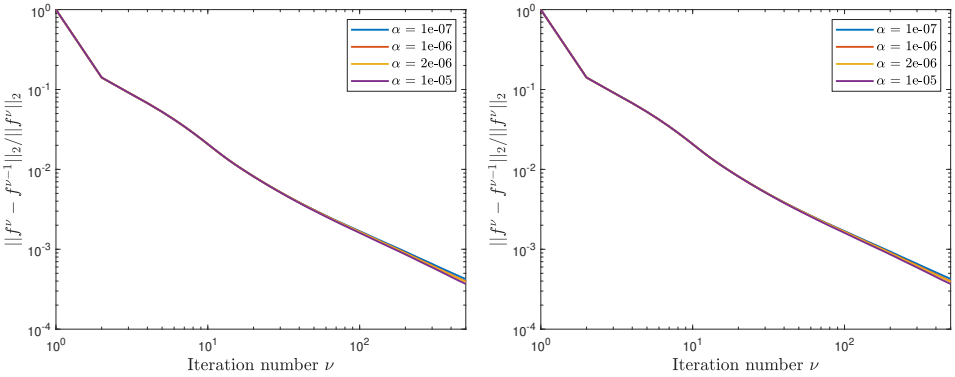
Figure 5.4 shows the convergence of the relative distance between iterations for CBCT reconstructions of the low dose walnut dataset using ATV-NN-PDFP and TV-NN-PDFP, respectively. Figure 5.5 shows the same data for the ultra low dose cervical vertebrae dataset. The numerical iterations exhibit stable convergence for both 360° and short-scan CBCT imaging, in accordance with the theoretical convergence properties of the PDFP optimization scheme.

Figure 5.6 shows a comparison of FDK and ATV-NN-PDFP reconstructions for the low dose walnut dataset, using the center slices of the volume in the xy-, yz-, and xz-directions. The 100% dose FDK reconstruction exhibits crisp boundaries, low noise, and good contrast. The 10% dose FDK reconstruction is extremely noisy, and the contrast



(a) ATV-NN-PDFP **(b)** TV-NN-PDFP

Fig. 5.4. Convergence of the relative distance between iterations in the ATV-NN-PDFP and TV-NN-PDFP algorithms for CBCT for the low dose walnut dataset.



(a) ATV-NN-PDFP **(b)** TV-NN-PDFP

Fig. 5.5. Convergence of the relative distance between iterations in the ATV-NN-PDFP and TV-NN-PDFP algorithms for CBCT for the ultra low dose cervical vertebrae dataset.

between materials as well as fine detail detectability are severely degraded. The appearance of the 10% dose ATV-NN-PDFP reconstructions depends radically on the value of the regularization parameter α. When α is too small, there is very little regularization, and no benefit to using iterative reconstruction. For $\alpha = 2 \cdot 10^{-6}$ the noise is suppressed without blurring of boundaries, and some contrast is recovered. When α is too high, the reconstruction becomes overly smooth, and small structures and features are lost. Furthermore, for large α the reconstruction is beginning to exhibit an effect where transitions are preferably aligned with the Cartesian coordinate axes, leading to staircase-like image artifacts. This is an undesireable feature of anisotropic total variation. In all ATV-NN-PDFP reconstructions the vertical top and bottom slices exhibit iterative CBCT artifacts, which occur in MBIR algorithms when the scanned object extends beyond the field of view in the z-direction [237, 238]. It should be emphasized that the values of α were chosen *ad hoc* to demonstrate the effect of regularization, and $\alpha = 2 \cdot 10^{-6}$ should not be assumed to be the optimal parameter choice. Nevertheless, even subjective image quality evaluation makes it evident that the regularized image reconstruction scheme can hugely improve low dose CBCT scans.

Figure 5.7 shows a comparison of FDK and TV-NN-PDFP reconstructions for the low dose walnut dataset. The same observations can be made as with the ATV-NN-PDFP reconstruction results, but with one crucial difference. Regardless of the value of α, the reconstructions at no point exhibit any spatial preference for the direction of boundaries. This is a significant advantage of isotropic total variation, as the regularization strength can be freely chosen without fear of the emergence of such a phenomenon.

Figures 5.8 and 5.9 show comparison of FDK and the ATV-NN-PDFP and TV-NN-PDFP reconstruction algorithms for the cervical vertebrae data. The same observations as made for the low dose walnut data still hold, but a number of important additional remarks must be made. First, even the lowest regularization parameter tried here still resulted in some noise suppression. Second, the MBIR methods have removed the shadow-like artifacts surrounding bony structures in the FDK reconstructions. Third, the noise texture or *granularity* is significantly altered by the regularized reconstruction process. This can be significant, as many radiologists prefer to look at an FBP-type noise structure [239]. Fourth, the z-direction artifacts are significantly worse, making the top few slices somewhat unusable. Finally, the regularized reconstructions preserve the average grayscale values with fidelity, so that the conversion to Hounsfield units is not affected. Overall, ultra low dose regularized reconstruction with a suitable value of α is able to recover visibility of low contrast features and small details when compared to FDK.

Table 5.2 shows the average iteration times for each of the test cases. The ATV-NN-PDFP appears to be up to approximatimately a third faster than TV-NN-PDFP. This is most likely caused by differences in the proximal operators of the ATV/TV penalties, as the latter involves computing a large number of square roots. Nevertheless, the results are in the same order of magnitude, and considering the higher spatial fidelity in (isotropic) TV, it would appear to be the superior choice for gradient-based regularization. For comparison, the reconstruction times for FDK were not even measured, as in all cases they took mere seconds on the same reconstruction computer.

Studying the average iteration times and the number of iterations used to compute the reconstructions affirms the considerations in Section 5.3.2: model-based iterative reconstruction is many orders of magnitude slower than FDK. In this case, computing each reconstruction took hours. The improvements in image quality therefore come at a great price. There are however, a number of alleviations that can be made. Using a high-performance computational cluster will lead to shorter reconstruction times. In many

cases, the reconstructed volumes may be substantially smaller than in the examples used here, as the number of voxels grows in the third power when the edge length of a cube-shaped reconstruction volume is extended. Also, it may be possible to optimize the reconstruction code to perform better. And while MATLAB is an excellent tool for developing and prototyping computational algorithms, the final computational implementation could be done with a faster performing language, such as C++. Finally, one must ask the very important question *when is the MBIR output good enough*? Figure 5.10 shows the TV-NN-PDFP reconstructions of the low dose walnut data at various values of ν when $\alpha = 2 \cdot 10^{-6}$. The iterations start with the noisy FDK reconstruction as f^0. Evaluating the image quality from an end-user's subjective point of view, it is clear that during the first few hundred iterations the reconstruction changes significantly. However, there would appear to be negligible difference between the image quality at $\nu = 500$ and $\nu = 1000$. Therefore, in the practical application of MBIR algorithms, it is imperative from a practical standpoint to establish what number of iterations is needed to strike a good bargain between improved image quality and acceptable reconstruction times.

Table 5.2: Average iteration times in seconds for the ATV-NN-PDFP and TV-NN-PDFP methods for the test data.

		Average iteration time (s)				
		α				
Dataset	Method	$1 \cdot 10^{-7}$	$5 \cdot 10^{-7}$	$1 \cdot 10^{-6}$	$2 \cdot 10^{-6}$	$1 \cdot 10^{-5}$
Dose 10% walnut	ATV-NN-PDFP	40.1	41.1	42.2	46.9	44.4
Dose 10% walnut	TV-NN-PDFP	61.0	65.3	69.0	65.7	68.5
ULD vertebrae	ATV-NN-PDFP	43.8	N/A	45.0	45.7	41.1
ULD vertebrae	TV-NN-PDFP	55.3	N/A	56.8	55.7	57.0

5.5.4 TV-CGS

Variational regularization which has the ℓ^1 norm in the regularization term induces solutions that are sparse in some basis, as discussed in Section 2.3. This sparsity can be quantified. We define the counting function

$$\#_\kappa f = \#\{l \mid 1 \leq l \leq N,\ |f_l| < \kappa\}, \tag{5.22}$$

where $f \in \mathbb{R}^N$ and κ is a small non-negative number. $\#_\kappa f$ counts the number of non-zero entries in f using tolerance κ around zero, which manages floating-point arithmetic issues and extremely low-level noise. We now define the *sparsity level* of f as

$$\mathcal{C}_f = \frac{\#_\kappa f}{N}, \tag{5.23}$$

which is the ratio of non-zero elements to all elements.

Next, we define the *gradient sparsity level* $\mathcal{C}_{\nabla f}$ of f. In anistropic TV, one is interested in tracking the ratio of non-zero gradient coefficients to the total number of gradient coefficients, which yields the definition

$$\mathcal{C}_{\nabla f,\,\mathrm{ATV}} = \frac{\#_\kappa Df}{nN}, \tag{5.24}$$

where n is the dimensionality of the underlying physical model, which in this work is always 2 or 3. In isotropic TV, one is interested in tracking the elements of f where

the corresponding gradient magnitude is non-zero. Using the unconventional notation $||\nabla f||_2 \in \mathbb{R}^N$ to signify the vector containing the gradient magnitude values of f as defined in Eqs. (3.7) and (3.13), we define

$$\mathcal{C}_{\nabla f,\,\mathrm{TV}} = \frac{\#_\kappa ||\nabla f||_2}{N}. \tag{5.25}$$

The gradient sparsity levels $\mathcal{C}_{\nabla f}$ of the ATV-NN-PDFP and TV-NN-PDFP algorithms will both converge to some values $(0, 1)$. This follows directly from the properties of the optimization algorithm. If the requirements of the PDFP algorithm are met, it is guaranteed to converge to a solution. This solution will by definition have some gradient sparsity, towards which the gradient sparsities of the iterates will converge.

Figure 5.11 shows the convergence of $\mathcal{C}_{\nabla f,\,\mathrm{ATV}}$ and $\mathcal{C}_{\nabla f,\,\mathrm{TV}}$ for the low dose walnut test case. Figure 5.12 shows the same data for the ultra low dose cervical vertebrae test case. The value $\kappa = 10^{-6}$ was used in the computations. In all cases, the gradient sparsity is seen to converge towards a certain value. In most cases, the convergence behaves as mathematical intuition would dictate: as the value of α increases, the regularization becomes stronger, leading to fewer non-zero elements in the gradient coefficients (ATV-NN-PDFP) or the gradient magnitude values (TV-NN-PDFP), therefore decreasing the value of $\mathcal{C}_{\nabla f}$. The TV-NN-PDFP results deviate from this expectation for $\alpha \leq 5 \cdot 10^{-7}$ in the walnut data. It can be hypothesized that this is due to the extreme amount of noise, which can lead to unpredictable outcomes in numerics. However, as seen in the previous section, such extremely small values of α lead to negligible regularization, and as such are not relevant from the point of view of image reconstruction.

The previous results show that TV-NN-PDFP (and ATV-NN-PDFP to a lesser extent) has potential for improving image quality in CBCT reconstruction. Nevertheless, this image reconstruction method formulated on the minimization problem in Eq. (4.30) has its issues. As seen from the reconstruction results in Section 5.5.3, the possible range of values for α spans many orders of magnitude, and choosing a suitable value for the regularization parameter is known to be a highly non-trivial task [144]. Computing many reconstructions and then choosing a suitable one becomes extremely expensive and time-consuming from a computational point of view. Furthermore, a suitable regularization parameter choice is highly case-specific, and depends also on the noise level or sparsity of the projection data. Finally, the actual numerical value of α does not offer much in the way of interpretability.

The reconstruction method *Total Variation using Controlled Gradient Sparsity* flips the previous results around: the reconstruction algorithm is designed in such a way that the gradient magnitude sparsity in (isotropic) total variation regularized reconstruction converges towards some *a priori* defined value $\mathcal{C}\mathrm{pr} \in (0, 1)$. TV-CGS is built upon the TV-NN-PDFP method, but α is now treated as a dynamic variable that can change at every iteration.

The underlying assumption in TV-CGS is that the gradient magnitude sparsity of the imaged object, the so-called true image, remains constant regardless of factors such as image noise. Thus, if $\mathcal{C}_{\nabla f}$ is driven to converge towards a certain value, the value of α will automatically adjust to the noise level. Of course, one still needs to know the prior gradient magnitude sparsity. We hypothesize that for similar objects the sparsity remains approximately constant, allowing a gradient magnitude sparsity atlas to be built for various imaging settings. Furthermore, defining regularization in terms of $\mathcal{C}\mathrm{pr} \in (0, 1)$ significantly increases interpretability of the reconstruction process, as the numerical value dictating the regularization strength now has an intuitive meaning. Thus, even when the regularization strength needs to be adjusted, it is easier to choose values from the interval

$(0, 1)$ than from the whole range of positive real numbers. The convergence plots for the gradient sparsity would seem to suggest that the controlled gradient sparsity strategy is very stable for anisotropic total variation, whereas isotropic total variation requires a certain level of regularization to achieve predictable behaviour.

The TV-CGS image reconstruction method and results are described in detail in Publication IV.

