Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Department of Information Technology
and Electrical Engineering

Diss. ETH No. 23945

Fast reconstruction methods for underconstrained datasets of synchrotron-based tomographic microscopy

A thesis submitted to attain the degree of
DOCTOR of SCIENCES of ETH ZURICH
(Dr.sc. ETH Zurich)

presented by

Filippo Arcadu

MSc., University of Florence, Italy
born on the 24th August 1987
citizen of Italy

accepted on the recommendation of

Examiner: Prof. Dr. Marco Stampanoni
Co-examiner: Prof. Dr. Jean-Philippe Thiran
Co-examiner: Prof. Dr. Joost Batenburg
Co-examiner: Dr. Federica Marone

2016

Dedicated to my parents.

"Frustra fit per plura quod fieri potest per pauciora."
William of Ockham,

"True learning must not be content with ideas, which are, in fact, signs,
but must discover things in their individual truth."
Umberto Eco, The Name of the Rose

Abstract

Synchrotron radiation based X-ray tomographic microscopy (SRXTM) allows non-destructive 3D investigation of a large variety of samples at micrometer and submicrometer scale in few minutes. Several pioneering studies in the field of biology, material science, chemistry and paleontology were based on data acquired with this X-ray imaging technique.

If the specimen under study is either sensitive to X-ray radiation or fast-evolving, the dose irradiated to the sample or the total scan time are bounded to be under a certain limit, respectively. To fulfill this constraint, the exposure time per projection and the total number of views have to be reduced, leading in most cases to the acquisition of noisy, undersampled datasets, labeled as "underconstrained", that present a major challenge for analytical tomographic reconstruction.

This thesis focuses on the development of methods and strategies to address the reconstruction of underconstrained SRXTM absorption-contrast, in-line phase-contrast and differential phase-contrast datasets, acquired in low-dose scans of standard and full interior tomography. Considering the large amount of raw data (several terabytes) created by high-throughput SRXTM acquisitions and the heterogeneity of the samples investigated with this imaging technique, the guiding thread of this work was designing algorithmical solutions characterized by: (i) a good trade-off between reconstruction accuracy and computational efficiency; (ii) versatility in dealing with different types of specimens and imaging modalities; (iii) a minimum amount of supervised a-priori knowledge and required input hyper-parameters to foster the practicality of the tomographic reconstruction. Several standard and state-of-the-art iterative methods, already published in literature and proved effective for medical or non-medical underconstrained datasets, had to be discarded due to the lack of versatility, computational efficiency or "user-friendliness".

To guarantee computational efficiency and the flexibility necessary to address a large variety of different datasets and imaging modalities, the alternate direction method of multipliers plug-and-play (ADMP) has been identified as one of the most suitable iterative reconstruction scheme. The ADMP provides reconstructions with satisfactory trade-off between contrast and resolution after very few iterations. At the same time, it allows the direct incorporation of any forward tomographic operator and any denoising method for the problem regularization.
The computational efficiency of the ADMP has been greatly boosted by the usage of the gridding projectors with minimal oversampling. The low complexity of these Fourier operators enables very fast iterative reconstructions on a single core, with speed comparable to what can be achieved with tomographic operators implemented on graphics processing units. The ADMP working with gridding projectors and parallel beam geometry is computationally suited to run on central processing unit (CPU) clusters.
Here, it is also shown that the degree of coupling between the implementations of the forward and backprojector plays a crucial role in the performance of an iterative algorithm: the absence of coupling between the tomographic operators leads in the best case scenario to sub-optimal accuracy and slower convergence rates, in the worst case scenario it causes the algorithm to diverge or reconstructions to be affected by severe artifacts.

Furthermore, a procedure, called "virtual strategy", is proposed to perform efficient iterative reconstruction of full interior tomographic datasets. The virtual strategy transforms the interior dataset into a standard one, avoiding to use differentiation or edge-padding throughout the entire iterative reconstruction process, which reduces the amount of computations and required memory.
The Helgason-Ludwig sinogram filter is, finally, introduced as a fast unsupervised method to boost the accuracy of analytical reconstructions of strongly undercontrained datasets in standard tomography. This filter allows to reach an accuracy which is halfway between what achievable with analytical and iterative reconstruction.

Studies conducted on SRXTM datasets of mouse lung tissue in full interior tomography at a resolution of 1.1μm show that the proposed iterative reconstruction scheme enables a dose reduction per scan by a factor between 2.5 and 3. For other SRXTM datasets of high structural complexity in standard tomography, dose can be reduced by a factor 5 or even more, depending on the specimen under study and the envisaged quantification goal. Considering a CPU cluster with 50 cores and an underconstrained SRXTM dataset of typical size (e.g. 500 views $\times$ 2048 detector pixels $\times$ 2048 slices), the run time expected for the proposed iterative reconstruction algorithm is around 25-35 minutes.

Other aspects were also addressed, including the design of a reliable simulation framework to test reconstruction algorithms and assess their performance.

Apart from the single results, the hope with this doctoral project is to convince the reader that a goal-oriented vision and approaching iterative algorithms as "puzzles", whose pieces deserve both an independent and inter-mixed inspection, are indispensable ingredients to further expand the frontiers of tomographic image reconstruction and to see complex methods becoming practical and effective tools in the hand of scientists from diverse fields.

Zusammenfassung

Synchrotron-basierte tomographische Röntgen-Mikroskopie (Synchrotron-based X-ray Tomographic Microscopy, SRXTM) liefert hochauflösende dreidimensionale Abbildungen von mannigfaltigsten Materialien auf Grundlage von Durchleuchtungsbildern. Sie ermöglicht so zerstörungsfreie Untersuchung eines Probestücks im Mikrometer- und Sub-Mikrometer-Bereich. Zahlreiche wegweisende Studien, etwa in der Biologie, den Werkstoffwissenschaften, der Chemie oder der Paläontologie, basieren auf derartigen Daten.

Wenn dabei eine zu untersuchende Probe strahlungsempfindlich ist, oder wenn sie sich während der Aufnahme verändert, dann führt das zu Einschränkungen im Bildgebungsverfahren. In ersterem Fall darf die Strahlendosis einen gewissen Höchstwert nicht überschreiten, in letzterem Fall ist die zur Messung verfügbare Zeit begrenzt. Dementsprechend werden von solchen Proben tendenziell weniger Bilder in kürzerer Zeit aufgenommen. Es stehen zur Rekonstruktion des räumlichen Modells folglich weniger und aufgrund Rauschens schlechtere Rohdaten zur Verfügung. Ein derartiger Datensatz wird als “unterbestimmt” bezeichnet, und seine Verarbeitung stellt eine grosse Herausforderung dar, insbesondere bei Nutzung analytischer Rekonstruktionsverfahren.

Diese Doktorarbeit befasst sich mit der Entwicklung von Methoden und Strategien für die tomographische Rekonstruktion aus derartig unterbestimmten und mit niedrigen Strahlungsdosen gemessen SRXTM-Rohdaten. Sie beschäftigt sich dabei mit Absorptions- und (ausbreitungsbasiertem wie differenziellem) Phasenkontrast, und sie behandelt sowohl klassische als auch innere Rekonstruktionsprobleme. Angesichts der enormen Datenmengen von mehreren Terabytes, die bei SRXTM in kurzer Zeit anfallen, und aufgrund der Heterogenität der abzubildenden Proben orientieren sich die präsentierten algorithmischen Lösungen an folgenden Anforderungen: (i) gute Balance zwischen Genauigkeit und Berechnungseffizienz; (ii) Anpassungsfähigkeit an verschiedene zu untersuchende Materialien und Messprozeduren; (iii) minimale Nutzung von a-priori-Wissen oder “magischen” Parametern, insbesondere nicht um überhaupt Konvergenz zu erzielen. Mehrere Standardverfahren und spezialisierte hochmoderne iterative Methoden mussten angesichts dieser Vorgaben mangels Anpassungsfähigkeit, Berechnungseffizienz oder Nutzerfreundlichkeit verworfen werden, auch wenn deren grundsätzlicher Nutzen für medizinische und nicht-medizinische Anwendungen bereits in der Literatur bewiesen wurde.

Hinsichtlich Berechnungseffizienz und Flexibilität lässt sich die Alternate Direction Method of Multipliers in der “Plug and Play”-Variante (ADMP) als geeignetes iteratives Rekonstruktionsverfahren identifizieren. ADMP liefert nämlich nach bereits sehr wenigen Durchläufen Bilder mit ausreichendem Kontrast und guter Auflösung. Allgemein gesprochen berechnet ADMP die tomographische Rekonstruktion als regularisierte Näherungslösung des inversen Problems, das sich aus dem Bildgebungsverfahren ergibt: Gesucht ist nämlich das räumliche Bild, dessen künstliche Projektionen möglichst mit den gemessenen Bildern übereinstimmen. ADMP gestattet dabei freie Wahl sowohl hinsichtlich de verwendeten Projektionsoperators als auch in Bezug auf das zur Regularisierung genutzte Entrauschungsverfahren.

Die Berechnungseffizienz von ADMP ist in dieser Arbeit durch Verwendung von neuartigen Grid-

dingprojektoren mit minimaler Überabtastung massiv erhöht worden. Die niedrige numerische Komplexität dieser Operatoren ermöglicht sehr schnelle iterative Rekonstruktion auf einzelnen Cores, wobei die Geschwindigkeit vergleichbar ist mit komplexen GPU-basierten Ray-Casting-Projektoren. ADMP mit Griddingprojektoren ist insofern optimal geeignet für die Ausführung auf einem CPU-Cluster, wie er an Großforschungsanlagen vorhanden ist.

Weiterhin zeigt diese Arbeit experimentell, dass die Wahl übereinstimmender Vorwärts- und Rückwärtsprojektoren wesentlich für die Leistung der iterativen Methode ist: Nicht zueinander passende tomographische Operatoren verursachen im Bestfall suboptimale Genauigkeit und langsame Konvergenz, im schlechtesten Fall konvergiert das Verfahren entweder gar nicht, oder das Rekonstruktionsergebnis weist starke Bildfehler auf.

Hinsichtlich innerer Rekonstruktionsprobleme schlägt diese Arbeit eine "virtuelle Strategie" vor, um effizient iterative Rekonstruktionen zu berechnen. Problem in diesem Fall ist ja, dass die gemessenen Projektionen durch Bildanteile "verfälscht" sind, die nicht rekonstruierbar sind, da die verursachenden Strukturen außerhalb des Rekonstruktionsbereichs liegen. Ziel der Strategie ist folglich die virtuelle Umwandlung solcher "innerer" Rohdaten in äquivalente Standarddaten ohne Verwendung bekannter rechen- und speicherintensiver Verfahren zur Auffüllung fehlender Information. Ein solches Vorgehen macht durch Einsparung überflüssiger Rechenoperationen und Speicherzugriffe iterative Methoden letztlich überhaupt erst praktisch anwendbar.

In einem letzten Teil widmet sich die Arbeit schließlich dem Helgason-Ludwig-Sinogramm-Filter. Es handelt sich dabei um eine schnelle, parameterfreie Methode zur Verbesserung der Genauigkeit der analytischen Rekonstruktion aus stark unterbestimmten Daten. Dieser Filter ermöglicht die analytische Berechnung einer Rekonstruktion, deren Qualität etwa halbwegs die eines iterativ berechneten Ergebnisses erreicht, obwohl der Aufwand für die Filterung nur unwesentlich zu Buche schlägt.

Zusätzlich zu diesen Aspekten behandelt die Arbeit auch die Gestaltung eines zuverlässigen Simulationsrahmens zum Testen von Rekonstruktionsalgorithmen. Diese Regeln wurden bei Experimenten mit synthetischen Daten angewendet, um reproduzierbar zu verlässlichen Qualitätsaussagen zu kommen.

Zur praktischen Überprüfung der in dieser Arbeit vorgeschlagenen Verbesserungen wurden Studien mit SRXTM-Rohdaten durchgeführt. Für lokale Röntgenbilder von Mäuselungen mit einer Auflösung von 1.1 µm lässt sich zeigen, dass bei Verwendung der vorgeschlagenen iterativen Rekonstruktionsmethode eine 2,5- bis 3-fache Dosisreduktion pro Scan möglich ist. Für andere SRXTM-Datensätze mit höherer struktureller Komplexität kann die Dosis um Faktor 5 oder höher reduziert werden, abhängig von der untersuchten Probe und dem Quantifizierungsziel. Auf einem CPU-Cluster mit 50 Cores lässt sich ein typischer SRXTM-Datensatz (etwa 500 Projektionen × 2048 Detektorpixel × 2048 Schichten) mit den vorgeschlagenen iterativen Verfahren in ungefähr 25-35 Minuten verarbeiten.

Abgesehen von den konkreten Ergebnissen dieser Doktorarbeit besteht die Hoffnung, den Leser davon zu überzeugen, dass eine zielorientierte Herangehensweise an iterative Verfahren notwendig ist. Insbesondere sollten derartige Methoden als Puzzle verstanden werden, und die Teile sollten individuell und im Zusammenspiel analysiert werden. So können die bestehenden Grenzen tomographischer Bildrekonstruktion erweitert werden, und es wird möglich, komplexe numerische Methoden als praktische und effektive Werkzeuge in die Hände von Wissenschaftlern aus den verschiedensten Bereichen zu geben.

Acknowledgements

I would like to express here my gratitude to the many people that have characterized this long and beautiful journey.

I thank Prof. Marco Stampanoni for granting me the opportunity to conduct this doctoral project at TOMCAT, for his trust starting from the beginning and for having always been supportive to this work even in its darkest moments.

I thank my supervisor Dr. Federica Marone, from whom I learnt the importance of a practical spirit in steering the research and being constructively critic. Thanks for the countless hours that you spent behind my manuscripts, reports, presentations or posters and for even working at night, in the weekends or on holidays just to give me feedbacks and crucial remarks as soon as possible. I owe you for sure my scientific writing.

I thank Alain Studer for the incredible help when I started to approach scientific computing in Python. I will be always grateful for the many long and insightful discussions on inverse problems and signal processing, that we had at Time-out. I was lucky to find someone so passionate for his job and so capable of thinking outside the box.

I thank Dr. Goran Lovric and Dr. Silvia Peter for having been such good friends and having shared with me the happy and less happy moments since the start. Thanks to Goran for two unforgettable trips to Graz and Vodice and for all the advices, when I had barely a clue of how many things work. Thanks to Silvia for sharing with me tons of experiences like the wonderful concerts of Einsturzende and Massive Attack, our trip to Milan and Berlin. We ended this journey almost together and I wish you luck with all the will come.

I thank my dear friend Fabio D'Isidoro, who is probably one of the most sincere and unique person I have ever met so far.

I thank my dear friend Dr. Alberto Astolfo, that always reminded to take life from the bright side and not to take myself too seriously. Man, you don't know it yet, but you have been my personal Socratis, somehow.

I thank Dr. Rajmund Mokso for all the good time we had at conferences and workshops and to enjoy the convivial aspect of these scientific meetings.

I thank Dr. Pablo Villanueva for the insightful discussions on physics and image reconstruction, it was a pleasure to work with you and I hope will have other occasions to do.

I thank Dr. Jakob Vogel, that I "annoyed" with questions and discussion right after he sat at his new desk in our group. Many thanks for having revised this thesis and to be so meticolous in your work, I really enjoyed collaborating with you.

I thank Ioannis Vogiatzis for being always so prone to help in any moment. Good luck with your PhD journey!

I thank Matteo Abis, Dr. Peter Modregger, Dr. Kevin Mader, Carolina Arboleda, Maria Büchner, Mathias Kagias, Dr. Zhentian Wang, Dr. Spyridon Gkoumas, Dr. Alekhin Mikhail, Dr. Alessandra Patera and Dr. Anne Bonin for having been such good and funny colleagues. Doing almost three hours per day of commuting to come to PSI was not so bad after all, thinking that I would have shared the day with you. I hope to find in the future a team that can resemble you guys, good luck to you as well!

My deepest gratitude, however, goes to you, Olivera, for having been at my side during this journey and for being such a good "scientific collaborator" in this project called life.

Contents

List of Figures

List of Tables

Notation

$\mathbb{N}$	Set of natural numbers
$\mathbb{Z}$	Set of integer numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathcal{F}$	Fourier transform
$\mathcal{R}$	Radon transform
$\mathcal{A}$	Generic linear operator
$\mathcal{A}^*$	Adjoint of the linear operator $\mathcal{A}$
$\lfloor \ldots \rceil$	Rounding to the next integer
$\lfloor \ldots \rfloor$	Flooring operator
$*$	Convolution operator
$*^{-1}$	Deconvolution operator
rect	Rectangular function
$\mathbb{Z}_\mathrm{o}$	Set of odd integer numbers
$\mathbb{Z}_\mathrm{e}$	Set of even integer numbers
$\mathcal{B}$	Backprojection operator
$\mathcal{N}$	Normal distribution
Ш	Shah or comb function
Θ	Heaviside or box function
$\mathcal{L}$	Lagrangian
$\mathbf{I}$	Identity matrix

Abbreviations

CT	Computed Tomography
SRXTM	Synchrotron Radiation based X-ray Tomography Microscopy
CNR	Contrast-to-Noise-Ratio
TOMCAT	TOmographic Microscopy and Coherent rAdiology experiments
SLS	Swiss Light Source
PSI	Paul Scherrer Institut
DCMM	Double Crystal Multilayer Monochromator
CMOS	Complementary Metal-Oxide Semiconductor
FOV	Field-Of-View
PBPC	Propagation-Based Phase Contrast
TIE	Transport of Intensity Equation
GI	Grating Interferometer
DPC	Differential Phase Contrast
GPU	Graphics Processing Units
PBG	Parallel Beam Geometry
FST	Fast Slice Theorem
FFT	Fast Fourier Transform
FBP	Filtered BackProjection
DFT	Discrete Fourier Transform
std	standard deviation
SL	Shepp-Logan phantom
SLO	Shepp-Logan Original phantom
SLM	Shepp-Logan Modified phantom
SI	Spatial information
SF	Sampling factor
SNR	Signal-to-Noise Ratio
CNR	Contrast-to-Noise Ratio
ROI	Region Of Interest
PSF	Point-Spread-Function
erf	ERror Function
FRC	Fourier Ring Correlation

FSC --- Fourier Shell Correlation
MSE --- Mean Squared Error
RMSE --- Root Mean Squared Error
dB --- deciBels
PSNR --- Peak Signal-to-Noise Ratio
SSIM --- Structural SIMilarity index
MSSIM --- Mean Structural SIMilarity index
NMI --- Normalized Mutual Information
MNMI --- Mean Normalized Mutual Information
FGP --- Forward gridding projector
ADMM --- Alternate Direction Method of Multipliers
ADMP --- Alternate Direction Method of multipliers Plug-and-play
CS --- Compressed Sensing
TV --- Total variation
TGV --- Total Generalized Variation
OMP --- Orthogonal Matching Pursuit
DCT --- Discrete Cosine Transform
NLM --- Non-Local Means
NLTV --- Non-Local-Total-Variation
SIRT --- Simultaneous Iterative Reconstruction Technique
MLEM --- Maximum Likelihood Expectation Maximization
PWLS --- Penalized Weighted Least Squares
INT --- INterior Tomography
FINT --- Full INterior Tomography
DBP --- Differentiated BackProjection
HLSF --- Helgason-Ludwig Sinogram Filter
MLP --- Multi-Layer Perceptron

Chapter 1

Scientific Aim & Outline

Synchrotron radiation based X-ray tomographic microscopy (SRXTM) is an imaging technique that can provide non-destructive 3D information of a large variety of samples at micrometer and submicrometer scale, owing to the high brilliance and high coherence of the used radiation. Ultra-fast and in-vivo SRXTM studies are recent achievements enabled by the latest developments in detector technology and other instrumentation. Numerous cutting-edge investigations in the field of biology (e.g. [1–3]), material science (e.g. [4, 5]), geology (e.g. [6, 7]), energy research (e.g. [8, 9]) and paleontology (e.g. [10, 11]) have been made possible by this X-ray imaging technique.

This doctoral project complements the recent hardware advancements of SRXTM with software designed to address the reconstruction of tomographic datasets acquired in low-dose and fast scans. When the specimen investigated through SRXTM is radiation-sensitive, limiting the total dose of the scan is of primary importance. To fulfill the dose requirement, the exposure time per view and the number of projections need to be reduced, leading to the acquisition of so-called "underconstrained" datasets, whose accurate reconstruction poses a non-trivial problem.

On paper, the topic of this doctoral project is certainly not new. The first studies in tomographic reconstruction algorithms date back to the 70s. Since then, more than 60,000 papers have been published just under the keyword "iterative tomographic reconstruction", according to Google Scholar. Despite the amount of ink already poured in journals and scientific magazines, it was interesting to discover how much work is still needed to fill the gap between theory of tomographic reconstruction, algorithm implementation, experimental limitations and quantification goals.