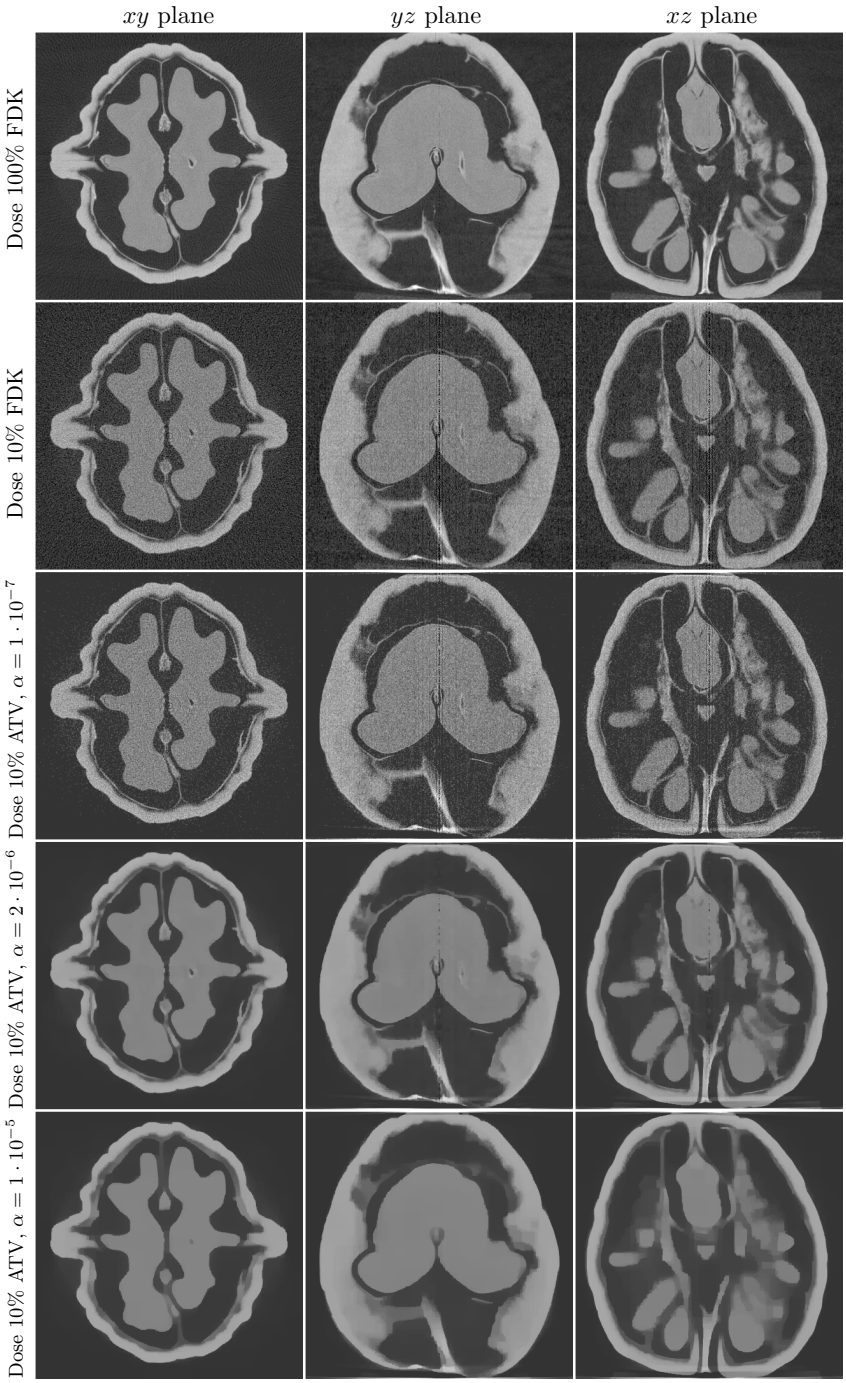


Fig. 5.6. Comparison of FDK and ATV-NN-PDFP reconstructions for the low dose walnut dataset. The high dose FDK reconstruction acts as a reference. Dataset from [229]. All images use identical windowing and gamma correction.

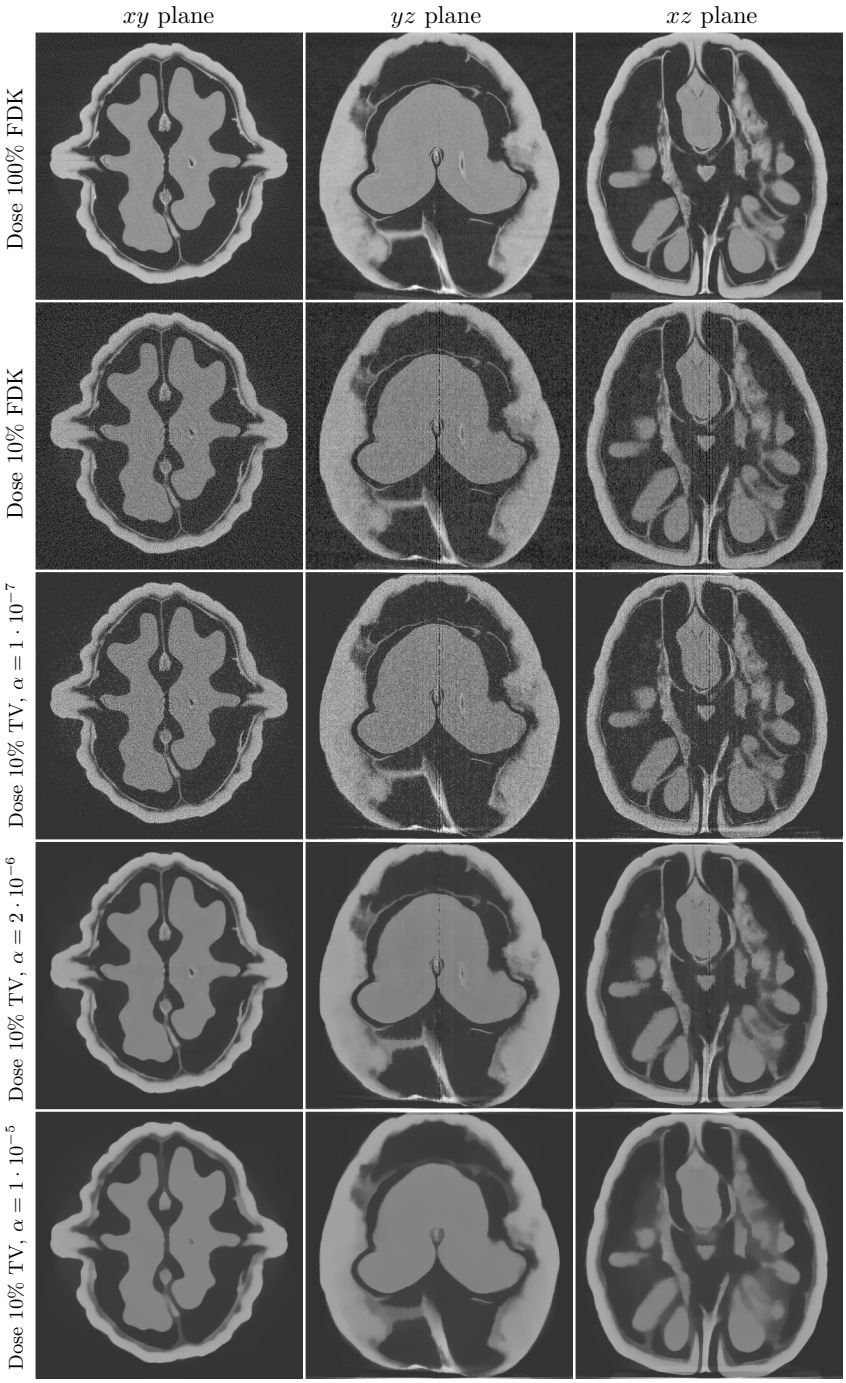


Fig. 5.7. Comparison of FDK and TV-NN-PDFP reconstructions for the low dose walnut dataset. The high dose FDK reconstruction acts as a reference. Dataset from [229]. All images use identical windowing and gamma correction.

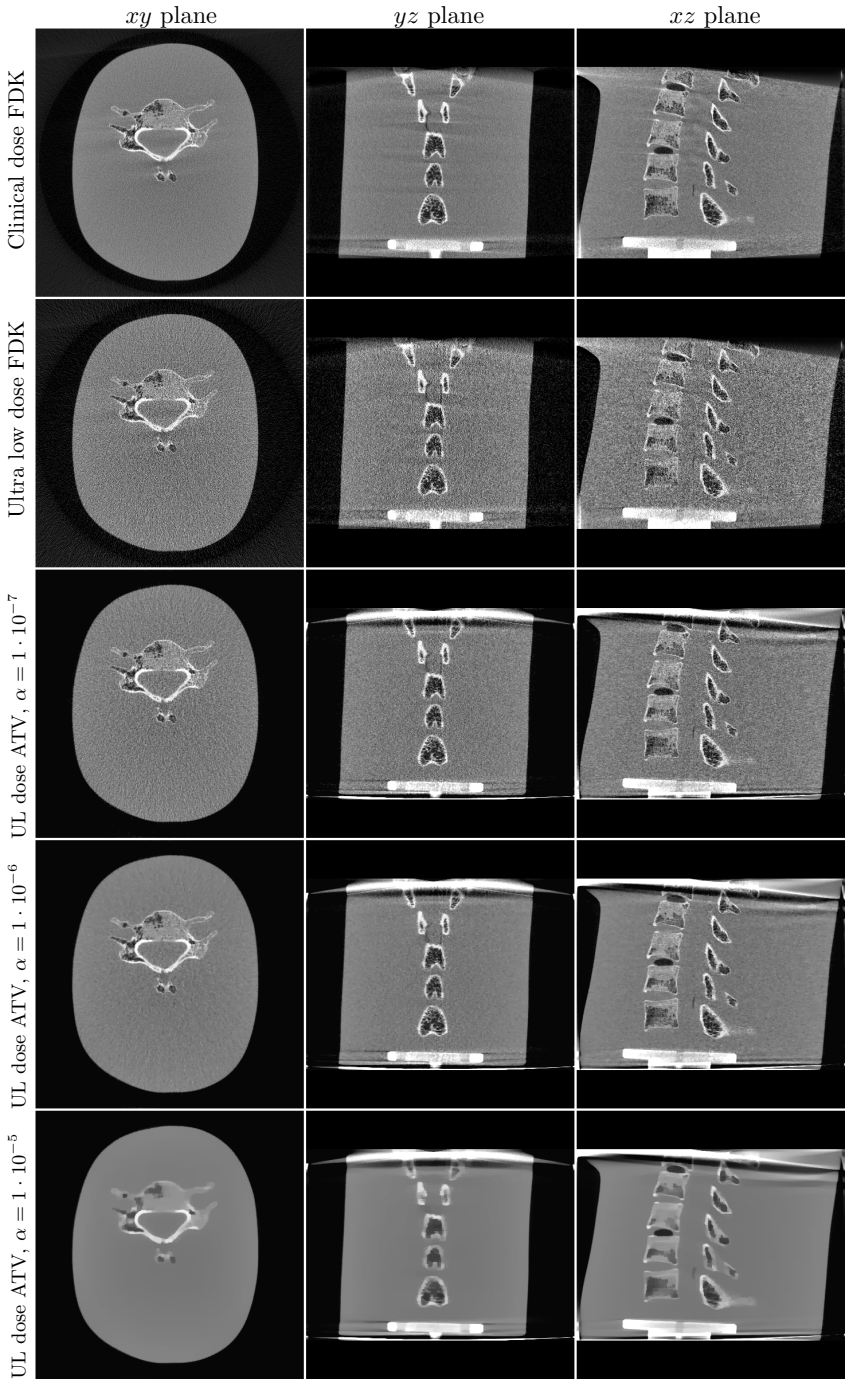


Fig. 5.8. Comparison of FDK and ATV-NN-PDFP reconstructions for the ultra low dose cervical vertebrae dataset. The high dose FDK reconstruction acts as a reference. All images use the HU windowing $[-1050, 1000]$.

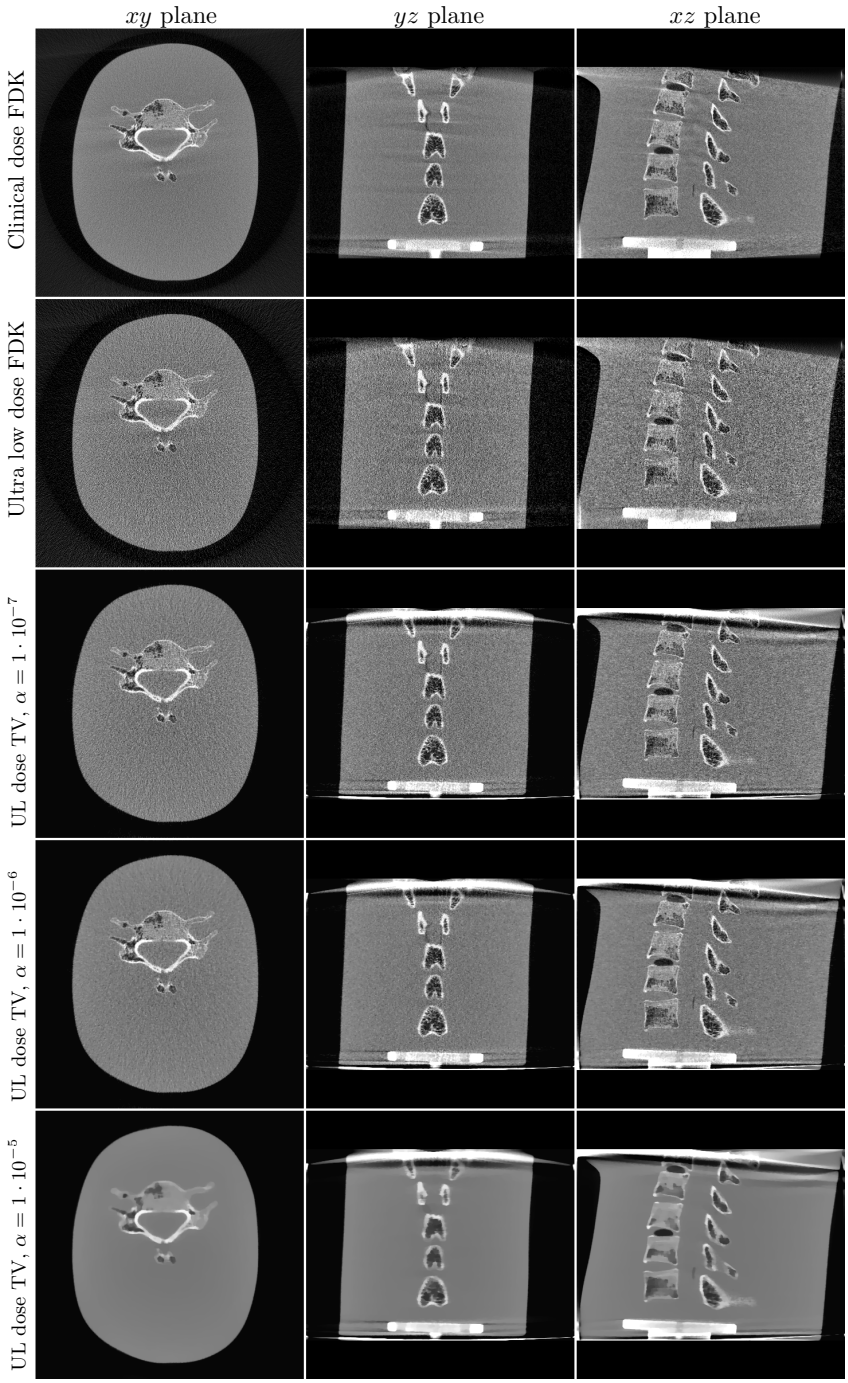


Fig. 5.9. Comparison of FDK and TV-NN-PDFP reconstructions for the ultra low dose cervical vertebrae dataset. The high dose FDK reconstruction acts as a reference. All images use the HU windowing $[-1050, 1000]$.

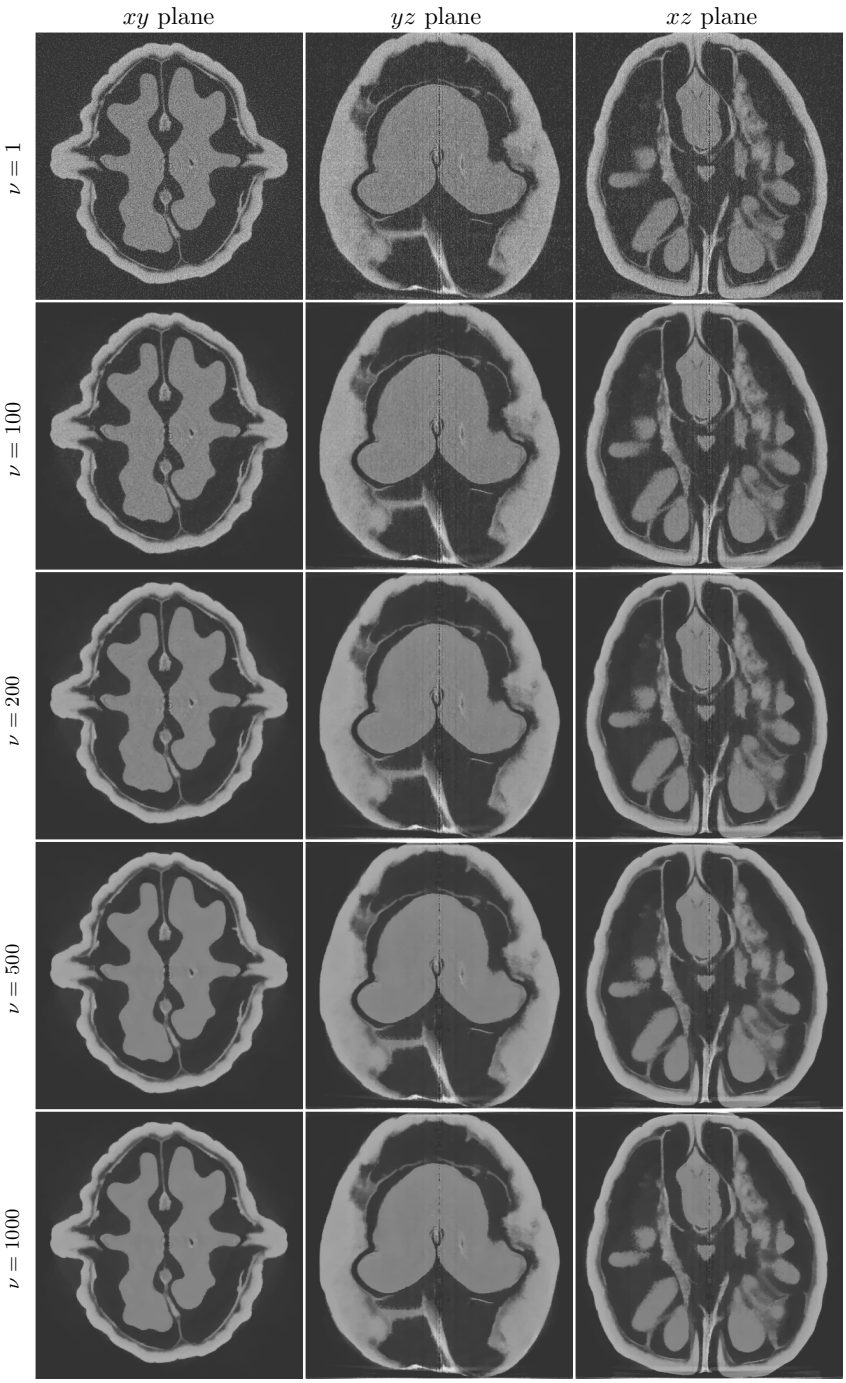
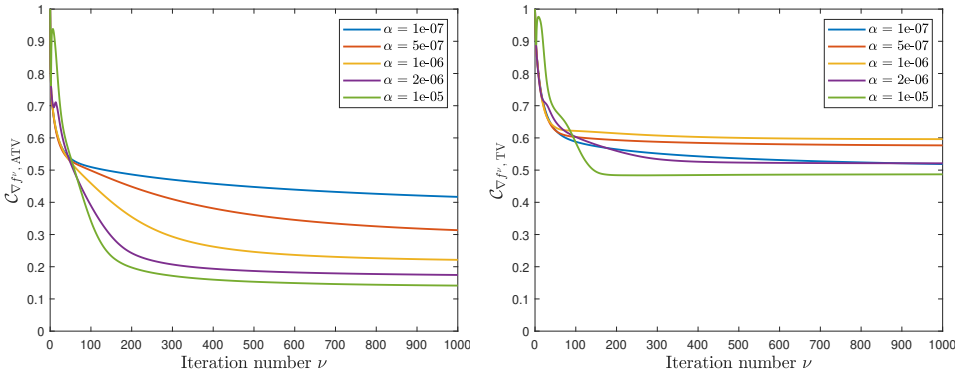
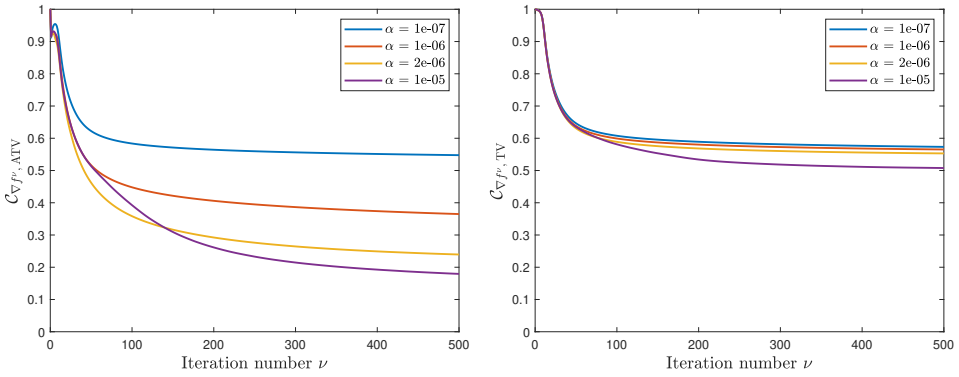


Fig. 5.10. Evolution of the TV-NN-PDFP reconstruction of the low dose walnut data at various iterations ν when $\alpha = 2 \cdot 10^{-6}$.



(a) ATV-NN-PDFP **(b)** TV-NN-PDFP

Fig. 5.11. Convergence of (a) the gradient coefficient sparsity in ATV-NN-PDFP, and (b) the gradient magnitude sparsity in TV-NN-PDFP for the low dose walnut dataset.



(a) ATV-NN-PDFP **(b)** TV-NN-PDFP

Fig. 5.12. Convergence of (a) the gradient coefficient sparsity in ATV-NN-PDFP, and (b) the gradient magnitude sparsity in TV-NN-PDFP for the ultra low dose cervical vertebrae dataset.

6 Open Software and Data for CT Reconstruction

6.1 Reconstruction Algorithms in the Real World

In the endeavour to create CT reconstruction algorithms that are robust, reliable, and fast enough for practical uses, the real work only begins after the mathematical formulation of the solution method has been written down. The algorithm must discretized, implemented as computer code, and rigorously evaluated using a broad range of CT data, as computational inversion can be a complicated and nebulous affair. Many algorithms that in principal have excellent properties can turn out to be so slow to converge that they are practically useless when working with the data sizes encountered in clinical imaging, or they can otherwise behave in an unstable and unpredictable way when implemented in practice.

In order to cultivate a flourishing community of CT algorithm developers, it is vitally important that applied mathematicians and other image reconstruction experts have access to a broad range of computational tools and realistic data that facilitate rapid algorithm prototyping and testing, without the arduous task of building a framework from the ground up, being limited to overly simple test data that is of little use in gauging algorithm performance in practical scenarios, or spending significant effort on fastidiously mapping the mathematical properties of algorithms that turn out to have little practical use [37]. This research approach has also been referred to as plug-and-play algorithm development.

6.2 Reconstruction Tools

Today's algorithm designer fortunately does not need to start from zero when implementing their research into computer code. Within the last decade or so, multiple open source frameworks have been released specifically for the purpose of CT image reconstruction. These include the ASTRA Toolbox [234, 235, 240], CONRAD [241, 242], the Core Imaging Library (CIL) [243–245], elsa [246, 247], FreeCT [248–250], NeuTomPy [251, 252], The Reconstruction Toolkit (RTK) [253, 254], SCICO [255, 256], and TIGRE [257, 258], to name a few. Each one of these has its own set of available imaging geometries, features, and supported programming languages. Standard reconstruction algorithms and forward modelling are usually included, and other features such as data pre-processing, regularization, and optimization tools may be available. Many frameworks have GPU support for efficient computation of forward and backwards projections. The prospective algorithm designer would be well advised to gauge which computational framework is best suited to their needs, as the benefits of choosing the right software tools can be substantial. The Tomopedia Wiki hosts more information on tomography-related algorithms and software packages [259].

Throughout this work, the framework for CT reconstruction algorithm development was for the most part the MATLAB version of the ASTRA Toolbox. Furthermore, a MATLAB workflow toolbox was designed and implemented in-house to facilitate a smooth workflow, with quick and seamless integration of the ASTRA toolbox functionalities, algorithm testing, and CT data processing. *HelTomo - Helsinki Tomography Toolbox* is available via Mathworks File Exchange [233], and it has been designed to interface with the open datasets published during this work.

6.3 Open CT Projection Datasets

During the process of CT reconstruction algorithm development, a natural first approach to proof of concept testing and sanity testing is to use simple computational phantoms which can be quickly and easily generated. This could mean, for instance, simple geometric shapes, or the ubiqutous Shepp-Logan phantom and its variations [179, 260–262]. However, algorithm testing with such phantoms is of limited use when gauging how well a reconstruction algorithm will perform in a real, practical situation, and more realistic data is therefore needed. Regrettably for algorithm developers working in the field of medical CT, it is well known that commercial CT scanner manufacturers rigorously guard access to raw data from their machines. Moreover, using real patient data is restricted by patient privacy laws.

Fortunately, advanced software has been developed for creating synthetic human phantom data [263, 264]. Using a dedicated *CT simulator* allows simulating realistic X-ray projections from computational human phantoms [265, 266]. CT simulators often utilize highly advanced models of the measurement circumstances, taking into account factors such as photon noise, scatter, and detector physics. Furthermore, using a dedicated CT simulator to create X-ray projections allows the algorithm developer to avoid committing inverse crime when generating test data. Publications I and II of this work used simulated human X-ray data created using the XCAT software [267, 268]. XCAT includes tools for anthropomorphic phantom design as well as a dedicated CT projector.

Ultimately, the performance of a CT reconstruction algorithm should always be assessed by using real, measured X-ray projection data. Although CT simulators can be used to create highly realistic X-ray projection data, it is impossible to fully capture all the myriad phenomena found in real data, including complex and diverse radiation-matter interactions, detector physics, eletronic phenomena, A/D conversion, and modeling error [37–39, 111].

Until quite recently, very little real data had been available for CT reconstruction algorithm developers. However, there is currently an increasing emphasis on open data and open datasets in science [269–271], and a growing number of real projection datasets have been published. These include non-medical CBCT datasets acquired with non-medical scanners, such as [272, 273]. The TomoBank repository provides access to a number of experimental and simulated datasets for synchrotron computed tomography [274, 275]. It is important to point out that such non-medical datasets can also be of high value to a medical CT algorithm designer, as the underlying physics and mathematics remain the same. In helical CT, a vendor-neutral open data format for CT projection data has been developed [38]. This DICOM-CT-PD format is compatible with the DICOM file standard used in medical imaging. However, to the best knowledge of the author, only one dataset containing clinical patient data from helical CT scanners has been published at the time of writing [39]. Even this dataset consists partially of simulated data. Furthermore, the data has been provided only after vendor-specific preprocessing and logarithm conversion, which

makes it impossible to use the data for testing preprocessing algorithms for projection data.