Most of the studies on computed tomography (CT) reconstruction have been performed in the context of medical imaging. In the 70s, the filtered backprojection (FBP) algorithm [12] for analytical tomographic reconstruction and the family of algebraic algorithms [13–15] based on the Kaczmarz method [16] for iterative tomographic reconstruction were introduced. FBP has the advantage of being fast and easy to implement, while providing high-quality reconstructions when the dataset is well-sampled and features high signal-to-noise ratio. Algebraic methods are slower than FBP and yield reconstructions that generally look less sharp, but they may outperform FBP in terms of contrast-to-noise ratio when dealing with undersampled or noisy datasets. In the last 20 years, statistical iterative algorithms (e.g. [17, 18]), that incorporate the model ruling the signal formation at the detector, have probably represented the leading line of research in the field. This is exemplified by the fact that, since few years, some of the biggest manufacturers of CT scans like GE Healthcare, Philips Healthcare and Toshiba Medical Systems provide their own specific statistical iterative reconstruction methods [19–21]. When the physics of the experimental setup is well understood, statistical methods outperform algebraic methods in terms of spatial resolution and provide better reconstructions than both algebraic methods and FBP, when dealing with highly noisy datasets. Starting from the pioneering works of [22] and [23], research in the field has also witnessed an incredible explosion in the use of compressed sensing techniques combined to iterative methods (both algebraic and statistical) to better address the reconstruction of underconstrained tomographic datasets. Compressed sensing is used as regularization for iterative reconstruction to promote sparsity of the final result in a certain space. Sparsity is a property that commonly characterize natural and medical images and can be exploited to steer an iterative method towards better results, especially when the input dataset is undersampled and affected by noise. Several published works [24–26] have consistently demonstrated the advantage of using compressed sensing combined to an iterative reconstruction method to achieve higher accuracy. Altogether, the mainstream research line in CT reconstruction at the beginning of this PhD thesis (2012-2013) was focusing on statistical iterative algorithms incorporating compressed sensing based regularization schemes.

Our challenge consisted in designing general algorithms for the reconstruction of underconstrained datasets with high structural complexity and able to operate with minimal changes on projections acquired with different modalities available in modern synchrotron imaging beamlines: absorption-contrast, in-line phase contrast and differential phase contrast.

High-resolution SRXTM scans often require the acquisition of datasets in full interior tomography (FINT), where the field-of-view lies entirely inside the object support. The reconstruction of underconstrained FINT datasets represented another challenging task addressed by this doctoral project.

Furthermore, since SRXTM scans imply the acquisition of several terabytes of raw data, the computational feasibility and the trade-off between accuracy of the results and complexity of the investigated algorithms was a constant focus throughout this study. The target platform for the computations was a CPU cluster consisting of, at least, 50 cores.

Another point of interest was the selection of “user-friendly” algorithms, i.e. requiring the lowest amount of supervised a-priori knowledge and input parameters, in the spirit of Occam’s razor, to pursue the easiest and most practical among different viable solutions. “User-friendly” methods are more likely to become tools regularly exploited by scientists performing SRXTM experiments at imaging beamlines and not necessarily possessing an expertise in tomographic reconstruction.

This work was built, on one hand, upon recent advancements of compressed-sensing for image processing and, on the other hand, upon a rather unexplored algorithm for CT reconstruction: the alternate direction method of multipliers (ADMM) [27]. The target was, indeed, to discover whether the ADMM had some potential for CT iterative reconstruction (even just in few niches of application) compared to established algebraic and statistical methods.

The uniqueness of this work lies in the innovative way to approach tomographic reconstruction of underconstrained datasets.
First, algorithm design and characterization are driven by the quantification goals envisaged for specific datasets and summarizable in questions like: what trade-off between contrast and spatial resolution is provided by the proposed reconstruction method? Does it significantly facilitate segmentation? What is the achievable dose reduction factor? In this respect, the SRXTM datasets of mouse lung tissue in full interior tomography, acquired in the framework of the project "In-vivo lung imaging at micrometer scale" [1], has offered the perfect platform to pursue this work. Dose reduction represents, indeed, a crucial aspect to enable longitudinal studies, i.e. following the lung development at different stages of the animal growth. The well-defined quantification goals of the lung imaging project were fundamental to direct some of the studies conducted in this thesis.
Second, iterative reconstruction is analyzed for the first time as a "puzzle", whose components require to be inspected both separately and in relation to the other parts. This approach has lead to the following results: the design of the gridding projectors with minimal oversampling; the discovery that the accuracy of the tomographic projectors has limited influence on the accuracy of the iterative reconstruction algorithm; revealing the crucial importance of working with highly-coupled tomographic projectors to guarantee the optimal performance of an iterative algorithm; the characterization of the alternate direction method of multipliers, where the various components were separately analyzed; highlighting the paramount relevance of the preprocessing strategies to tackle the reconstruction of FINT datasets.

This thesis is organized as follows.

Chapter 2 is a short introduction to the concepts needed to understand the research conducted in this doctoral project. It starts from the Beer-Lambert law and briefly describes the physics behind absorption-contrast, inline phase-contrast and differential phase-contrast. The Radon transform and its most known implementations in real space are then introduced, followed by the Fourier slice theorem and a discussion about filtered backprojection. It concludes with the definition of "underconstrained" dataset and an overview of the different "families" of iterative reconstruction algorithms.

Chapter 3 offers a description of the simulation framework used to test tomographic reconstruction algorithms. The chapter opens by discussing how to choose simulated objects and what aspects should be considered when using such objects to compute forward projections. Several metrics for image quality are, then, introduced and their characteristics and limitations highlighted. In particular, signal-to-noise ratio, contrast-to-noise ratio, methods to estimate the spatial resolution and full reference metrics are here discussed and compared from a user perspective.

Chapter 4 presents the detailed study conducted on the forward gridding projector. This operator represents the key for fast iterative tomographic reconstruction without need for graphics processing units. The chapter discusses the gridding method, its mathematical formulation and optimal gridding kernels to guarantee a good trade-off between accuracy and efficiency. The gridding projectors with minimal oversampling and their utilization for iterative tomographic reconstruction are, then, introduced and analyzed in detail. The chapter presents also a contextualization of these operators in the current research of tomographic reconstruction and with future perspectives and closes with an experimental work regarding the ocupling effect between forward and backprojector. Three manuscripts produced in this doctoral project [28–30] are here presented.

Chapter 5 describes the usage of the alternate direction method of multipliers (ADMM) for iterative regularized reconstruction and introduces the "plug-and-play" algorithm (ADMP), as a versatile framework allowing the integration of any denoising method as regularization scheme. This chapter offers, first, an overview of regularization schemes based on compressed sensing and, then, presents a full experimental characterization of the ADMM, aimed at studying separate aspects like speed of convergence, influence of the number of conjugate-gradient sub-iterations on the overall performance of the algorithm, usage of physical constraints and warm initialization. A comparison on simulated data between the ADMM and the ADMP closes this chapter.

Chapter 6 discusses the interior tomography problem and the methods, namely differentiated backprojection and edge-padding, to perform analytical reconstruction of this kind of datasets. The chapter, then, focuses on the specific case of full interior tomography (FINT) without knowledge of the object support, characteristic of many SRXTM acquisitions. This part shows how the gridding projectors with minimal oversampling can be combined to differentiated backprojection and edge-padding to perform fast iterative reconstruction of underconstrained FINT datasets. The virtual strategy is also proposed as an equally effective, but much more efficient alternative to perform the same task. One manuscript produced in this doctoral project [31] is here presented.

Chapter 7 describes a non-iterative unsupervised algorithm, designed to double the number of views of a tomographic dataset by enforcing the Helgason-Ludwig consistency conditions. The method, abbreviated as HLSF (Helgason-Ludwig sinogram filter), is a fast filter, that works in the Fourier-Chebyshev domain and is applied prior to filtered backprojection. The HLSF improves the accuracy of analytical reconstructions, when dealing with heavily undersampled datasets in standard tomography. However, the quality of the reconstructions yielded by the HLSF followed by an analytical method is not comparable to what achievable with a finely tuned iterative reconstruction algorithm. One manuscript produced in this doctoral project [32] is here presented.

Chapter 8 summarizes the various contributions of this doctoral project and discusses future perspectives for the reconstruction of underconstrained SRXTM datasets.

Chapter 9 is entitled "Additional Contributions" and presents an application of the ADMM framework to a specific phase retrieval problem. One manuscript produced in this doctoral project [33] is here presented.

Chapter 2

Background

2.1 Synchrotron-based X-ray tomographic microscopy

X-ray radiography, discovered by Röntgen in 1895 [34], is a 2D non-destructive imaging technique, that has played a crucial role in various fields like medicine, material science, archeology, quality control and homeland security until these days. In the 1960s, Cormack had the intuition that the density distribution of an object could be retrieved from radiographic projections acquired around a single rotation axis [35] and in the 1970s Hounsfield devised the first X-ray tomographic scanner [36], upgrading radiography to 3D and giving birth to *computed tomography (CT)*. CT scans have since been extensively utilized for preventive medicine and disease screening [37]. This method was later adopted for non-destructive testing in material research starting from the 1980s [38, 39]. In the same years, X-ray transmission tomography at micrometer scale was studied with synchrotron radiation [40] and microfocus X-ray tubes [41].

Compared to laboratory sources, the major assets of X-rays produced at third-generation synchrotron facilities are: the possibility of working with monochromatic light, therefore miminizing beam hardening effects; a few orders of magnitude higher flux, allowing much faster scans even with monochromatic light; an almost exact parallel beam geometry, which greatly simplifies tomographic reconstruction; partial transverse coherence of the X-ray beam enabling the usage of phase-contrast techniques. For all these reasons, reconstructions of synchrotron radiation based X-ray tomographic microscopy (SRXTM) data are affected by fewer artifacts and feature a substantially improved contrast-to-noise-ratio (CNR) [42].

In the last 20 years, SRXTM has found broad application in biology [1, 43–47], material science [48–54] and paleontology [55, 56].

The experimental data used for the studies conducted in this thesis were acquired at the TOMCAT beamline of the Swiss Light Source (SLS) at the Paul Scherrer Institute (PSI), in operation since June 2006 [57]. X-rays are produced by a 2.9 T superbending magnet with a critical energy of 11.1 keV on a 2.4 GeV storage ring [58]. A fixed-exit double crystal multilayer monochromator (DCMM) located in the front-end allows to work with monochromatic light in the 6 to 45 keV energy range [59]. X-rays exiting the sample are, first, converted into visible light by a scintillator screen and, then, projected onto a digital camera (CCD, CMOS or sCMOS) detector with suitable optics [1]. Thanks to interchangeable cameras, the field-of-view (FOV) can range from 28.2× 23.8 mm^2 with 11 μm pixel size to 13.1× 13.1 mm^2 with 6.5 μm pixel size. Objectives placed between the scintillator screen and the camera can yield an up to 40-fold magnification of the FOV, allowing to reach an effective pixel size down to 0.16 μm.