Throughout this work, a philosophy was maintained whereby CT reconstruction algorithms were applied to real data as soon as possible. To this end, the *Industrial Mathematics Computed Tomography Laboratory*, first detailed in [195], was maintained and upgraded. The laboratory contains a cone-beam scanner consisting of a molybdenum[1] target X-ray tube (model XTF5011, Oxford Instruments plc, Abingdon, United Kingdom), a motorized rotation stage with variable geometric magnification, and two X-ray detectors: a conventional energy-integrating detector (Hamamatsu Photonics C7942CA-22) and an energy-sensitive photon-counting detector (XCounter Actaeon FX5) utilizing CdTe as a direct conversion semiconductor material. The photon-counting detector can be moved in front of the energy-integrating detector, facilitating both conventional CBCT and photon-counting CBCT measurements using the same scanner. Many of the real data measurements made during this work have been published as open datasets, either as 2D fan beam data or as fully 3D cone beam data.

Two open datasets are highlighted here. The first one is the Helsinki Tomography Challenge 2022 (HTC2022) open tomographic dataset [276]. This 2D fan beam dataset is one of the main focuses of Publication III of this work, where it is described in detail. Reconstruction challenges such as HTC2022 have become a highly popular way of engaging reconstruction algorithm developers and advancing algorithm development [277–281, 201, 282–284]. The second dataset of note is the multienergy bird dataset used in Publications I and II of this work [285]. It consists of a biological sample, specifically the chest of a small bird (the common quail, *Coturnix coturnix*). The multienergy bird dataset is the only dataset presented here that was not measured in the Industrial Mathematics Computed Tomography Laboratory. It was scanned in the University of Helsinki Micro-CT Laboratory using a GE phoenix nanotom microtomography scanner. The sample was sequentially scanned using three different X-ray spectra, making it suitable for use in multienergy CT reconstruction. The measurement setup and example FBP reconstructions for the different energies are shown in Figure 6.1. The bird phantom contains many fine bone structures, making it suitable for studying how algorithms perform in reconstructing small details.

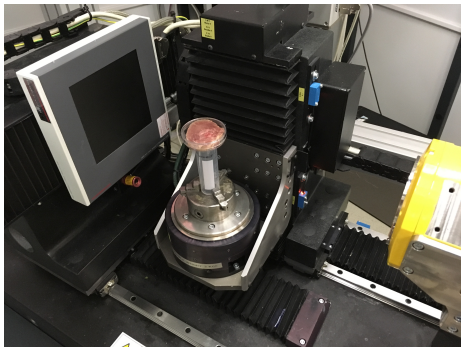
Other open datasets produced during this work include:

- Cone-Beam Computed Tomography Dataset of a Walnut Imaged at 4 Different Dose Levels [229].
- Cone-beam Gel phantom data for dynamic X-ray tomography [286].
- Cone-Beam Computed Tomography Dataset of a Chicken Bone Imaged at 4 Different Dose Levels [287].
- Cone-Beam Computed Tomography Dataset of a Walnut [288].
- Cone-Beam Computed Tomography Dataset of a Pine Cone [289].
- Cone-Beam Computed Tomography Dataset of a Seashell [290].
- Gel phantom data for dynamic X-ray tomography [291, 292].
- Tomographic X-ray data of time-dependent 3D cross phantom [293, 294].
- Tomographic X-ray data of carved cheese [295, 296].

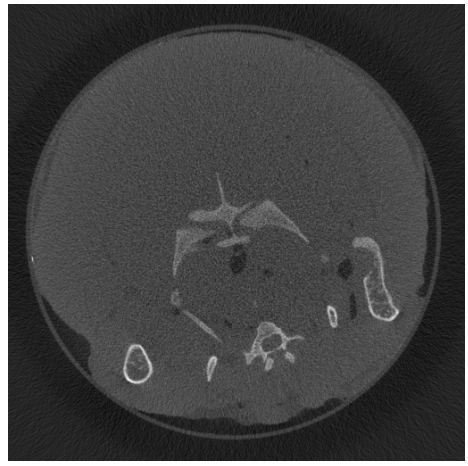
[1]Molybdenum is a somewhat unusual target material in X-ray CT, but it is occasionally utilized in microtomography.

- Tomographic X-ray data of 3D emoji [297, 298].

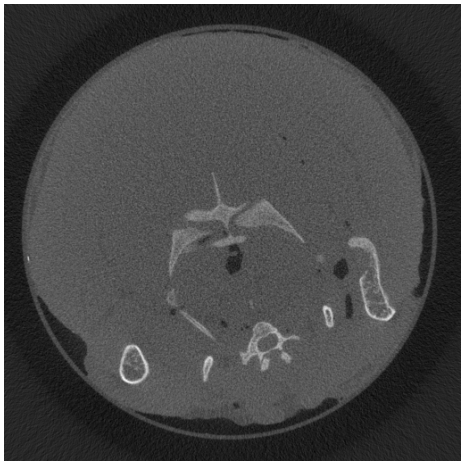
The Finnish Inverse Problems Society FIPS maintains links on its website to open X-ray CT datasets produced by its member organizations [299].



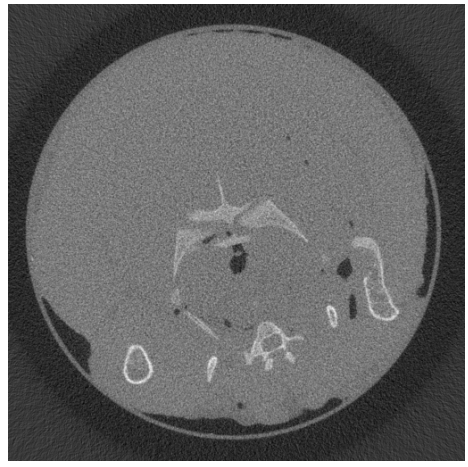
(a) Measurement setup



(b) Low energy FBP



(c) Middle energy FBP



(d) High energy FBP

Fig. 6.1. (a) The measurement setup for the bird phantom and (b)-(d) FBP example reconstructions for the X-ray data acquired using three different X-ray spectra. The tissue contrast changes significantly between the acquisitions made at different energies. The FBP computations were made using the Phoenix datos|x CT software provided with the microtomography scanner.

7 Summary of Results

7.1 Publication I

This work explored the use of different variational regularization strategies for *joint reconstruction* in 2D multispectral CT image reconstruction. Joint reconstruction is a technique where one combined inverse problem is created for computing the reconstructions for all the energy channels, using regularizers which are designed to promote structural similarity between the different energy channels. The regularizers studied in this work included joint total variation (JTV), linear parallel level sets (LPLS) [300], a spectral smoothness regularizer that penalizes differences along the energy direction (D1), and a structural prior based on the structural similarity index (SSIM) image quality metric but only using the structure part (S) [218, 219]. Additionally, adding channel-wise total variation penalties to the D1 prior (D1+TV) and to the structural prior (S+TV) was also studied. Furthermore, to provide a baseline for comparison, independent reconstructions were computed for each individual channel using a least-squares approach (no prior) and total variation regularization (TV). The priors TV, JTV, LPLS, S, D1+TV, and S+TV all used smoothing to make the objective functions everywhere differentiable.

The study used a novel low dose data acquisition protocol where sparse X-ray projections using three different energies were acquired from non-overlapping directions by rapidly cycling through different X-ray spectra. As test cases, simulated anthropomorphic chest data was created using the XCAT software, and real data of the chest of a common quail bird was measured using a cone beam micro-CT scanner. The reconstruction computations were implemented using the Polak-Ribière nonlinear conjugate gradient method, using steepest descent when the β_{PR} parameter became negative. The computations also used projection to the nonnegative orthant, in which case there are no global convergence guarantees for nonlinear CG, but all the numerical experiments proved to be stable.

The joint reconstruction techniques consistently outperformed channel-by-channel reconstruction, showing that utilizing measurement data across the energy channels can improve the image quality of the reconstructions for each individual channel. The combined regularization approaches D1+TV and S+TV yielded good results, as they provided a good balance between noise suppression and promoting similarity across the energy channels. Furthermore, the S+TV method was compared to channel-wise TV in a setting where the multienergy CT data fully came from overlapping projection directions. Joint reconstruction using S+TV outperformed channel-by-channel TV also here, suggesting that the S+TV regularization technique has potential for improving image quality in multienergy photon-counting detector CT reconstruction.

7.2 Publication II

This work explored a synergistic approach to sparse-angle multispectral CT image reconstruction in 2D, whereby information from the different energy channels were combined to improve the individual reconstructions. In multispectral CT, the individual energy channels often suffer from poor noise statistics. To counter this, all the information from the different energy channels was fused into a single polyenergetic sinogram from which a reconstruction was obtained and used as side information in directional total variation (dTV) regularized image reconstruction. The underlying assumption here is that although the different energy channels will have differing grayscale values and contrast, tissue boundaries and edges will be found in the same locations. For comparison, independent channel-by-channel total variation (TV) regularized reconstructions were also computed.

The study used a low dose data acquisition protocol where sparse X-ray projections using three different energies were acquired from the same (*i.e.* overlapping) directions for each energy. This makes the methodology amenable for use in multi-energy photon-counting detector CT reconstruction. As a real data test case, the same bird data as in Publication I was used, albeit with different, overlapping subsampling of the measurement data. As a simulated data test case, anthropomorphic chest data was created using the XCAT software, modelling an idealized energy-sensitive three-channel X-ray detector. The simulated XCAT data in this work was not based on that used in Publication I.

Two different methods were compared for solving the dTV and TV regularized image reconstruction tasks. A reconstruction problem using variational regularization was formulated and solved using the forward–backward splitting (FBS) algorithm [301], which is a proximal algorithm. Iterative regularization was also employed by using Linearized Bregman iterations [302, 303], which allow incorporating the dTV or TV prior into the image reconstruction task.

In all test cases the dTV prior outperformed the channel-wise TV prior. There were no substantial differences in image quality between FBS and linearized Bregman iterations. However, the Bregman iterations were faster in converging to the desired solution. The side information in the dTV regularized image reconstruction method is included in such a way that it is possible to use parallel computing for reconstructing the different energy channels. In terms of reconstruction speed, this may give the dTV method an edge over joint reconstruction methods such as total nuclear variation (TNV) [128], where the different energy channels are reconstructed simultaneously using a single combined very large inverse problem, at least when the resulting image quality is otherwise comparable.

7.3 Publication III

This work details the Helsinki Tomography Challenge 2022 (HTC2022), which was organized to promote development of novel new methods for CT reconstruction. HTC2022 was an international algorithm development challenge open to all research groups. Its focus was on 2D limited angle tomography, a highly challenging inverse problem where the ill-posedness of the reconstruction task becomes increasingly severe the more the view angle is restricted. This publication details both the challenge itself, and the resulting open dataset published after the challenge.

The goal of HTC2022 was to recover the shapes of 2D phantoms from limited angle CT data and to provide a segmentation into two components: air and polymeric material. HTC2022 relied exclusively on real fan-beam CT data, acquired with the cone-beam

scanner in the Industrial Mathematics Computed Tomography Laboratory. The imaging phantoms were circular disks consisting of a single polymeric material and containing holes of various amounts and shapes. They were designed using an algorithm based on Voronoi diagrams [304].

Altogether 26 physical phantoms were produced and scanned for HTC2022. These were divided into two categories: example/training data and test data. The training data consisted of five phantoms, with full data provided to the contestants to demonstrate the nature of the phantoms and the data format used in the challenge. The participants were encouraged to create more simulated data based on the example/training data. The challenge test data consisted of 21 phantoms, arranged into seven different difficulty levels, each containing three phantoms. As the difficulty level increased, the shapes of the holes in the phantoms became increasingly complex, non-convex and anisotropic, and the view angle also became increasingly constricted.

Many of the submitted algorithms yielded highly impressive results, recovering the structure of the phantoms with high fidelity in even severely limited angle settings. However, for the most challenging test cases, all the algorithms failed to recover the structure of the phantoms to any significant accuracy, indicating that HTC2022 spanned a broad range of difficulty. A wide variety of well-performing reconstruction approaches were seen in the challenge. In the top three best performing submissions, the winning team used a purely machine learning (ML) based approach, the runner-up used a hybrid approach aiding ML with traditional mathematics, and the second runner-up employed conventional variational regularization. Overall, the HTC2022 submissions demonstrated many promising directions for algorithm development in limited angle tomographic reconstruction.

After HTC2022 had closed, all of the challenge data was published as an open dataset. Both the limited angle sinograms used in the challenge and the respective full angle sinograms were included for the test data. In contrast to many open datasets of CT projection data, the simple binary structure of the phantoms in the HTC2022 dataset allows for a near absolute ground truth reference to be established. This may be useful for algorithm developers in establishing unambiguous image quality metrics for evaluating reconstruction performance while still using real X-ray data in algorithm performance assessment.

7.4 Publication IV

This work focuses on a reformulation of total variation (TV) regularized image reconstruction in 3D cone beam computed tomography (CBCT). In variational TV regularization, an appropriate choice for the value of the regularization parameter α needs to be decided on a case-by-case basis, often requiring many reconstructions to narrow in on a suitable value. The choice is affected by many factors, including data sparsity and measurement noise level. The new method presented in this work is based on three observations. Firstly, the actual numerical value of the regularization parameter α offers very little, if any, interpretability. Secondly, for any TV regularized reconstruction algorithm with guaranteed convergence, the relative sparsity of the reconstruction gradient magnitude will also converge to some number between zero and one. Lastly, it is reasonable to assume that the sparsity of the underlying "true" image remains constant, regardless of the noise level in the acquired measurement data.

A CT reconstruction algorithm named *Total Variation using Controlled Gradient Sparsity* (TV-CGS) is derived and presented. TV-CGS is based on variational regularization, specifically on the TV-NN-PDFP algorithm presented in Chapter 5, but the regularization parameter α is now treated as a dynamic variable whose value can change between

iterations. The value of α is altered during iteration so that the sparsity of the gradient magnitude of the reconstruction converges to an *a priori* defined value $\mathcal{C}\mathrm{pr} \in (0, 1)$. This results in automatic adjustment of the regularization parameter as the noise level or sparsity in the measurement data changes. We assert that this way of defining the regularization strength significantly increases its interpretability. Furthermore, we hypothesize that the gradient sparsity of similar objects in a given category remains approximately constant.

The TV-CGS algorithm was tested using synthetic CBCT data. We generated CBCT projections with realistic photon noise using the 3D Shepp-Logan phantom. For the synthetic data case, the TV-CGS algorithm behaved as predicted: with constant noise level, the gradient sparsities converged to the prior sparsity levels $\mathcal{C}_{\mathrm{pr}}$, and the regularization parameters α converged towards fixed values, in exact reverse order to the prior sparsity levels. One exception was that the extreme lowest gradient sparsity tried did not exhibit stable convergence, implying that outside a wide region of stability the TV-CGS algorithm can behave unpredicably. When a fixed value for $\mathcal{C}_{\mathrm{pr}}$ was used and the noise level was varied, in all cases the gradient magnitude sparsity converged toward the *a priori* defined value, and the regularization parameters were adjusted automatically so that increased noise resulted in a higher value towards which α converged in a stable manner.

The results with simulated data suggest the TV-CGS algorithm can be used for total variation regularized reconstruction in cone beam computed tomography, and further tests with measured data are warranted.

Bibliography

[1] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig*, 69:262–277, 1917.

[2] J. Radon. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4):170–176, 1986. Translated by P. C. Parks from the original German text.

[3] A. M. Cormack. Reconstruction of densities from their projections, with applications in radiological physics. *Physics in Medicine & Biology*, 18(2):195–207, 1973.

[4] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34(9):2722–2727, 1963.

[5] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications, II. *Journal of Applied Physics*, 35(10):2908–2913, 1964.

[6] G. N. Hounsfield. Computerized transverse axial scanning (tomography). Part 1. Description of system. *British Journal of Radiology*, 46(552):1016–1022, 1973.

[7] J. Ambrose. Computerized transverse axial scanning (tomography): Part 2. Clinical application. *British Journal of Radiology*, 46(552):1023–1047, 1973.

[8] The Nobel Foundation. The Nobel Prize in Physiology or Medicine 1979. `https://www.nobelprize.org/prizes/medicine/1979/`. (Accessed: 2024-04-26).

[9] V. R. Fuchs and H. C. Sox. Physicians' views of the relative importance of thirty medical innovations. *Health Affairs*, 20(5):30–42, 2001.

[10] G. D. Rubin. Computed tomography: Revolutionizing the practice of medicine for 40 years. *Radiology*, 273(2 Suppl):S45–74, 2014.

[11] C. H. McCollough and P. S. Rajiah. Milestones in CT: Past, present, and future. *Radiology*, 309(1):e230803, 2023.

[12] M. M. Lell and M. Kachelriess. Recent and upcoming technological developments in computed tomography. *Investigative Radiology*, 55(1):8–19, 2020.

[13] J. Hsieh and T. Flohr. Computed tomography recent history and future perspectives. *Journal of Medical Imaging*, 8(5):052109, 2021.

[14] R. A. Schulz, J. A. Stein, and N. J. Pelc. How CT happened: the early development of medical computed tomography. *Journal of Medical Imaging*, 8(5):052110, 2021.

[15] C. R. Crawford and K. F. King. Computed tomography scanning with simultaneous patient translation. *Medical Physics*, 17(6):967–982, 1990.

[16] W. A. Kalender, W. Seissler, and E. Klotz P. Vock. Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology*, 176(1):181–183, 1990.

[17] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, 53(5):1337—1349, 2002.

[18] R. Fahrig, D. A. Jaffray, I. Sechopoulos, and J. W. Stayman. Flat-panel conebeam CT in the clinic: history and current state. *Journal of Medical Imaging*, 8(5):052115, 2021.

[19] E. J. Morton, S. Webb, J. E. Bateman, L. J. Clarke, and C. G. Shelton. Three-dimensional x-ray microtomography for medical and biological applications. *Physics in Medicine & Biology*, 35(7):805–820, 1990.

[20] A. C. Thompson, J. Llacer, L. Campbell Finman, E.B. Hughes, J.N. Otis, S. Wilson, and H.D. Zeman. Computed tomography using synchrotron radiation. *Nuclear Instruments and Methods in Physics Research*, 222(1):319–323, 1984.

[21] H. Labriet, C. Nemoz, M. Renier, P. Berkvens, T. Brochard, R. Cassagne, H. Elleaume, F. Estève, C. Verry, J. Balosso, J. F. Adam, and E. Brun. Significant dose reduction using synchrotron radiation computed tomography: first clinical case and application to high resolution CT exams. *Scientific Reports*, 8:12491, 2018.

[22] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann. Photon-counting CT: Technical principles and clinical prospects. *Radiology*, 289(2):293–312, 2018.

[23] S. Leng, M. Bruesewitz, S. Tao, K. Rajendran, A. F. Halaweish, N. G. Campeau, J. G. Fletcher, and C. H. McCollough. Photon-counting detector CT: System design and clinical applications of an emerging technology. *RadioGraphics*, 39(3):729–743, 2019.

[24] A. Esquivel, A. Ferrero, A. Mileto, F. Baffour, K. Horst, P. S. Rajiah, A. Inoue, S. Leng, C. McCollough, and J. G. Fletcher. Photon-counting detector CT: Key points radiologists should know. *Korean Journal of Radiology*, 23(9):854–865, 2022.

[25] K. Rajendran, M. Petersilka, A. Henning, E. R. Shanblatt, B. Schmidt, T. G. Flohr, A. Ferrero, F. Baffour, F. E. Diehn, L. Yu, P. Rajiah, J. G. Fletcher, S. Leng, and C. H. McCollough. First clinical photon-counting detector CT system: Technical evaluation. *Radiology*, 303(1):130–138, 2022.

[26] G. van Kaick and S. Delorme. Computed tomography in various fields outside medicine. *European Radiology*, 15(Suppl 4):D74–81, 2005.

[27] E. N. Landis and D. T. Keane. X-ray microtomography. *Materials Characterization*, 61(12):1305–1316, 2010.

[28] R. Mizutani and Y. Suzuki. X-ray microtomography in biology. *Micron*, 43(2):104–115, 2012.

[29] L. De Chiffre, S. Carmignato, J.-P. Kruth, R. Schmitt, and A. Weckenmann. Industrial applications of computed tomography. *CIRP Annals - Manufacturing Technology*, 63(2):655–677, 2014.

[30] M. Beister, D. Kolditz, and W. A. Kalender. Iterative reconstruction methods in X-ray CT. *Physica Medica*, 28(2):94–108, 2012.

[31] L. L. Geyer, U. J. Schoepf, F. G. Meinel, J. W. Nance, G. Bastarrika, J. A. Leipsic, N. S. Paul, M. Rengo, A. Laghi, and C. N. De Cecco. State of the art: Iterative CT reconstruction techniques. *Radiology*, 276(2):339–357, 2015.

[32] S. Ravishankar, J. C. Ye, and J. A. Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1):86––109, 2020.

[33] M. J. Willemink and P. B. Noël. The evolution of image reconstruction for CT — from filtered back projection to artificial intelligence. *European Radiology*, 29:2185–2195, 2019.

[34] P. J. La Riviere and C. R. Crawford. From EMI to AI: a brief history of commercial CT reconstruction algorithms. *Journal of Medical Imaging*, 8(5):052111, 2021.

[35] C. M. McLeavy, M. H. Chunara, R. J. Gravell, A. Rauf, A. Cushnie, C. Staley Talbot, and R. M. Hawkins. The future of CT: deep learning reconstruction. *Clinical Radiology*, 76(6):407–415, 2021.

[36] T. P. Szczykutowicz, G. V. Toia, A. Dhanantwari, and B. Nett. A review of deep learning CT reconstruction: Concepts, limitations, and promise in clinical practice. *Current Radiology Reports*, 10:101–115, 2022.

[37] X. Pan, E. Y. Sidky, and M. Vannier. Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse Problems*, 25(12):123009, 2009.

[38] B. Chen, X. Duan, Z. Yu, S. Leng, L. Yu, and C. McCollough. Technical note: Development and validation of an open data format for CT projection data. *Medical Physics*, 42(12):6964–6972, 2015.

[39] T. R. Moen, B. Chen, D. R. Holmes III, X. Duan, Z. Yu, L. Yu, S. Leng, J. G. Fletcher, and C. H. McCollough. Low-dose CT image and projection dataset. *Medical Physics*, 48(2):902–911, 2021.

[40] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, XIII(4):49–52, 1902.

[41] V. Maz´ya and T. Shaposhnikova. *Jacques Hadamard: A Universal Mathematician*. American Mathematical Society, London Mathematical Society, 1998.

[42] J. L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. SIAM, 2012.

[43] D. Calvetti, F. Pitolli, E. Somersalo, and B. Vantaggi. Bayes meets Krylov: Statistically inspired preconditioners for CGLS. *SIAM Review*, 60(2):429–461, 2018.

[44] D. Poole. *Linear Algebra: A Modern Introduction*. Cengage, fourth edition, 2014.

[45] D. Calvetti and E. Somersalo. Inverse problems: From regularization to Bayesian inference. *WIREs Computational Statistics*, 10(3):e1427, 2018.

[46] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111, 2018.

[47] C. Lanczos. Linear systems in self-adjoint form. *The American Mathematical Monthly*, 65(9):665–679, 1958.

[48] R. J. Hanson. A numerical method for solving Fredholm integral equations of the first kind using singular values. *SIAM Journal on Numerical Analysis*, 8(3):616–622, 1971.

[49] J. M. Varah. On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems. *SIAM Journal on Numerical Analysis*, 10(2):257–267, 1973.

[50] J. M. Varah. A practical examination of some numerical methods for linear discrete ill-posed problems. *SIAM Review*, 21(1):100–111, 1979.

[51] P. C. Hansen. The truncated SVD as a method for regularization. *BIT Numerical Mathematics*, 27:534–553, 1987.

[52] P. C. Hansen. Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank. *SIAM Journal on Scientific and Statistical Computing*, 11(3):503–518, 1990.

[53] F. Natterer. *The Mathematics of Computerized Tomography*. John Wiley & Sons, 1986.

[54] T. Elfving, P. C. Hansen, and T. Nikazad. Semi-convergence properties of Kaczmarz's method. *Inverse Problems*, 30(5):055007, 2014.

[55] F. Benvenuto and C. Campi. A discrepancy principle for the Landweber iteration based on risk minimization. *Applied Mathematics Letters*, 96:1–6, 2019.

[56] L. Bungert, M. Burger, Y. Korolev, and C.-B. Schönlieb. Variational regularisation for inverse problems with imperfect forward operators and general noise models. *Inverse Problems*, 36(12):125014, 2020.

[57] M. Burger. Variational regularization in inverse problems and machine learning. In *European Congress of Mathematics, Portorož, 20–26 June, 2021*, pages 253–275, 2023.

[58] V. A. Morozov. Regularization of incorrectly posed problems and the choice of regularization parameter. *USSR Computational Mathematics and Mathematical Physics*, 6(1):242–251, 1966.

[59] V. A. Morozov. On the solution of functional equations by the method of regularization. *Doklady Mathematics*, 7:414–417, 1966.

[60] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[61] H. W. Engl. Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Journal of Optimization Theory and Applications*, 52(2):209–215, 1987.

[62] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularisation of non-linear ill-posed problems. *Inverse Problems*, 5(4):523–540, 1989.

[63] P. C. Hansen and D. P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.

[64] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, 1998.

[65] K. Kunisch and J. Zou. Iterative choices of regularization parameters in linear inverse problems. *Inverse Problems*, 14(5):1247–1264, 1998.

[66] T. Bonesky. Morozov's discrepancy principle and Tikhonov-type functionals. *Inverse Problems*, 25(1):015015, 2008.

[67] U. Hämarik, R. Palm, and T. Raus. Comparison of parameter choices in regularization algorithms in case of different information about noise level. *Calcolo*, 48:47–59, 2011.

[68] K. Ito, B. Jin, and J. Zou. A new choice rule for regularization parameters in Tikhonov regularization. *Applicable Analysis*, 90(10):1521–1544, 2011.

[69] U. Hämarik, R. Palm, and T. Raus. On minimization strategies for choice of the regularization parameter in ill-posed problems. *Numerical Functional Analysis and Optimization*, 30(9-10):924–950, 2009.

[70] U. Hämarik, R. Palm, and T. Raus. A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level. *Journal of Computational and Applied Mathematics*, 236(8):2146–2157, 2012.

[71] T. Sanders, R. B. Platte, and R. D. Skeel. Effective new methods for automated parameter selection in regularized inverse problems. *Applied Numerical Mathematics*, 152:29–48, 2020.

[72] L. Reichel and G. Rodriguez. Old and new parameter choice rules for discrete ill-posed problems. *Numerical Algorithms*, 63:65–87, 2013.

[73] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, third edition, 2009.

[74] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[75] D. L. Donoho. Compressed sensing. *IEEE Transaction on Information Theory*, 52(4):1289–1306, 2006.

[76] K. Bryan and T. Leise. Making do with less: An introduction to compressed sensing. *SIAM Review*, 55(3):547–566, 2013.

[77] A. Wirgin. The inverse crime, 2004. hal-00001084.

[78] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[79] J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198:493–504, 2007.

[80] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher. Dual- and multi-energy CT: Principles, technical approaches, and clinical applications. *Radiology*, 276(3):637–653, 2015.

[81] M. C. Jacobsen, S. L. Thrower, R. B. Ger, S. Leng, L. E. Court, K. K. Brock, E. P. Tamm, E. N. K. Cressman, D. D. Cody, and R. R. Layman. Multi-energy computed tomography and material quantification: Current barriers and opportunities for advancement. *Medical Physics*, 47(8):3752–3771, 2020.