2.2 X-ray projections

2.2.1 Beer-Lambert law

X-ray photons traveling through a material either interact with particles of mass or pass unaffected. Interactions occur every time a photon is absorbed, due to the photoelectric effect, or is scattered, due to the Compton (incoherent scattering) or Rayleigh (coherent scattering) effect. In all these scenarios, photons are considered removed from the original beam. The loss of photons caused by absorption and scattering is called *attenuation*.
Assuming that N_{in} X-ray photons move in parallel geometry along the z-axis and hit perpendicularly an infinitesimally thin (thickness $\Delta z \approx 0$) and homogeneous slab of material, the Beer-Lambert law states that the expected number of photons lost inside the body is directly proportional to N_{in} and Δz [60–62]:

$$\Delta N := N_{\text{out}} - N_{\text{in}} = -\mu N_{\text{in}} \Delta z \quad , \tag{2.1}$$

where μ is called *linear attenuation coefficient*, has units of an inverse length and represents a characteristic quantity of a given material. It holds that:

$$\frac{\mu}{\rho} = \frac{\sigma_{\text{tot}}}{uA} \quad , \tag{2.2}$$

where ρ is the density, u is the atomic mass unit, A is the relative atomic mass and σ_{tot} is the total cross section, accounting for all types of interactions between X-rays and the selected material. Considering the differential form of (2.1) and integrating over the thickness of the slab l, the Beer-Lambert law becomes:

$$N_{\text{out}} = N_{\text{in}} \exp\left(-\mu l\right) \quad . \tag{2.3}$$

If the body consists of n_s thin slabs with different materials whose thicknesses z_i sum up to l, (2.3) assumes a more general form:

$$N_{\text{out}} = N_{\text{in}} \exp\left(-\sum_{i=0}^{n_s} \mu_i z_i\right)\Bigg|_{\sum_{i=0}^{n_s} z_i = l} \overset{z_i \longrightarrow 0\,,\forall\, i}{\overset{n_s \longrightarrow \infty}{=}} N_{\text{in}} \exp\left(-\int_0^l \mu(z)\,dz\right) \quad . \tag{2.4}$$

Since, in practice, the X-ray source provides a flux of photons, it is convenient to rewrite (2.4) in terms of *intensities*, defined in general as the number of photons impinging on a selected region per unit of time:

$$\mathrm{I}_{\text{out}} = \mathrm{I}_{\text{in}} \exp\left(-\sum_{i=0}^{n_s} \mu_i z_i\right)\Bigg|_{\sum\limits_{i=0}^{n_s} z_i = l} \overset{z_i \longrightarrow 0\,,\forall\, i}{\overset{n_s \longrightarrow \infty}{=}} \mathrm{I}_{\text{in}} \exp\left(-\int_0^l \mu(z)\mathrm{d}z\right) \quad . \tag{2.5}$$

I_{out} is called *projection image* of μ. Relation (2.5) is valid only under the following conditions: (i) interactions act independently from each other; (ii) the incident radiation consists of parallel rays; (iii) the X-ray beam is monochromatic, otherwise (2.5) has to be rewritten for each single energy component $\mathcal{E}$ by replacing I_{out}, I_{in} and $\mu(z)$ with $\mathrm{I}_{\text{out}}(\mathcal{E})$, $\mathrm{I}_{\text{in}}(\mathcal{E})$ and $\mu(z, \mathcal{E})$.

At SLS, X-rays are shaped to have a divergence angle between 0.6 mrad and 2 mrad [58], allowing the parallel beam geometry to be a good approximation. The DCMM provides monochromatic radiation with $\Delta\lambda/\lambda$ (λ is the wavelength) of a few percent when multilayer crystals are used or 10^{-4} when the Si-111 crystals are used [63].

Despite its simplicity, the Beer-Lambert law is the physical principle providing the basis for the mathematical model of tomographic reconstruction.

2.2.2 Absorption-contrast imaging

Standard *X-ray absorption imaging* can be performed on materials and at energies where $\mu = \mu_a + \mu_s \approx \mu_a$, μ_a being the absorption coefficient and μ_s the scattering coefficient. The cross-section of the photoelectric effect, ruling the absorption of X-ray photons inside the sample, is $\propto Z^3/E^3$ [64], where Z is the atomic number and E the beam energy. X-ray absorption imaging can well discriminate between hard and soft tissues, as μ_a varies considerably for materials with so disparate densities. On the other hand, it fails to provide enough contrast between different types of soft tissues due to the little variation of μ_a.

2.2.3 Phase-contrast imaging

At X-ray energies, the refractive index n of a material is expressed as follows [65]:

$$n = 1 - \delta + i\beta \tag{2.6}$$

$$\text{where} \qquad \beta = \frac{\mu_a \lambda}{4\pi} \tag{2.7}$$

$$\text{and} \qquad \delta = \frac{\lambda^2 r_e n_e}{2\pi} \quad , \tag{2.8}$$

where r_e is the classical radius of the electron and n_e the electron density of the material. The terms β and δ are related to different interactions between X-rays and the sample: β accounts for the absorption, which affects the amplitude of the X-ray wave front [66]; δ accounts for the coherent Rayleigh scattering [65], which affects the phase of the X-ray front wave [66].

Phase-contrast is an imaging modality sensitive to variations of δ, thus, of electron density according to (2.8). Soft tissues can be effectively discriminated when $\lambda < 0.1$ nm and phase-contrast dominates absorption contrast [65, 66]. Phase-contrast techniques measure phase variations of a spatially coeherent X-ray beam. Since it is not possible to fabricate detectors with response time comparable to the oscillation of optical or quantum fields, the phase of the wave front cannot be measured directly [67]. There are various systems to convert X-ray phase variations into intensity variations measurable at the detector. Examples are propagation-based phase contrast [53, 68], differential phase contrast based on a Talbot grating interferometer [69], analyzer-based phase contrast [70] and crystal-based interferometry [71].

In this work, we addressed the reconstruction of projections created by propagation-based and differential phase contrast.

Propagation-based phase-contrast projections

Propagation-based phase-contrast (PBPC) (also called "in-line phase-contrast") is yielded by the free-space propagation of the X-ray wave front coherently scattered by the sample [53, 68] and is based on the Fresnel approximation of the Kirchoff diffraction formula [72]. For absorption contrast imaging, the detector is located directly behind the object. The opposite is true for PBPC, where a certain distance is necessary to obtain edge enhancement.

Methods based on the *contrast-transfer function (CTF)* [73] and the *transport of intensity equation (TIE)* [74, 75] offer a way to retrieve the phase from the intensity data. CTF works under the condition that the sample is weakly absorbing [76]; it also requires high coherence and the acquisition of projections at multiple distances [54]. TIE is a preferable solution when the beam coeherence is limited [77].

The phase of the PBPC projections used in this work are retrieved by the Paganin algorithm [77], that inverts the TIE equation and is applicable under the following assumptions: (i) "near-field" condition, i.e. short detector-sample distance; (ii) the *projection* and *paraxial* approximations are valid [78]; (iii) the sample consists of a single material. Calling k_x and k_y the dual Fourier variables of x and y, $I(x,y,z=\delta z)$ the intensity measured at distance δz from the sample, the projected phase ϕ is retrieved by the Paganin algorithm as [77]:

$$\phi(x,y,z=0) = \frac{2\pi}{\lambda}\frac{\delta}{\mu}\ln\left(\mathcal{F}^{-1}_{x,y}\left\{\mu\frac{\mathcal{F}_{x,y}\left\{\frac{I(x,y,z=dz)}{I_{\text{in}}}\right\}}{dz\,\delta\sqrt{k_x^2+k_y^2}+\mu}\right\}\right) \quad , \tag{2.9}$$

with

$$\mathcal{F}_{x,y}\left\{f(x,y)\right\} = \int_0^\infty dx \int_0^\infty dy\, f(x,y)\exp\left(i\left(k_x x + k_y y\right)\right) \quad . \tag{2.10}$$

being the Fourier transform. Formula (2.9) clearly shows that Paganin phase retrieval works as a low-pass filter on the input intensity data $I(x,y,z=dz)$, since the kernel $1/\left(a_1\sqrt{k_x^2+k_y^2}+a_2\right)$ attenuates more frequency components corresponding to high values of k_x and k_y.

$\phi(x, y, z = 0)$ represents the projected phase of the object. Once phase projections are acquired at different angular positions, tomographic reconstruction techniques can provide the entire 3D phase map of the object.

Compared to other methods, Paganin phase retrieval has the advantage of being fast and requiring single projections. Although quantitative results are not achievable when the sample consists of two or more materials, Paganin can still be applied to enhance the contrast-to-noise ratio of the projections.

Differential phase-contrast projections

A *grating interferometer (GI)* [69, 79] is a setup consisting of two gratings located after the source and the sample. The first grating, G1 with period p_1, splits the incident beam into several diffraction orders, which interfere constructively at specific distances if the beam features sufficient spatial coherence. The second grating, G2 with period p_2, is made of highly absorbing materials and is positioned in front of the detector to analyze the interference pattern, whose period would be too small to be resolved otherwise.

GI is based on the occurrence of a rectangular-shaped interference pattern due to the *Talbot effect* [80]. Once the sample is placed inside the incident wave, the phase sensitivity of GI is given by the lateral shift of the interference pattern. Assuming the gratings aligned along the x-axis (i.e. the line of the gratings are vertical), the lateral shift Δx is proportional to the refraction angle α, which, in turn, is proportional to the first derivative of the phase of the wave of front after having traversed the sample [67, 81]:

$$\alpha = -\frac{1}{k}\frac{\partial \phi}{\partial x} \quad , \tag{2.11}$$

where $k = 2\pi/\lambda$.