[82] P. Rajiah, A. Parakh, F. Kay, D. Baruah, A. R.Kambadakone, and S. Leng. Update on multienergy CT: Physics, principles, and applications. *RadioGraphics*, 40(5):1284–1308, 2020.

[83] C. H. McCollough, K. Boedeker, D. Cody, X. Duan, T. Flohr, S. S. Halliburton, J. Hsieh, R. R. Layman, and N. J. Pelc. Principles and applications of multienergy CT: Report of AAPM Task Group 291. *Medical Physics*, 47(7):e881–e912, 2020.

[84] C. H. McCollough. Erratum to Medical physics article 'Principles and applications of multienergy CT: Report of AAPM Task Group 291'. *Medical Physics*, 48(5):2694–2694, 2021.

[85] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1–4):259–268, 1992.

[86] P. Getreuer. Rudin–Osher–Fatemi total variation denoising using split Bregman. *Image Processing On Line*, 2:74–95, 2012.

[87] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19(6):S165–S187, 2003.

[88] A. Pock. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[89] L. Condat. Discrete total variation: New definition and minimization. *SIAM Journal on Imaging Sciences*, 10(3):1258–1290, 2017.

[90] Institute for Pure & Applied Mathematics. Rudin-Osher-Fatemi model captures infinity and beyond. `https://www.ipam.ucla.edu/news/rudin-osher-fatemi-model-captures-infinity-and-beyond/`. 2019-04-15. (Accessed: 2024-03-20).

[91] A. H. Delaney and Y. Bresler. Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography. *IEEE Transactions on Image Processing*, 7(2):204–221, 1998.

[92] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Transactions on Nuclear Science*, 46(6):2202–2210, 1999.

[93] M. Persson, D. Bone, and H. Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine & Biology*, 46(3):853–866, 2001.

[94] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: II. application to dental radiology. *Physics in Medicine & Biology*, 48(10):1465–1490, 2003.

[95] V. Kolehmainen, A. Vanne, S. Siltanen, S. Järvenpää, J. P. Kaipio, M. Lassas, and M. Kalke. Parallelized Bayesian inversion for three-dimensional dental X-ray imaging. *IEEE Transactions on Medical Imaging*, 25(2):218–228, 2006.

[96] E. Y. Sidky, C. M. Kao, and X. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-Ray Science and Technology*, 14(2):119–139, 2006.

[97] G. T. Herman and R. Davidi. Image reconstruction from a small number of projections. *Inverse Problems*, 24(4):045011, 2008.

[98] X. Duan, L. Zhang, Y. Xing, Z. Chen, and J. Cheng. Few-view projection reconstruction with an iterative reconstruction-reprojection algorithm and TV constraint. *IEEE Transactions on Nuclear Science*, 56(3):1377–1382, 2009.

[99] H. Yu and G. Wang. A soft-threshold filtering approach for reconstruction from a limited number of projections. *Physics in Medicine & Biology*, 55(13):3905–3916, 2010.

[100] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelriess. Improved total variation-based CT image reconstruction applied to clinical data. *Physics in Medicine & Biology*, 56(6):1545–1561, 2011.

[101] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang. Low-dose CT reconstruction via edge-preserving total variation regularization. *Physics in Medicine & Biology*, 56(18):5949–5967, 2011.

[102] T. L. Jensen, J. H. Jørgensen, P. C. Hansen, and S. H. Jensen. Implementation of an optimal first-order method for strongly convex total variation regularization. *BIT Numerical Mathematics*, 52:329–356, 2012.

[103] Y. Liu, J. Ma, Y. Fan, and Z. Liang. Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. *Physics in Medicine & Biology*, 57(23):7923—-7956, 2012.

[104] K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen. Sparse tomography. *SIAM Journal on Scientific Computing*, 35(3):B644–B665, 2013.

[105] K. Hämäläinen, L. Harhanen, A. Hauptmann, A. Kallonen, E. Niemi, and S. Siltanen. Total variation regularization for large-scale X-ray tomography. *International Journal of Tomography and Simulation*, 25(1), 2014.

[106] E. Y. Sidky, R. Chartrand, J. M. Boone, and X. Pan. Constrained TpV minimization for enhanced exploitation of gradient sparsity: Application to CT image reconstruction. *IEEE Journal of Translational Engineering in Health and Medicine*, 2:1800418, 2014.

[107] J. S. Jørgensen, C. Kruschel, and D. A. Lorenz. Testable uniqueness conditions for empirical assessment of undersampling levels in total variation-regularized X-ray CT. *Inverse Problems in Science and Engineering*, 23(8):1283–1305, 2015.

[108] J. S. Jørgensen and E. Y. Sidky. How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography. *Philosophical Transactions of the Royal Society A*, 373:20140387, 2015.

[109] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777–4807, 2008.

[110] J. Tang, B. E. Nett, and G. H. Chen. Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms. *Physics in Medicine & Biology*, 54(19):5781–5804, 2009.

[111] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan. Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Physics in Medicine & Biology*, 55(22):6575–6599, 2010.

[112] J. Bian, K. Yang, J. M. Boone, X. Han, E. Y. Sidky, and X. Pan. Investigation of iterative image reconstruction in low-dose breast CT. *Physics in Medicine & Biology*, 59(11):2659–2685, 2014.

[113] A. Cai, L. Wang, H. Zhang, B. Yan, L. Li, X. Xi, M. Guan, and J. Li. 3D alternating direction TV-based cone-beam CT reconstruction with efficient GPU implementation. *Computational and Mathematical Methods in Medicine*, 2014:982695, 2014.

[114] H. C. Lee, B. Song, J. S. Kim, J. J. Jung, H. H. Li, S. Mutic, and J. C. Park. An efficient iterative CBCT reconstruction approach using gradient projection sparse reconstruction algorithm. *Oncotarget*, 7(52):87342–87350, 2016.

[115] C. Yang, P. Wu, S. Gong, J. Wang, Q. Lyu, X. Tang, and T. Niu. Shading correction assisted iterative cone-beam CT reconstruction. *Physics in Medicine & Biology*, 62(22):8495–8520, 2017.

[116] Y. Chen, F. F. Yin, Y. Zhang, Y. Zhang, and L. Ren. Low dose CBCT reconstruction via prior contour based total variation (PCTV) regularization: a feasibility study. *Physics in Medicine & Biology*, 63(8):085014, 2018.

[117] C.de Molina, E. Serrano, J. Garcia-Blas, J. Carretero, M. Desco, and M. Abella. GPU-accelerated iterative reconstruction for limited-data tomography in CBCT systems. *BMC Bioinformatics*, 19:171, 2018.

[118] Y. Chen, F. F. Yin, Y. Zhang, Y. Zhang, and L. Ren. Low dose cone-beam computed tomography reconstruction via hybrid prior contour based total variation regularization (hybrid-PCTV). *Quantitative Imaging in Medicine and Surgery*, 9(7):1214–1228, 2019.

[119] H. W. Tseng, S. Vedanthama, and A. Karellas. Cone-beam breast computed tomography using ultra-fast image reconstruction with onstrained, total-variation minimization for suppression of artifacts. *Physica Medica*, 73:117–124, 2020.

[120] S. Zhi, M. Kachelriess, and X. Mou. High-quality initial image-guided 4D CBCT reconstruction. *Medical Physics*, 47(5):2099–2115, 2020.

[121] H. W. Tseng, A. Karellas, and S. Vedanthama. Cone-beam breast CT using an offset detector: effect of detector offset and image reconstruction algorithm. *Physics in Medicine & Biology*, 67(8), 2022.

[122] M. Sabaté Landman, A. Biguri, S. Hatamikia, R. Boardman, J. Aston, and C.-B. Schönlieb. On Krylov methods for large-scale CBCT reconstruction. *Physics in Medicine & Biology*, 68(15):155008, 2023.

[123] P. Zhang, S. Ren, Y. Liu, Z. Gui, H. Shangguan, Y. Wang, H. Shu, and Y. Chen. A total variation prior unrolling approach for computed tomography reconstruction. *Medical Physics*, 50(5):2816–2834, 2023.

[124] Y. Wang, G. Chen, T. Xi, Z. Bian, D. Zeng, H. Zaidi, J. He, and Jianhua J. Ma. Helical ct reconstruction from sparse-view data through exploiting the 3d anatomical structure sparsity. *IEEE Access*, 9:15200–15211, 2021.

[125] D. S. Rigie and P. J. La Rivière. Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. *Physics in Medicine & Biology*, 60(5):1741–1762, 2015.

[126] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan. An algorithm for constrained one-step inversion of spectral CT data. *Physics in Medicine & Biology*, 61(10):3784–3818, 2016.

[127] B. Chen, Z. Zhang, E. Y. Sidky, D. Xia, and X. Pan. Image reconstruction and scan configurations enabled by optimization-based algorithms in multispectral CT. *Physics in Medicine & Biology*, 62(22):8763–8793, 2017.

[128] D. S. Rigie, A. A. Sanchez, and P. J. La Rivière. Assessment of vectorial total variation penalties on realistic dual-energy CT data. *Physics in Medicine & Biology*, 62(8):3284–3298, 2017.

[129] T. G. Schmidt, R. F. Barber, and E. Y. Sidky. A spectral CT method to directly estimate basis material maps from experimental photon-counting data. *IEEE Transactions on Medical Imaging*, 36(9):1808–1819, 2017.

[130] D. Kazantsev, J. S. Jørgensen, M. S. Andersen, W. R. B. Lionheart, P. D. Lee, and P. J. Withers. Joint image reconstruction method with correlative multi-channel prior for x-ray spectral computed tomography. *Inverse Problems*, 34(6):064001, 2018.

[131] S. Niu, Y. Zhang, Y. Zhong, G. Liu, S. Lu, X. Zhang, S. Hu, T. Wang, G. Yu, and J. Wang. Iterative reconstruction for photon-counting CT using prior image constrained total generalized variation. *Inverse Problems*, 103:167–182, 2018.

[132] M. A. K. Juntunen, P. Sepponen, K. Korhonen, V.-M. Pohjanen, J. Ketola, A. Kotiaho, M. T. Nieminen, and S. I. Inkinen. Interior photon counting computed tomography for quantification of coronary artery calcium: pre-clinical phantom study. *Biomedical Physics & Engineering Express*, 6:055011, 2020.

[133] M. A. K. Juntunen, A. O. Kotiaho, M. T. Nieminen, and S. I. Inkinen. Optimizing iterative reconstruction for quantification of calcium hydroxyapatite with photon counting flat-detector computed tomography: a cardiac phantom study. *Journal of Medical Imaging*, 8(5):052102, 2021.

[134] R. Warr, E. Ametova, R. J. Cernik, G. Fardell, S. Handschuh, J. S. Jørgensen, E. Papoutsellis, E. Pasca, and P. J. Withers. Enhanced hyperspectral tomography for bioimaging by spatiospectral reconstruction. *Scientific Reports*, 11:20818, 2021.

[135] T. G. Schmidt, B. A. Sammut, R. F. Barber, X. Pan, and E. Y. Sidky. Addressing CT metal artifacts using photon-counting detectors and one-step spectral CT image reconstruction. *Medical Physics*, 49(5):3021–3040, 2022.

[136] Y. He, L. Zeng, Q. Xu, Z. Wang, H. Yu, Z. Shen, Z. Yang, and R. Zhou. Spectral CT reconstruction via low-rank representation and structure preserving regularization. *Physics in Medicine & Biology*, 68:025011, 2023.

[137] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing Using MATLAB®*. McGraw Hill, second edition, 2009.

[138] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Pearson Education, third edition, 2008.

[139] A. Chambolle, S. E. Levine, and B. J. Lucier. An upwind finite-difference method for total variation–based image smoothing. *SIAM Journal on Imaging Sciences*, 4(1):277–299, 2011.

[140] R. Abergel and L. Moisan. The shannon total variation. *Journal of Mathematical Imaging and Vision*, 59:341–370, 2017.

[141] X. Hou, Y. Teng, Y. Kang, and S. Qi. A separable quadratic surrogate total variation minimization algorithm for accelerating accurate CT reconstruction from few-views and limited-angle data. *Medical Physics*, 45(2):535–548, 2018.

[142] Y. Liu, Y. Liu, P. Zhang, R. Yan, L. Wang, W. Liu, and Z. Gui. Cone-beam computed tomography based on truncated adaptive-weight total variation. *NDT & E International*, 133:102755, 2023.

[143] Z. Chen, X. Jin, L. Li, and G. Wang. A limited-angle CT reconstruction method based on anisotropic TV minimization. *Physics in Medicine & Biology*, 58(7):2119, 2013.

[144] K. Niinimäki, M. Lassas, K. Hämäläinen, A. Kallonen, V. Kolehmainen, E. Niemi, and S. Siltanen. Multiresolution parameter choice method for total variation regularized tomography. *SIAM Journal of Imaging Sciences*, 9(3):938–974, 2016.

[145] D.-J. Ji, G.-R. Qu, C.-H. Hu, B.-D. Liu, J.-B. Jian, and X.-K. Guo. Anisotropic total variation minimization approach in in-line phase-contrast tomography and its application to correction of ring artifacts. *Chinese Physics B*, 26(6):060701, 2017.

[146] Z. Sun, D. Ji, and S. Yang. Anisotropic total variation minimization approach in in-line phase-contrast tomography. In *The Fourth International Symposium on Image Computing and Digital Medicine*, pages 10–13, 2021.

[147] P. Blomgren and T. F. Chan. Color TV: total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[148] M. J. Ehrhardt and M. M. Betcke. Multicontrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences*, 9(3):1084–1106, 2016.

[149] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[150] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[151] C. T. Kelley. *Iterative Methods for Optimization*. SIAM, 1999.

[152] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[153] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, 1993.

[154] J. Xu and F. Noo. Convex optimization algorithms in medical image reconstruction—in the age of AI. *Physics in Medicine & Biology*, 67(7):07TR01, 2022.

[155] S. Wang, G. Cao, Y. Wang, S. Liao, Q. Wang, J. Shi, C. Li, and D. Shen. Review and prospect: Artificial intelligence in advanced medical imaging. *Frontiers in Radiology*, 1, 2021.

[156] J. A. Fessler. Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms. *IEEE Signal Processing Magazine*, 37(1):33–40, 2020.

[157] J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3):502–506, 1953.

[158] M. Avriel and D. J. Wilde. Optimality proof for the symmetric Fibonacci search technique. *The Fibonacci Quarterly*, 4:265–269, 1966.

[159] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.

[160] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988.

[161] M. Raydan. On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, 13(3):321–326, 1993.

[162] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization*, 7(1):26–33, 1997.

[163] N. Agrawal. Improve your gradient descent: The epic quest for the optimal stride. *Towards Data Science*. `https://towardsdatascience.com/improve-your-gradient-descent-the-epic-quest-for-the-optimal-stride-4711dc6f5dba`, 2023. (Accessed: 2024-06-13).

[164] S. Ruder. An overview of gradient descent optimization algorithms. `https://www.ruder.io/optimizing-gradient-descent/`, 2016. (Accessed: 2024-06-13).

[165] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.

[166] J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain, edition $1\frac{1}{4}$. Technical report, Carnegie Mellon University, 1994.

[167] W. W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization*, 2:35–58, 2006.

[168] E. Polak and G. Ribière. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle*, 3:35–43, 1969.

[169] B. T. Polayk. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, 9(4):94–112, 1969. Translated by D. E. Brown.

[170] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer New York, 2011.

[171] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

[172] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists. *SIAM Review*, 65(2):375–435, 2023.

[173] 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*, 2016:54, 2016.

[174] A. Gramfort, M. Kowalski, and M. Hämäläinen. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine & Biology*, 57(7):1937–1961, 2012.

[175] P. Chen, J. Huang, and X. Zhang. A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29:025011, 2013.

[176] Core Imaging Library (CIL) documentation: Optimisation framework. `https://tomographicimaging.github.io/CIL/nightly/optimisation/`. (Accessed: 2024-02-15).

[177] T. M. Buzug. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer, 2008.

[178] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.

[179] L. A. Shepp and B. F. Logan. The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974.

[180] J. A. Seibert, J. M. Boone, and K. K. Lindfors. Flat-field correction technique for digital detectors. In *Medical Imaging 1998: Physics of Medical Imaging*, volume 3336, pages 348–354. SPIE, 1998.

[181] V. Van Nieuwenhove, J. De Beenhouwer, F. De Carlo, L. Mancini, F. Marone, , and J. Sijbers. Dynamic intensity normalization using eigen flat fields in x-ray imaging. *Optics Express*, 23:27975–27989, 2015.

[182] F. E. Boas and D. Fleischmann. CT artifacts: Causes and reduction techniques. *Imaging in Medicine*, 4(2):229–240, 2012.

[183] R. A. Brooks and G. Di Chiro. Beam hardening in X-ray reconstructive tomography. *Physics in Medicine & Biology*, 21(3):390–398, 1976.

[184] H. S. Park, Y. E. Chung, and J. K. Seo. Computed tomographic beam-hardening artefacts: mathematical characterization and analysis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2043):20140388, 2015.

[185] L. Gjesteby, B. De Man, Y. Jin, H. Paganetti, J. Verburg, D. Giantsoudi, and G. Wang. Metal artifact reduction in CT: Where are we after four decades? *IEEE Access*, 4:5826–5849, 2016.

[186] M. Danielsson, M. Persson, and M. Sjölin. Photon-counting x-ray detectors for CT. *Physics in Medicine & Biology*, 66(3):03TR01, 2021.

[187] J. van der Bie, M. van Straten, R. Booij, D. Bos, M. L. Dijkshoorn, A. Hirsch, S. P. Sharma, E. H.G. Oei, and R. P. J. Budde. Photon-counting CT: Review of initial clinical results. *European Journal of Radiology*, 163:110829, 2023.

[188] C. Mory, B. Sixou, S. Si-Mohamed, L. Boussel, and S. Rit. Comparison of five one-step reconstruction algorithms for spectral CT. *Physics in Medicine & Biology*, 63(23):235001, 2018.

[189] Q. Wang, W. Wu, S. Deng, Y. Zhu, and H. Yu. Locally linear transform based three-dimensional gradient l_0-norm minimization for spectral CT reconstruction. *Medical Physics*, 47(10):4810–4826, 2020.

[190] T. Sartoretti, A. Landsmann, D. Nakhostin, M. Eberhard, C. Roeren, V. Mergen, K. Higashigaito, R. Raupach, H. Alkadhi, and A. Euler. Quantum iterative reconstruction for abdominal photon-counting detector CT improves image quality. *Radiology*, 303(2):339–348, 2022.

[191] L. Shen, Y. Xing, and L. Zhang. Joint reconstruction and spectrum refinement for photon-counting-detector spectral CT. *IEEE Transactions on Medical Imaging*, 42(9):2653–2665, 2023.

[192] P.-A. Rodesch, S. A. Si-Mohamed, J. Lesaint, P. C. Douek, and S. Rit. Image quality improvement of a one-step spectral CT reconstruction on a prototype photon-counting scanner. *Physics in Medicine & Biology*, 69(1):015005, 2024.

[193] J. Jakubek. Data processing and image reconstruction methods for pixel detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 576(1):223–234, 2007.

[194] R. E. Alvarez. Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis. *Medical Physics*, 38(5):2324–2334, 2011.

[195] A. Meaney. Design and construction of an X-ray computed tomography imaging system. Master's thesis, University of Helsinki, 2015.

[196] M. E. Davison. The ill-conditioned nature of the limited angle tomography problem. *SIAM Journal on Applied Mathematics*, 43(2):428–448, 1983.

[197] G.-H. Chen, J. Tang, and S. Leng. Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical Physics*, 35(2):660–663, 2008.

[198] H. Kudo, T. Suzuki, and E. A. Rashed. Image reconstruction for sparse-view CT and interior CT—introduction to compressed sensing and differentiated backprojection. *Quantitative Imaging in Medicine and Surgery*, 3(3):147–161, 2013.

[199] K. Mei, F. K. Kopp, R. Bippus, T. Köhler, B. J. Schwaiger, A. S. Gersing, A. Fehringer, A. Sauter, D. Münzel, F. Pfeiffer, E. J. Rummeny, J. S. Kirschke, P. B. Noël, and T. Baum. Is multidetector CT-based bone mineral density and quantitative bone microstructure assessment at the spine still feasible using ultra-low tube current and sparse sampling? *European Radiology*, 27(12):5261–5271, 2017.

[200] A. P. Sauter, F. K. Kopp, R. Bippus, J. Dangelmaier, D. Deniffel, A. A. Fingerle, Felix Meurer, D. Pfeiffer, R. Proksa, E. J. Rummeny, and P. B. Noël. Sparse sampling computed tomography (SpSCT) for detection of pulmonary embolism: a feasibility study. *European Radiology*, 29(11):5950–5960, 2019.

[201] E. Y. Sidky and X. Pan. Report on the AAPM deep-learning sparse-view CT grand challenge. *Medical Physics*, 49(8):4935–4943, 2021.

[202] J. T. Dobbins III and D. J. Godfrey. Digital x-ray tomosynthesis: current state of the art and clinical potential. *Physics in Medicine & Biology*, 48(19):R65–R106, 2003.

[203] T. Wu, R. H. Moore, E. A. Rafferty, and D. B. Kopans. A comparison of reconstruction algorithms for breast tomosynthesis. *Medical physics*, 31(9):2636–2647, 2004.

[204] S. Vedantham, A. Karellas, G. R. Vijayaraghavan, and D. B. Kopans. Digital breast tomosynthesis: state of the art. *Radiology*, 277(3):663–684, 2015.

[205] E. Loli Piccolomini and E. Morotti. A fast Total Variation-based iterative algorithm for digital breast tomosynthesis image reconstruction. *Journal of Algorithms & Computational Technology*, 10(4):277–289, 2016.

[206] G. Landi, E. Loli Piccolomini, and J. G. Nagy. A limited memory BFGS method for a nonlinear inverse problem in digital breast tomosynthesis. *Inverse Problems*, 33(9):095005, 2017.

[207] G. Landi, E. Loli Piccolomini, and J. G. Nagy. Nonlinear conjugate gradient method for spectral tomosynthesis. *Inverse Problems*, 35(9):094003, 2019.

[208] S. M. Mauriello, A. M. Broome, E. Platin, A. Mol, C. Inscoe, J. Lu, O. Zhou, and K. Moss. The role of stationary intraoral tomosynthesis in reducing proximal overlap in bitewing radiography. *Dentomaxillofacial Radiology*, 49(8):20190504, 2020.

[209] J. W. Casselman, K. Gieraerts, D. Volders, J. Delanote, K. Mermuys, B. De Foer, and G. Swennen. Cone beam CT: non-dental applications. *Journal of the Belgian Society of Radiology*, 96(6):333–353, 2013.

[210] T. Kaasalainen, M. Ekholm, T. Siiskonen, and M. Kortesniemi. Dental cone beam CT: An updated review. *Physica Medica*, 88:193–217, 2021.

[211] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984.

[212] H. K. Tuy. An inversion formula for cone-beam reconstruction. *SIAM Journal on Applied Mathematics*, 43(3):546–552, 1983.

[213] B. D. Smith. Image reconstruction from cone-beam projections: Necessary and sufficient conditions and reconstruction methods. *IEEE Transactions on Medical Imaging*, 4(1):14–25, 1985.

[214] X. Tang, E. A. Krupinski, H. Xie, and A. E. Stillman. On the data acquisition, image reconstruction, cone beam artifacts, and their suppression in axial MDCT and CBCT – a review. *Medical Physics*, 45(9):e761–e782, 2018.

[215] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3):471–481, 1970.

[216] Peter Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1):105–117, 1972.

[217] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

[218] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[219] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

[220] G. Prieto Renieblas, A. Turrero Nogués, A. Muñoz González, N. Gómez-León, and E. Guibelalde del Castillo. Structural similarity index family for image quality assessment in radiological images. *Journal of Medical Imaging*, 4(3):035501, 2017.

[221] J.-F. Pambrun and R. Noumeir. Limitations of the SSIM quality metric in the context of diagnostic imaging. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 2960–2963, 2015.

[222] A. Mason, J. Rioux, S. E. Clarke, A. Costa, M. Schmidt, V. Keough, T. Huynh, and S. Beyea. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE Transactions on Medical Imaging*, 39(4):1064–1072, 2020.

[223] S. Kastryulin, J. Zakirov, N. Pezzotti, and D. V. Dylov. Image quality assessment for magnetic resonance imaging. *IEEE Access*, 11:14154–14168, 2023.

[224] E. J. I. Hoeijmakers, B. Martens, B. M. F. Hendriks, C. Mihl, R. L. Miclea, W. H. Backes, J. E. Wildberger, F. M. Zijta, H. A. Gietema, P. J. Nelemans, and C. R. L. P. N. Jeukens. How subjective CT image quality assessment becomes surprisingly reliable: pairwise comparisons instead of Likert scale. *European Radiology*, pages 1–10, 2024.

[225] L. S. Chow and R. Paramesran. Review of medical image quality assessment. *Biomedical Signal Processing and Control*, 27:145–154, 2016.

[226] L. Zhang, C. Cavaro-Ménard, and P. Le Callet. An overview of model observers. *IRBM*, 35(4):214–224, 2014.

[227] F. R. Verdun, D. Racine, J. G. Ott, M. J. Tapiovaara, P. Toroi, F. O. Bochud, W. J. H. Veldkamp, A. Schegerer, R. W. Bouwman, I. Hernandez Giron, N. W. Marshall, and S. Edyvean. Image quality in CT: From physical measurements to model observers. *Physica Medica*, 31(8):823–843, 2015.

[228] C. H. McCollough, G. H. Chen, W. Kalender, S. Leng, E. Samei, K. Taguchi, G. Wang, L. Yu, and R. I. Pettrigrew. Achieving routine submillisievert CT scanning: Report from the summit on management of radiation dose in CT. *Radiology*, 264(2):567–580, 2012.

[229] A. Meaney. Cone-beam computed tomography dataset of a walnut imaged at 4 different dose levels. `https://zenodo.org/doi/10.5281/zenodo.11397265`, 2024. (Accessed: 2024-05-31).

[230] Erler-Zimmer. X-ray phantom head with cervical vertebrae, transparent. `https://erler-zimmer.de/en/X-ray-phantom-head-with-cervical-vertebrae-transparent/7300`. (Accessed: 2024-04-26).

[231] Planmeca. Planmeca Viso®. `https://www.planmeca.com/imaging/3d-imaging/planmeca-viso/`. (Accessed: 2024-04-26).

[232] The MathWorks Inc. Matlab version: 23.2 (r2023b). `https://www.mathworks.com`, 2023.

[233] A. Meaney. HelTomo - Helsinki Tomography Toolbox. `https://se.mathworks.com/matlabcentral/fileexchange/74417-heltomo-helsinki-tomography-toolbox`, 2022. (Accessed: 2024-04-26).

[234] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers. The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, 157:35–47, 2015.

[235] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers. Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics Express*, 24(22):25129–25147, 2016.

[236] E. van den Berg and M. P. Friedlander. Spot – a linear-operator toolbox, version 1.2. `https://www.cs.ubc.ca/labs/scl/spot/index.html`, 2013. (Accessed: 2024-02-09).

[237] H. Yan, X. Wang, F. Shi, T. Bai, M. Folkerts, L. Cervino, S. B. Jiang, and X. Jia. Towards the clinical implementation of iterative low-dose cone-beam CT reconstruction in image-guided radiation therapy: Cone/ring artifact correction and multiple GPU implementation. *Medical Physics*, 41(11):111912, 2014.