Phase-stepping is a widely adopted procedure to analyze the signal at the detector: one grating is shifted by fractions of its pitch with respect to the other grating and at each step a projection is acquired [67]. The result is an oscillatory triangular function for each pixel, that looks closer to a sinusoid due to the finite spatial coherence of the system. The phase stepping curve, $\varphi(\phi)$, is of the form [82]:

$$\varphi(\phi) = a_1 + a_2 \sin(\phi - p) \quad , \tag{2.12}$$

a_1, being the mean of φ, is connected to the absorption contrast; a_2, being the amplitude of φ, is connected to the scatter contrast; s is the lateral offset of the phase steps; p is the phase of φ and relates to the differential phase contrast (DPC). In particular, if $\varphi(\phi)_1$ and $\varphi(\phi)_0$ are the phase stepping curves acquired with and without sample, DPC is given by [67]:

$$\text{DPC} = \arg\left(\mathcal{F}\{\varphi(\phi)_1\}(q_n)\right) - \arg\left(\mathcal{F}\{\varphi(\phi)_0\}(q_n)\right) \quad , \tag{2.13}$$

where q_n is the n-th Fourier harmonic and n is the number of phase steps. DPC is proportional to α according to [67]:

$$\text{DPC} = 2\pi\frac{d_m \alpha}{p_2} \quad , \tag{2.14}$$

where d_m is the selected integer or fractional Talbot distance. Considering (2.11), (2.14) and that:

$$\phi(x,z) = \int dz\, \delta(x,z) \tag{2.15}$$

where the z-axis is the direction of propagation of the X-rays, DPC is proportional to the derivative of the line integral of δ along the grating direction (x-axis, in this case).

2.2.4 Corrected projections

In the experimental practice, two are the main sources of systematic noise imparing the quality of X-ray projections: fix-pattern and camera noise. Fix-pattern noise stems from beam inhomogeneities, non-uniform detector response due to variations in the photon conversion yield, losses in charge transport, charge trapping, non-constant performance of the readout or simply dust/scratches accumulated on the surface of the scintillator or other optical elements [83]. The camera noise is mainly due to the dark current, which is proportional to the exposure time, and to the ADC pedestal or digitization offset, which is, instead, independent from the exposure time.

Flat-field and *dark-field corrections* are used to greatly reduce the impact of these sources of noise. The flat field or white field is the intensity map of the X-ray beam without sample and serves to capture the fix-pattern noise. The dark field corresponds to an image acquired by the detector without X-ray illumination, which accounts for the camera noise. Calling the raw projection P, the flat-field F and the dark-field D, the corrected projection, P_c, is computed as follows:

$$P_c = \frac{P - D}{F - D} \quad . \tag{2.16}$$

The correction computed by (2.16) is effective when the X-ray beam, scintillator response and camera sensitivity can be considered stationary, an assumption that often is only approximately met [84]. When beam inhomogeneities or fix-pattern noise are absent, the flat-field correction is still required by the Beer-Lambert law to calculate the unknown attenuation coefficient, since $\mu l = \ln(N_{\text{in}}/N_{\text{out}}) = \ln(F/P)$.

2.3 Radon transform

2.3.1 Definition and properties

The Radon transform represents the mathematical forward model for tomographic reconstruction. Studied since the first decade of the 20th century by J. Radon [85], this transform has been applied also to electron microscopy [86], reflection seismology [87], hyperbolic partial differential equations [88, 89], barcode scanners [90] and line detection on natural images [91].

Considered a function $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$, the Radon transform $\mathcal{R}$ integrates f over an hyperplane

$HY(\boldsymbol{\theta}, t) = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x} \cdot \boldsymbol{\theta} = t\}$, where $\boldsymbol{\theta} \in S^{n-1}$ and $t \in \mathbb{R}$ is the signed distance from the origin [92]:

$$\mathcal{R}\{f\}(\boldsymbol{\theta}, t) := \int_{HY} d\mathbf{x}\, f(\mathbf{x}) = \int_{\mathbb{R}^n} d\mathbf{x}\, \delta(t - \mathbf{x} \cdot \boldsymbol{\theta}) f(\mathbf{x}) = \int_{\boldsymbol{\theta}^\perp} d\mathbf{x}\, f(t\boldsymbol{\theta} + \mathbf{x}) \quad , \tag{2.17}$$

where δ is the Dirac function and $\boldsymbol{\theta}^\perp = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x} \cdot \boldsymbol{\theta} = 0\}$ is the subspace orthogonal to $\boldsymbol{\theta}$. (2.17) encloses three different equivalent definitions of $\mathcal{R}$. $\mathcal{R}\{f\}$ is a function on the unit cylinder in $\mathbb{R}^n$:

$$CY^n = \left\{(\boldsymbol{\theta}, t) \,:\, \boldsymbol{\theta} \in S^{n-1} \,,\, t \in \mathbb{R}\right\} \tag{2.18}$$

The domain of the Radon transform is defined as follows:

$$\Omega\,(CY^n) = \left\{\forall\, g \in C^\infty(CY^n) \,:\, t^l \frac{\partial^k}{\partial t^k} g(\boldsymbol{\theta}, t) \;\text{ is bounded}\,,\;\; l, k = 0, 1, \ldots\right\} \quad . \tag{2.19}$$

In the two dimensional case, the Radon transform integrates the function along lines. For $n = 2$, $f(\mathbf{x}) = f(x_1, x_2)$, $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$, HY is a line of equation $\mathbf{x} \cdot \boldsymbol{\theta} = x_1 \cos\theta + x_2 \sin\theta = t$, CY^2 is a circle of unit radius and the second definition in (2.17) simplifies to:

$$\mathcal{R}\{f\}(\theta, t) := \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, f(x_1, x_2)\, \delta(x_1 \cos\theta + x_2 \sin\theta - t) \quad . \tag{2.20}$$

From this point onward, the two dimensional Radon transform (2.20) is considered.

Property (1): symmetry. Both parameter sets $\{t \in \{0, +\infty\} \,\cup\, \theta \in \{0, 2\pi\}\}$ and $\{t \in \{-\infty, +\infty\} \,\cup\, \theta \in \{0, \pi\}\}$ describe every element of the Radon transform, as it holds that:

$$\mathcal{R}\{f\}(\theta, t) = \mathcal{R}\{f\}(\theta + \pi, -t) \quad . \tag{2.21}$$

Property (2): linearity. Given $\alpha_1, \alpha_2 \in \mathbb{R}$ and two functions $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$, it holds that:

$$\mathcal{R}\{\alpha_1 f_1 + \alpha_2 f_2\} = \alpha_1\, \mathcal{R}\{f_1\} + \alpha_2\, \mathcal{R}\{f_2\} \quad . \tag{2.22}$$

Property (3): shifting. Given $g(x_1, x_2) = f(x_1 - k_1, x_2 - k_2)$, it results that:

$$\mathcal{R}\{g\}\,(\theta, t) = \mathcal{R}\{f\}\,(\theta, t - k_1 \cos\theta - k_2 \sin\theta) \quad . \tag{2.23}$$

Property (4): rotation. Working in polar coordinates (φ, r) and given $g(\varphi, r) = f(\varphi - \varphi_0, r)$, it results that:

$$\mathcal{R}\{g\}(\theta, t) = \mathcal{R}\{f\}(\varphi_0 - \theta, t) \quad . \tag{2.24}$$

Property (5): scaling. Given $a, b \in \mathbb{R}_0$ and $g(x_1, x_2) = f(x_1/a, x_2/b)$, it results that:

$$\begin{gathered}\mathcal{R}\{g\}(\theta, t) = \mathcal{R}\{f\}(\theta', t') \\ t' = \frac{t}{\sqrt{a^2 \cos^2\theta + b^2 \sin^2\theta}} \qquad \theta' = \tan^{-1}\left(\frac{a}{b}\tan\theta\right) \quad .\end{gathered} \tag{2.25}$$

Property (6): convolution. Given two functions $f_1(x_1, x_2)$, $f_2(x_1, x_2)$ and indicating with $\overset{x,y,z,\dots}{*}$ the convolution with respect to the variables $x, y, z, \dots$, it results that:

$$\mathcal{R}\{f_1 \overset{x_1,x_2}{*} f_2\} = \mathcal{R}\{f_1\} \overset{t}{*} \mathcal{R}\{f_2\} \quad . \tag{2.26}$$

The relation between the Beer-lambert law and the Radon transform is straightforward if (2.5) is rewritten in the form:

$$\ln\left(\frac{I_{in}}{I_{out}}\right) = \int\limits_{\text{ray path}} \mu(z) dz \quad , \tag{2.27}$$

which states that the logarithmic ratio between the exiting and incident intensity corresponds to the Radon transform of the map of attenuation coefficients along the ray direction. For absorption, PBPC and DPC projections, (2.27) represents the Radon transform of the map of absorption coefficients, the Radon transform of the map of δ and the derivative of the Radon transform of the map of δ, respectively.

2.3.2 Real space implementations

Slant stacking

Formula (2.20) can be explicitly rewritten as a line integral by introducing the variable s, which runs over the X-ray path. On the basis of Fig. 2.1, $x_1 = t\cos\theta - s\sin\theta$, $x_2 = t\sin\theta + s\cos\theta$, therefore, (2.20) becomes:

$$\begin{aligned}\mathcal{R}\{f\}(\theta, t) &= \int\limits_{-\infty}^{+\infty} ds\, f\left(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta\right) \\ &\approx \Delta s \sum_{l=0}^{L-1} f\left[t_j\cos\theta_k - s_l\sin\theta_k, t_j\sin\theta_k + s_l\cos\theta_k\right] \quad .\end{aligned} \tag{2.28}$$

The last term in (2.28) is the discretized version of $\mathcal{R}\{f\}(\theta, t)$; for this reason, f has, now, square brackets and all variables t, θ and s are indexed. The implementation of formula (2.28) necessitates the use of two-dimensional interpolation, as the required points $[t_j \cos\theta_k - s_l \sin\theta_k, t_j \sin\theta_k + s_l \cos\theta_k]$, in general, never coincide with the given image points $[x_{1,m}, x_{2,n}]$. Two-dimensional interpolation can be avoided by approximating the Radon transform as *slant stacking* (in seismics also known as τ-p transform) [93], which entails only 1D interpolations. The slant stacking approximation of the discrete Radon transform divides the interval $[0, \pi]$ in two regions: one for nearly-horizontal lines $\theta \leqslant \pi/4$ and $3\pi/4 \leqslant \theta \leqslant \pi$; one for nearly-vertical lines $\pi/4 \leqslant \theta \leqslant 3\pi/4$. The slant stacking approximation with nearest neighbor interpolation results [93]:

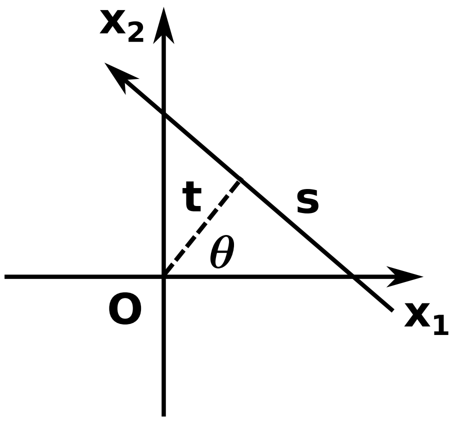


Figure 2.1: Reference system for formula (2.28).

$$\mathcal{R}\{f\}\left[\theta_k, t_j\right] = \begin{cases} \dfrac{\Delta x_1}{|\sin\theta_k|} \displaystyle\sum_m f\left[x_{1,m}, \lfloor a x_{1,m} + b \rceil\right] & \text{for } |\sin\theta| > \frac{\sqrt{2}}{2} \\ \dfrac{\Delta x_2}{|\cos\theta_k|} \displaystyle\sum_n f\left[\lfloor a' x_{2,n} + b' \rceil, x_{2,n}\right] & \text{for } |\sin\theta| \leqslant \frac{\sqrt{2}}{2} \end{cases}, \tag{2.29}$$

where $x_{1,m} = x_1^{\text{min}} + m\Delta x_1$ for $m = 0, 1, \dots, M-1 \in \mathbb{N}$, $x_{2,n} = x_2^{\text{min}} + n\Delta x_2$ for $n = 0, 1, \dots, N-1 \in \mathbb{N}$, $\lfloor \dots \rceil$ represents the operator rounding to the next integer and

$$\begin{aligned} a &= -\frac{\Delta x_1}{\Delta x_2} \frac{\cos\theta}{\sin\theta} & b &= \frac{t - x_1^{\text{min}} \cos\theta - x_2^{\text{min}} \sin\theta}{\Delta x_2 \sin\theta} \\ a' &= \frac{1}{a} & b' &= \frac{t - x_1^{\text{min}} \cos\theta - x_2^{\text{min}} \sin\theta}{\Delta x_1 \cos\theta} \quad . \end{aligned} \tag{2.30}$$

Formula (2.29) clearly shows that interpolation is needed only along the s-direction, being x_1 or x_2 depending on the selected θ. The slant stacking approximation with linear interpolation becomes [93]:

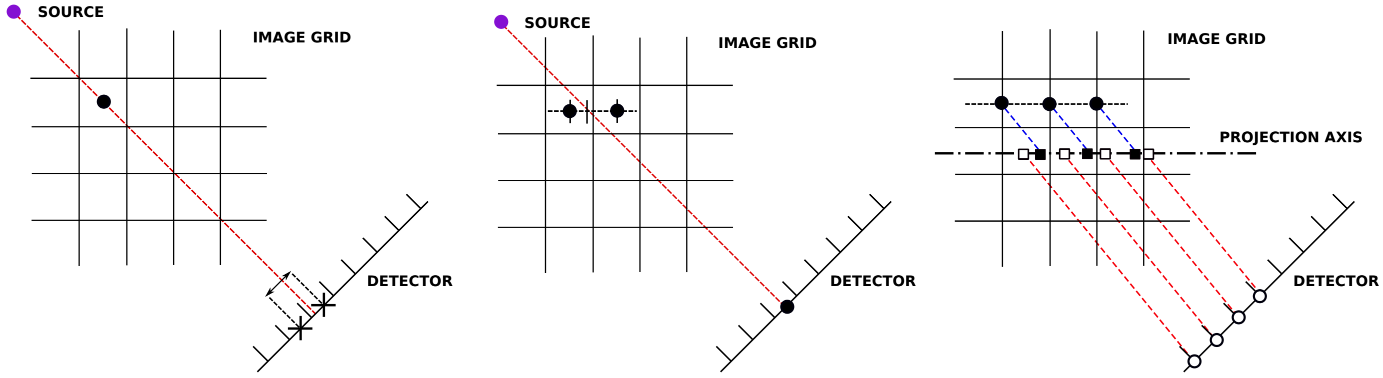
$$\mathcal{R}\{f\}\left[\theta_k, t_j\right] = \begin{cases} \dfrac{\Delta x_1}{|\sin\theta_k|} \displaystyle\sum_m (1-\omega) f\left[x_{1,m}, \lfloor a x_{1,m} + b \rfloor\right] + \omega f\left[x_{1,m}, \lfloor a x_{1,m} + b \rfloor + 1\right] \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } |\sin\theta| > \frac{\sqrt{2}}{2} \\ \dfrac{\Delta x_2}{|\cos\theta_k|} \displaystyle\sum_n (1-\omega') f\left[\lfloor a' x_{2,n} + b' \rfloor, x_{2,n}\right] + \omega' f\left[\lfloor a x_{2,m} + b \rfloor + 1, x_{2,m}\right] \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } |\sin\theta| \leqslant \frac{\sqrt{2}}{2} \end{cases} \tag{2.31}$$

where a, a', b, b' are defined in (2.30), $\lfloor ... \rfloor$ is the flooring operator, $\omega = a x_{1,m} + b - \lfloor a x_{1,m} + b \rfloor$ and $\omega' = a' x_{2,n} + b' - \lfloor a' x_{2,n} + b' \rfloor$. Given a square image $N \times N$ pixels and N views in $[0, \pi]$, the slant slacking implementation of the discrete radon transform with either nearest neighbor or linear interpolation is characterized by a complexity $\mathcal{O}(N^3)$.

Pixel-, ray- and distance-driven projectors

In the pixel-driven approach [95], the source point is connected to the selected pixel center until intersection with the detector line, as displayed in Fig. 2.2a. A linear interpolation scheme distributes the pixel value to the two detector cells that enclose the ray end point (they are indicated with a cross in Fig. 2.2a). Pixel-driven forward projection is rarely utilized due its characteristic high-frequency artifacts [96].

The ray-driven approach [97] connects the source to the center of a selected detector cell (Fig. 2.2b). The latter point is projected for each row (column) onto the horizontal (vertical) axis joining the centers of the two image pixels, that surround the ray (this axis is represented in Fig. 2.2b as a



(a) Pixel-driven approach (b) Ray-driven approach (c) Distance-driven approach

Figure 2.2: Schematic representation of the different mechanisms characterizing the pixel-, ray- and distance-driven approach for forward projection. This figure is readapted from [94].

dashed horizontal line within the image grid): a linear interpolation scheme weighs the contributions for each couple of image pixels. The ray-driven projection suffers from the same kind of high-frequency artifacts affecting the pixel-driven approach [98].

The distance-driven approach [94, 98] in Fig. 2.2c projects the pixel boundaries (black dots) of each image row/column and the detector cell boundaries (white dots) onto a common axis (this axis is represented in Fig. 2.2c as a dashed horizontal line within the image grid). The overlap between the interval defined by the projected boundaries of an image pixel (black squares) and the one defined by the projected boundaries of a detector cell (white squares) weighs the contribution of the selected image pixel to the selected detector cell (and vice versa). This method differs substantially from the pixel- and ray-driven approaches for two reasons: (i) the distance-driven strategy is faster, as the main loop runs over the projected boundaries on the projection axis, rather than running over the image pixels or detector cells; (ii) the linear interpolation scheme is replaced with the calculation of overlapping intervals. This different approach prevents projections from being affected by the aforementioned high-frequency artifacts [94, 98].

Alternative basis for image representation

The previous implementations of the Radon transform employ the most common basis for image representation: an array of abutting square pixels (cubic voxels, if the 3D case is considered). This basis function assumes that each pixel has unit value inside and is zero outside. The major drawback of a pixel basis is that combinations of such piecewise-constant elements may represent a rather poor approximation of smoothly varying functions.

Given an image $\mathbf{f} \in \mathbb{R}^{M,N}$ sampled at nodes of a square grid $G = \{\forall \mathbf{x} \in \mathbb{R}^2 \mid \mathbf{x} = \mathbf{x}_{m,n} = \mathbf{x}_0 + (m,n)^T\ ,\ m = 1,2,...,M\ ,\ n = 1,2,...,N\}$ and a generic set of functions $\Phi_q : \mathbb{R}^2 \longrightarrow \mathbb{R}$, the continuous representation of the image on the selected basis results:

$$\tilde{f}(\mathbf{x}) = \sum_{q=1}^{Q} c_q\ \Phi_q(\mathbf{x})\ \ ,\ \ \forall\, \mathbf{x} \in \mathbb{R}^2 \quad \text{such that} \quad \tilde{f}(\mathbf{x}_{m,n}) = f_{m,n}\ \ , \tag{2.32}$$

where c_q's are the basis coefficients. Due to linearity of the Radon transform, it follows that:

$$p\left[\theta_k, t_j\right] \cong \mathcal{R}\left\{\sum_{q=1}^{Q} c_q\ \Phi_q(\mathbf{x})\right\} = \sum_{q=1}^{Q} c_q\ \mathcal{R}\{\Phi_q(\mathbf{x})\}\ \ , \tag{2.33}$$

where $\left[\theta_k, t_j\right]$ refers to the discrete projection points of $\mathbf{f}$. Once $\Phi_q(\mathbf{x})$ are chosen, $\mathcal{R}\{\Phi_q(\mathbf{x})\}$ is determined regardless of the object under study (the object information is entirely encoded by the coefficients c_q's).