[238] S. Aootaphao, S. S. Thongvigitmanee, P. Puttawibul, and P. Thajchayapong. Truncation effect reduction for fast iterative reconstruction in cone-beam CT. *BMC Medical Imaging*, 22:160, 2022.

[239] J. Hsieh, E. Liu, B. Nett, J. Tang, J.-B. Thibault, and S. Sahney. A new era of image reconstruction: TrueFidelity^{TM} technical white paper on deep learning image reconstruction, 2019.

[240] The ASTRA Toolbox. `https://astra-toolbox.com/`. (Accessed: 2024-04-24).

[241] A. Maier, H. G. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J.-H. Choi, C. Riess, A. Keil, and R. Fahrig. CONRAD—A software framework for cone-beam imaging in radiology. *Medical Physics*, 40(11):111914, 2013.

[242] CONRAD. `https://www5.cs.fau.de/conrad/`. (Accessed: 2024-04-24).

[243] J. S. Jørgensen, E. Ametova, G. Burca, G. Fardell, E. Papoutsellis, E. Pasca, K. Thielemans, M. Turner, R. Warr, W. R. B. Lionheart, and P. J. Withers. Core Imaging Library - Part I: a versatile Python framework for tomographic imaging. *Philosophical Transactions of the Royal Society A*, 379(2204):20200192, 2021.

[244] E. Papoutsellis, E. Ametova, C. Delplancke, G. Fardell, J. S. Jørgensen, E. Pasca, M. Turner, R. Warr, W. R. B. Lionheart, and P. J. Withers. Core Imaging Library - Part II: multichannel reconstruction for dynamic and spectral tomography. *Philosophical Transactions of the Royal Society A*, 379(2204):20200193, 2021.

[245] The Core Imaging Library. `https://ccpi.ac.uk/cil/`. (Accessed: 2024-04-24).

[246] D. Frank, J. Jelten, and T. Lasser. elsa: an elegant framework for tomographic reconstruction. *Journal of Open Source Software*, 9(94):6174, 2024.

[247] elsa - an elegant framework for tomographic reconstruction. `https://ciip.cit.tum.de/elsadocs/`. (Accessed: 2024-04-24).

[248] J. Hoffman, S. Young, F. Noo, and M. McNitt-Gray. Technical note: FreeCT_wFBP: A robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT. *Medical Physics*, 43(3):1411–1420, 2016.

[249] J. M. Hoffman, F. Noo, S. Young, S. S. Hsieh, and M. McNitt-Gray. Technical note: FreeCT_ICD: An open-source implementation of a model-based iterative reconstruction method using coordinate descent optimization for CT imaging investigations. *Medical Physics*, 45(8):3591–3603, 2018.

[250] FreeCT. `https://gitlab.com/freect/freect`. (Accessed: 2024-04-26).

[251] D. Micieli, T. Minniti, and G. Gorini. NeuTomPy toolbox, a Python package for tomographic data processing and reconstruction. *SoftwareX*, 9:260–264, 2019.

[252] Neutompy toolbox. `https://neutompy-toolbox.readthedocs.io/`. (Accessed: 2024-04-24).

[253] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). *Journal of Physics: Conference Series*, 489(1):012079, 2014.

[254] The Reconstruction Toolkit (RTK). `https://www.openrtk.org/`. (Accessed: 2024-04-26).

[255] T. Balke, F. Davis, C. Garcia-Cardona, S. Majee, M. McCann, L. Pfister, and B. Wohlberg. Scientific computational imaging code (scico). *Journal of Open Source Software*, 7(78):4722, 2022.

[256] Scientific Computational Imaging Code (SCICO). `https://scico.readthedocs.io/`. (Accessed: 2024-04-26).

[257] A. Biguri, M. Dosanjh, S. Hancock, and M. Soleimani. TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction. *Biomedical Physics & Engineering Express*, 2:055010, 2016.

[258] TIGRE: Tomographic iterative GPU-based reconstruction toolbox. `https://github.com/CERN/TIGRE`. (Accessed: 2024-04-26).

[259] Tomopedia - A Wiki for tomography-related algorithms and software packages. `https://tomopedia.github.io/`. (Accessed: 2024-04-26).

[260] H. Yu, S. Zhao, and G. Wang. A differentiable Shepp–Logan phantom and its applications in exact cone-beam CT. *Physics in Medicine & Biology*, 50(23):5583–5595, 2005.

[261] C. G. Koay, J. E. Sarlls, and E. Özarslan. Three-dimensional analytical magnetic resonance imaging phantom in the Fourier domain. *Magnetic Resonance in Medicine*, 58(2):430–436, 2007.

[262] Matthias Schabel. 3D Shepp-Logan phantom. `https://www.mathworks.com/matlabcentral/fileexchange/9416-3d-shepp-logan-phantom`. MATLAB Central File Exchange. (Accessed: 2024-04-26).

[263] W. P. Segars and B. M. W. Tsui. MCAT to XCAT: The evolution of 4-D computerized phantoms for imaging research. *Proceedings of the IEEE*, 97:1954–1968, 2009.

[264] W. Kainz, E. Neufeld, W. E. Bolch, C. G. Graff, C. H. Kim,, N. Kuster, B. Lloyd, T. Morrison, P. Segars, Y. S. Yeom, M. Zankl, X. G. Xu, and B. M. W. Tsui. Advances in computational human phantoms and their applications in biomedical engineering—a topical review. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(1):1–23, 2019.

[265] X. Jia, H. Yan, L. Cerviño, M. Folkerts, and S. B. Jiang. A GPU tool for efficient, accurate, and realistic simulation of cone beam CT projections. *Medical Physics*, 39(12):7368–7378, 2012.

[266] E. Abadi, B. P. Harrawood, S. Sharma, A. J. Kapadia, W. P. Segars, and E. Samei. DukeSim: A realistic, rapid, and scanner-specific simulation framework in computed tomography. *IEEE Transactions on Medical Imaging*, 38:1457–1465, 2019.

[267] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui. Realistic CT simulation using the 4D XCAT phantom. *Medical Physics*, 35(8):3800–3808, 2008.

[268] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui. 4D XCAT phantom for multimodality imaging research. *Medical Physics*, 37(9):4902–4915, 2010.

[269] Nature Editorial Announcement. Data-access practices strengthened. *Nature*, 515:312, 2014.

[270] Nature Editorial Announcement. Where are the data? *Nature*, 537:138, 2016.

[271] J. F. Williamson, S. K. Das, M. S. Goodsitt, and J. O. Deasy. Introducing the medical physics dataset article. *Medical Physics*, 44(2):349–350, 2017.

[272] J. S. Jørgensen, S. B. Coban, W. R. B. Lionheart, S. A. McDonald, and P. J. Withers. SparseBeads data: benchmarking sparsity regularized computed tomography. *Measurement Science and Technology*, 28:124005, 2017.

[273] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. B. Coban, and K. J. Batenburg. A cone-beam X-ray computed tomography data collection designed for machine learning. *Scientific Data*, 6:215, 2019.

[274] F. De Carlo, D. Gü"rsoy, D. J. Ching, K. J. Batenburg, W. Ludwig, L. Mancini, F. Marone, R. Mokso, D. M. Pelt, J. Sijbers, and M. Rivers. TomoBank: a tomographic data repository for computational x-ray science. *Measurement Science and Technology*, 29(3):034004, 2018.

[275] TomoBank. `https://tomobank.readthedocs.io/`. (Accessed: 2024-04-26).

[276] A. Meaney, F. S. Moura, M. Juvonen, and S. Siltanen. Helsinki Tomography Challenge 2022 (HTC2022) open tomographic dataset. `https://zenodo.org/doi/10.5281/zenodo.6937615`, 2022. (Accessed: 2024-04-26).

[277] C. H. McCollough, A. C. Bartley, R. E. Carter, B. Chen, T. A. Drees, P. Edwards, D. R. Holmes III, A. E. Huang, F. Khan, S. Leng, K. L. McMillan, G. J. Michalak, K. M. Nunez, L. Yu, and J. G. Fletcher. Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge. *Medical Physics*, 44(10):e339–e352, 2017.

[278] A. Shieh. SPARE challenge: The sparse-view reconstruction challenge for four-dimensional cone-beam CT. `https://image-x.sydney.edu.au/spare-challenge/`, 2018. (Accessed: 2023-06-13).

[279] C. C. Shieh, Y. Gonzalez, B. Li, X. Jia, S. Rit, C. Mory, M. Riblett, G. Hugo, Y. Zhang, Z. Jiang, X. Liu, L. Ren, and P. Keall. SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan. *Medical Physics*, 46(9):3799–3811, 2019.

[280] J. Leuschner, M. Schmidt, and A. Denker. LoDoPaB-CT Challenge. https://lodopab.grand-challenge.org/, 2020. (Accessed: 2024-04-26).

[281] J. Leuschner, M. Schmidt, D. Otero Baguer, and P. Maass. LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction. *Scientific Data*, 8:109, 2021.

[282] American Association of Physicists in Medicine. Deep Learning for Inverse Problems: Spectral Computed Tomography Image Reconstruction (DL-spectral CT) - An AAPM Grand Challenge. https://www.aapm.org/GrandChallenge/DL-spectral-CT/, 2022. (Accessed: 2024-04-26).

[283] American Association of Physicists in Medicine. Truth-Based CT (TrueCT) Reconstruction Challenge - An AAPM Grand Challenge. https://www.aapm.org/GrandChallenge/TrueCT/, 2022. (Accessed: 2024-04-26).

[284] American Association of Physicists in Medicine. CT Metal Artifact Reduction (CT-MAR) - An AAPM Grand Challenge. https://www.aapm.org/GrandChallenge/CT-MAR/, 2024. (Accessed: 2024-04-26).

[285] A. Meaney, J. Toivanen, S. Siltanen, and V. Kolehmainen. Multienergy fan beam computed tomography dataset of a bird chest imaged with 3 different X-ray spectra. https://zenodo.org/records/11388131, 2024. (Accessed: 2024-05-31).

[286] T. Heikkilä, H. Help, and A. Meaney. Cone-beam gel phantom data for dynamic X-ray tomography. https://zenodo.org/doi/10.5281/zenodo.7739984, 2023. (Accessed: 2024-04-26).

[287] A. Meaney. Cone-beam computed tomography dataset of a chicken bone imaged at 4 different dose levels. https://zenodo.org/doi/10.5281/zenodo.6990763, 2022. (Accessed: 2024-04-26).

[288] A. Meaney. Cone-beam computed tomography dataset of a walnut. https://zenodo.org/doi/10.5281/zenodo.4279548, 2022. (Accessed: 2024-04-26).

[289] A. Meaney. Cone-beam computed tomography dataset of a pine cone. https://zenodo.org/doi/10.5281/zenodo.4279612, 2022. (Accessed: 2024-04-26).

[290] E. Kamutta, S. Mäkinen, and A. Meaney. Cone-beam computed tomography dataset of a seashell. https://zenodo.org/doi/10.5281/zenodo.6979744, 2022. (Accessed: 2024-04-26).

[291] T. Heikkilä, H. Help, and A. Meaney. Gel phantom data for dynamic X-ray tomography. https://zenodo.org/doi/10.5281/zenodo.3696816, 2020. (Accessed: 2024-04-26).

[292] T. Heikkilä, H. Help, and A. Meaney. Gel phantom data for dynamic X-ray tomography. https://arxiv.org/abs/2003.02841, 2020. arXiv documentation. (Accessed: 2024-04-26).

[293] S.-M. Latva-Äijö, A. Meaney, and S. Siltanen. Tomographic X-ray data of time-dependent 3D cross phantom. https://zenodo.org/doi/10.5281/zenodo.1341456, 2018. (Accessed: 2024-04-26).

[294] S.-M. Latva-Äijö, A. Meaney, and S. Siltanen. Tomographic X-ray data of time-dependent 3D cross phantom. `https://arxiv.org/abs/1809.00166`, 2018. (arXiv documentation. Accessed: 2024-04-26).

[295] T. Bubba, M. Juvonen, J. Lehtonen, M. März, A. Meaney, Z. Purisha, and S. Siltanen. Tomographic X-ray data of carved cheese. `https://zenodo.org/doi/10.5281/zenodo.1254209`, 2017. (Accessed: 2024-04-26).

[296] T. Bubba, M. Juvonen, J. Lehtonen, M. März, A. Meaney, Z. Purisha, and S. Siltanen. Tomographic X-ray data of carved cheese. `https://arxiv.org/abs/1705.05732`, 2017. arXiv documentation. (Accessed: 2024-04-26).

[297] Z. Purisha, A. Meaney, and S. Siltanen. Tomographic X-ray data of 3D emoji. `https://zenodo.org/doi/10.5281/zenodo.1183531`, 2017. (Accessed: 2024-04-26).

[298] Z. Purisha, A. Meaney, and S. Siltanen. Tomographic X-ray data of 3D emoji. `https://arxiv.org/abs/1802.09397`, 2017. arXiv documentation. (Accessed: 2024-04-26).

[299] Finnish Inverse Problems Society. Open X-ray tomographic datasets. `https://fips.fi/category/open-datasets/`. (Accessed: 2024-12-12).

[300] M. J. Ehrhardt and S. R. Arridge. Vector-valued image processing by parallel level sets. *IEEE Transactions on Image Processing*, 23(1):9–18, 2014.

[301] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[302] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[303] W. Yin. Analysis and generalizations of the linearized Bregman method. *SIAM Journal on Imaging Sciences*, 3(4):856–877, 2010.

[304] F. Aurenhammer, R. Klein, and D. T. Lee. *Voronoi Diagrams And Delaunay Triangulations*. World Scientific Publishing Company, 2013.