According to [99], the optimal $\Phi_q(\mathbf{x})$ has to be smooth and both space- and band-limited. If the basis functions are space-limited, the computation of $\tilde{f}(\mathbf{x})$ can be performed very efficiently. Requiring the basis function to be band-limited has to do with the finite nature of the data (one should not retrieve frequencies above the band limit of the data) and with the null space of the Radon transform. It has been shown that the functions living in the null space of the Radon transform (known as *ghosts*) have

Fourier transform characteristics complementary to those of band-limited functions [100]. In the field of signal processing, functions limited in some space are also called *windows*.
A window easy to compute and being (nearly) optimally concentrated in real and Fourier domain is the Kaiser-Bessel [101]:

$$\Phi_q^{(l)}(r) = \begin{cases} \dfrac{\left(\sqrt{1-(\frac{r}{a})^2}\right)^m I_l\left(\alpha\sqrt{1-(\frac{r}{a})^2}\right)}{I_l(\alpha)} & 0 \leqslant r \leqslant a \quad , \\ 0 & \text{otherwise} \end{cases} \tag{2.34}$$

where I_l denotes the l-th order modified Bessel function of the first kind, a is the support radius and α is the tapering parameter, controlling the trade-off between the width of the main lobe and the amplitude of the side lobes of the Fourier transform of the function. Since the Kaiser-Bessel window (2.34) is a radially symmetric function, the expression of its Radon transform does not depend on θ [99]:

$$\mathcal{R}\left\{\Phi_q^l(r)\right\}(t) = \frac{a}{I_l(\alpha)}\left(\frac{2\pi}{\alpha}\right)^{\frac{1}{2}} I_{l+1/2}\left(\alpha\sqrt{2-\left(\frac{t}{a}\right)^2}\right)\left[\sqrt{1-\left(\frac{t}{a}\right)^2}\right]^{l+1/2} \quad . \tag{2.35}$$

Another basis recently introduced for tomographic image reconstruction is the tensor-product B-spline [102], that has shown potential particularly for the case of DPC data [103]. Defining $\Delta_h^n f(x)$ the n-fold iteration of the finite difference operator $\Delta_h f(x) = (f(x+h/2) - f(x-h/2))/h$ and $x_+ = \max\{0, x\}$, $\beta^{(l)}$ is a centered univariate B-spline of degree l:

$$\beta^{(l)}(x) = \frac{\Delta_1^{l+1}}{l!} x_+^l \quad . \tag{2.36}$$

In this case, $\Phi_q^{(l)}(\mathbf{x}) = \beta^{(l)}(x_1)\beta^{(l)}(x_2)$ and the expression of the n-th derivative of the Radon transform is [102, 103]:

$$\mathcal{R}^{(n)}\left\{\Phi_q^{(l)}(\mathbf{x})\right\}(\theta, t) = \frac{\Delta_{\cos\theta}^{l+1}\,\Delta_{\sin\theta}^{l+1}}{(2l-n+1)}\, t_+^{2l-n+1} \quad . \tag{2.37}$$

2.3.3 Parallel beam geometry

All the techniques developed and/or analyzed in this work address the reconstruction of tomographic data in parallel beam geometry (PBG), The assumption is that X-rays are traveling parallel to each other on planes perpendicular to the rotation axis r. This allows to greatly simplify the X-ray projection model, since the contribution to the forward projection at a certain height $\bar{r}$ originates from the object information lying on the plane perpendicular to the r-axis and intersecting it at $(0, 0, \bar{r})$. Each 2D projection can, therefore, be regarded as a stack of independent projection lines being perpendicular to r. The ensemble of 1D projections at a fixed height collected in $[0, \pi)$ is called *sinogram*. Fig. 2.3

is an example of a sinogram, whose projections are stacked along the vertical axis (namely, the "view direction"; the horizontal axis is also called "channel direction").

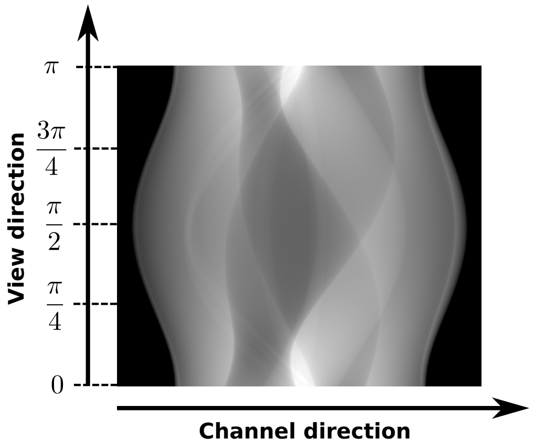


Figure 2.3: Example of a sinogram.

2.4 Analytical reconstruction

2.4.1 Fourier slice theorem

The Fourier slice theorem (FST) states that the 1D Fourier transform of a projection line acquired at angle θ with respect to a certain axis corresponds to the 2D Fourier transform of the object, sampled along a line that forms the same angle with respect to the same axis [12]. In mathematical terms, it means that:

$$\mathcal{F}_t\{\mathcal{R}\{f\}\}(\theta,\omega) = \mathcal{F}_{x_1,x_2}\{f\}(u_1,u_2)\Big|_{\substack{u_1=\omega\cos\theta\\ u_2=\omega\sin\theta}} \quad , \tag{2.38}$$

where

$$\begin{aligned} \mathcal{F}_t\{\mathcal{R}\{f\}\}(\theta,\omega) &= \int_{-\infty}^{+\infty} dt\ \mathcal{R}\{f\}(\theta,t)\ \exp(-2\pi i\omega t) \quad , \\ \mathcal{F}_{x_1,x_2}\{f\}(u_1,u_2) &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 f(x_1,x_2)\ \exp(-2\pi i(x_1u_1+x_2u_2)) \quad . \end{aligned} \tag{2.39}$$

The derivation (2.38) makes use of expression (2.28) of the Radon transform and of the coordinate change from (t,s) to (x_1,x_2) [12]:

$$\begin{aligned}
\mathcal{F}_t\{\mathcal{R}\{f\}\}(\theta,\omega) &= \int_{-\infty}^{+\infty} dt\ \mathcal{R}\{f\}(\theta,t)\ \exp(-2\pi i\omega t) = \int_{-\infty}^{+\infty} dt \left(\int_{-\infty}^{+\infty} ds\, f(t,s)\right) \exp(-2\pi i\omega t) \\
&= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, f(x_1,x_2)\ \exp(-2\pi i\omega(x_1\cos\theta + x_2\sin\theta)) \qquad . \\
&= \mathcal{F}_{x_1,x_2}\{f\}(\omega\cos\theta,\omega\sin\theta) \quad .
\end{aligned} \tag{2.40}$$

Equation (2.38) clearly shows that sampling the 1D Fourier transform of a projection line is equivalent to sampling the 2D Fourier transform of the object in a polar fashion. Theoretically speaking, the reconstruction of f demands the following three steps: (i) acquisition of an infinite number of projections in $[0,\pi)$; (ii) 1D Fourier transform of the projections; (iii) 2D inverse Fourier transform. In practice, implementing such an approach is not so trivial, because the 2D Fourier space of the object is sampled in a polar fashion by the 1D FFT-transformed projections and it is necessary to interpolate those samples on a Cartesian grid to apply the IFFT 2D. Fourier methods for tomographic reconstruction are treated in detail in Chapter 4.

2.4.2 Filtered backprojection

Filtered backprojection (FBP) is a reconstruction algorithm that descends directly from the FST. substitution with polar coordinates; split of the integral in two parts, one where $\theta \in [0,\pi/2]$, the other where $\theta \in [\pi/2,\pi]$; transformation of the second integral; recomposition of the two integrals to form again a single one; application of property 2.21; utilization of the FST.

$$\begin{aligned}
f(x_1,x_2) &= \int_{-\infty}^{+\infty} du_1 \int_{-\infty}^{+\infty} du_2\ \mathcal{F}\{f\}(u_1,u_2)\ \exp(-2\pi i(x_1u_1 + x_2u_2)) \\
&= \int_0^{2\pi} d\theta\ \underbrace{\int_0^{+\infty} d\omega\ \omega\ \mathcal{F}\{f\}(\theta,\omega)\ \exp(-2\pi i\omega(x_1\cos\theta + x_2\sin\theta))}_{:=D} = \int_0^{\pi} d\theta\, D + \int_{\pi}^{2\pi} d\theta\, D \\
&= \int_0^{\pi} d\theta\, D + \int_0^{\pi} d\theta' \int_0^{+\infty} d\omega\ \omega\ \mathcal{F}\{f\}(\theta'+\pi,\omega)\ \exp(+2\pi i\omega(x_1\cos\theta' + x_2\sin\theta')) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{substituted } \theta'=\theta-\pi \\
&= \int_0^{\pi} d\theta\, D + \int_0^{\pi} d\theta' \int_0^{+\infty} d\omega\ \omega\ \underbrace{\mathcal{F}\{f\}(\theta',-\omega)}_{\text{used property (23)}}\ \exp(+2\pi i\omega(x_1\cos\theta' + x_2\sin\theta'))
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\pi d\theta\, D + \int_0^\pi d\theta \int_{-\infty}^{0} d\omega' \,(-\omega')\ \mathcal{F}\{f\}\left(\theta,\omega'\right)\ \exp\left(-2\pi i\omega'\left(x_1\cos\theta + x_2\sin\theta\right)\right) \\
&\qquad\qquad\qquad\qquad \text{substituted } \omega'{=}-\omega \text{ and used again both } \theta\,,\,\omega \\
&= \int_0^\pi d\theta \int_{-\infty}^{+\infty} d\omega\, |\omega|\ \underbrace{\mathcal{F}\{f\}\,(\theta,\omega)}_{\text{apply FST}}\ \exp\left(-2\pi i\omega\left(x_1\cos\theta + x_2\sin\theta\right)\right) \\
&= \int_0^\pi d\theta \int_{-\infty}^{+\infty} d\omega\, |\omega|\ \mathcal{F}_t\left\{\mathcal{R}\{f\}\right\}(\theta,\omega)\ \exp\left(-2\pi i\omega\left(x_1\cos\theta + x_2\sin\theta\right)\right) && (2.41) \\
&= \int_0^\pi d\theta\ \mathcal{F}^{-1}\left\{|\omega|\right\}(t) * \mathcal{R}\{f\}(\theta,t)\ \ ,\ \ \text{with } t = x_1\cos\theta + x_2\sin\theta && (2.42) \\
&= \int_0^\pi d\theta \int_{-\infty}^{+\infty} dt'\ \mathcal{F}^{-1}\left\{|\omega|\right\}(t-t')\ \mathcal{R}\{f\}(\theta,t')\ \ ,
\end{aligned}
$$

where $*$ is the convolution operator. The passage from (2.41) to (2.42) is made possible by the convolution property of the Fourier transform, namely, $f * g = \mathcal{F}^{-1}\{\mathcal{F}\{f\}\mathcal{F}\{g\}\}$. (2.41) and (2.42) are equivalent continuous formulas of FBP. The difference lies in the inner integral corresponding to the *filtering* step, which is performed in Fourier domain for the first formula and in the real domain for the second one. The outer integral corresponds to the *backprojection* step.

The discretized form of (2.42), for example, is:

$$
\begin{aligned}
f\left[x_{1,m},x_{2,n}\right] &= \Delta\theta\Delta t \sum_{k=0}^{K-1}\ \sum_{j=-J/2}^{J/2-1} \mathcal{F}^{-1}\{|\omega|\}\left[x_{1,m}\cos\theta_k + x_{2,n}\sin\theta_k - t'_j\right] * \mathcal{R}\{f\}\left[\theta_k,t'_j\right] \\
&= \Delta\theta\Delta t \sum_{k=0}^{K-1} p^{(f)}\left[x_{1,m}\cos\theta_k + x_{2,n}\sin\theta_k\right]\ \ ,
\end{aligned}
\tag{2.43}
$$

where $p^{(f)}$ indicates the filtered projection and $\Delta\theta = \pi/K$ provided that the projectons are acquired at equispaced angular intervals. Generally, the argument of $p^{(f)}$ in (2.43) does not exactly coincide with the center of a detector cell and, therefore, an interpolation scheme is required.

One of the reasons why FBP has become a broadly adopted reconstruction algorithm for CT since its introduction (and for several decades since) is that interpolation is performed entirely in the real domain, differently from the Fourier methods briefly mentioned in 2.4.1. Interpolation errors in the real domain are localized in a neighborhood; interpolation errors in the Fourier domain are smeared back onto the entire image, once the inverse Fourier transform is computed. Considered an image

$\mathbf{f} \in \mathbb{R}^{M\times N}$ and its discrete Fourier transform (DFT) $\hat{\mathbf{f}}$, an error in Fourier domain can be thought as an impulse located at $(\bar{j}, \bar{k})$ and added to $\hat{\mathbf{f}}$:

$$\begin{aligned}
\tilde{f}[m,n] &= \sum_{j=0}^{M}\sum_{k=0}^{N} \exp\left(2\pi i j m/M\right) \exp\left(2\pi i k n/N\right) \left(\hat{f}\left[j,k\right] + a\delta\left[j-\bar{j}, k-\bar{k}\right]\right) \\
&= f[m,n] + a\, \exp\left(2\pi i \bar{j} m/M\right) \exp\left(2\pi i \bar{k} n/N\right) \\
&\underbrace{\sim f[m,n] + a\cos\left(2\pi * \left(\frac{\bar{j}m}{M} + \frac{\bar{k}n}{N}\right)\right)}_{\text{neglecting the imaginary part}} \quad . \qquad (2.44)
\end{aligned}$$

(2.44) clearly shows that the initial impulse located at $(\bar{j}, \bar{k})$ is transformed by the IDFT into a perturbation periodically added onto the entire image $\forall\, m = 0, 1, ..., M-1$ and $\forall\, n = 0, 1, ..., N-1$.

2.4.3 Filtering step

The filtering step acts separately on each projection, suppressing low frequencies (those closer to the origin of the Fourier domain), as shown by (2.41). The filter function $|\omega|$, coming from the Jacobian of the polar-Cartesian coordinate transformation, is called *ramp* or *Ram-Lak* [12].

To derive the discretized form of the ramp filter, one has first to assume the projections $p_\theta[t_j] = \mathcal{R}\{f\}[\theta, t_j]$[1] band-limited with bandwidth W, i.e. $\hat{p}_\theta[\omega_j] = 0$ if $|\omega_j| > W$. The *sampling theorem* [104] states that for a sampling period $\tau \leqslant 1/2W$ it follows [105]:

$$p_\theta(t) = \sum_{j=-\infty}^{+\infty} p_\theta\left[j\tau\right] \operatorname{sinc}\left(2\pi W(t - j\tau)\right) \quad , \qquad (2.45)$$

where $\operatorname{sinc} x := \sin x/x$. (2.45) is substituted inside the expression of the filtered projection, $p_\theta^{(f)}(t)$, in (2.41), where the integration interval is now limited to $[-W, W]$ [105]:

$$\begin{aligned}
p_\theta^{(f)}(t) &= \int_{-W}^{+W} d\omega\, |\omega|\, \mathcal{F}\{p_\theta(t)\} \exp\left(-2\pi i\omega t\right) = \int_{-W}^{+W} d\omega\, |\omega| \left(\int_{-\infty}^{+\infty} dt\, p_\theta(t) \exp\left(2\pi i \omega t\right)\right) \exp\left(-2\pi i\omega t\right) \\
&= \int_{-W}^{+W} d\omega\, |\omega| \left(\int_{-\infty}^{+\infty} dt' \sum_{j=-\infty}^{+\infty} p_\theta\left[j\tau\right] \operatorname{sinc}\left(2\pi W(t' - j\tau)\right) \exp\left(2\pi i\omega t'\right)\right) \exp\left(-2\pi i\omega t\right) \\
&= \sum_{j=-\infty}^{+\infty} p_\theta\left[j\tau\right] \int_{-W}^{+W} d\omega\, |\omega| \underbrace{\left(\int_{-\infty}^{+\infty} dt' \operatorname{sinc}\left(2\pi W(t' - j\tau)\right) \exp\left(2\pi i\omega t'\right)\right) \exp\left(-2\pi i\omega t\right)}_{:=D} \quad .
\end{aligned}$$

(2.46)

---

[1] The variable θ is not indexed here to simplify the notation. The discretization of the ramp filter involves exclusively the radial variable, t.

Integral D is related to the Fourier transform of the sinc function given by:

$$\mathcal{F}\{\operatorname{sinc} t\}(\omega) = \int_{-\infty}^{+\infty} dt \, \operatorname{sinc} t \, \exp(2\pi i \omega t) = \operatorname{rect}(\omega) := \begin{cases} 0 & |\omega| > 1/2 \\ 1/2 & |\omega| = 1/2 \\ 1 & |\omega| < 1/2 \end{cases} . \tag{2.47}$$

Using (2.47) to simplify D and considering the discretized version of the filtered projection, (2.46) becomes [105]:

$$\begin{aligned} p_\theta^{(f)}[k\tau] &= \frac{1}{2W} \sum_{j=-\infty}^{+\infty} p_\theta[j\tau] \int_{-W}^{+W} d\omega \, |\omega| \exp(-2\pi i \omega (k\tau - j\tau)) \\ &= \frac{1}{2W} \sum_{j=-\infty}^{+\infty} p_\theta[j\tau] \int_{-W}^{+W} d\omega \, |\omega| \left(\cos(2\pi\omega(k\tau - j\tau)) - \underbrace{i \sin(2\pi\omega(k\tau - j\tau))}_{\text{the integral of this function is 0}} \right) \\ &= \frac{W}{2} p_\theta[k\tau] + \frac{1}{W} \sum_{\substack{j=-\infty \\ j \neq k}}^{+\infty} p_\theta[j\tau] \int_{0}^{+W} d\omega \, |\omega| \cos(2\pi\omega(k\tau - j\tau)) \\ &= 2W \left(\frac{1}{4} p_\theta[k\tau] - \sum_{\substack{j=-\infty \\ j-k \in \mathbb{Z}_o}}^{+\infty} \frac{p_\theta[j\tau]}{\pi^2 (j-k)^2} \right) = p_\theta[j\tau] * h[j\tau] \ , \quad h[j\tau] = \begin{cases} 1/4 & j = 0 \\ 0 & j \in \mathbb{Z}_e \\ -1/j^2\pi^2 & j \in \mathbb{Z}_o \end{cases} , \end{aligned} \tag{2.48}$$

where $\mathbb{Z}_o$ and $\mathbb{Z}_e$ are the sets of odd and even integer numbers, respectively. The filter h and its DFT, $\hat{h}$, are plotted in Fig. 2.4. $\hat{h}$ resembles the ramp filter in the continuous formulation of FBP ($|\omega|$) with the difference that $\hat{h}[0] \neq 0$ (although very small), so that the filtering operation does not suppress the DC term of the projections [12].

The ramp filter compensates for the non-homogeneous sampling of the Fourier domain [12], when dealing with discrete projections. For the FST, the Fourier domain of the object is sampled in a polar fashion by the DFT of the projections, thus, it is more sparsely sampled at higher frequencies.

As shown in (2.41) and (2.42), projections can be either filtered in Fourier or real domain. The first solution can be efficiently implemented with FFT, provided that the data are first zero-padded to avoid the *wrap-around effect*. The convolution theorem is, indeed, applicable only to cyclic convolutions, i.e. the signals are assumed to be periodic. Hence, if there is no periodicity and FFT is used to implement the convolution, the tails of the signals wrap around, leading to undesirable artifacts.

Due to the high-pass nature of the ramp filter, the filtering step enhances the noise affecting the projections. As a matter of fact, the "pure" ramp filter should be employed only with almost noiseless data. In general, the ramp filter is multiplied with a window, that ignores low frequencies and suppresses the highest frequencies to a certain extent. Well-known examples of such windows are

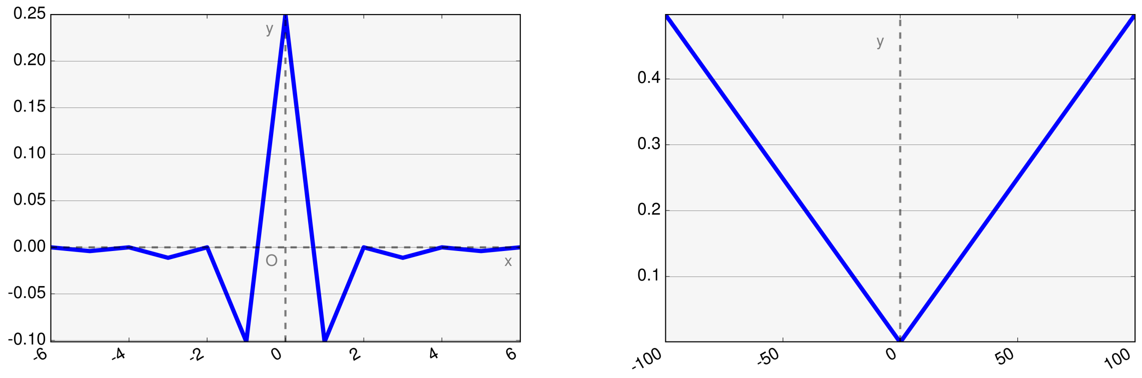
Shepp-Logan (abbr. "sl"), Hanning (abbr. "hn"), Hamming (abbr. "hm") and Parzen (abbr. "pz") [106], having the following expressions:

$$w_{\mathrm{sl}}(f) = \frac{2\dfrac{f}{f_c}}{\pi \sin\left(\dfrac{\pi|f|}{2f_c}\right)}$$

$$w_{\mathrm{hn}}(f) = \begin{cases} 0.5 + 0.5\cos\left(\dfrac{\pi f}{f_c}\right) & 0 \leqslant |f| \leqslant f_c \\ 0 & \text{otherwise} \end{cases}$$

$$w_{\mathrm{hm}}(f) = \begin{cases} 0.54 + 0.46\cos\left(\dfrac{\pi f}{f_c}\right) & 0 \leqslant |f| \leqslant f_c \\ 0 & \text{otherwise} \end{cases}$$

$$w_{\mathrm{pz}}(f) = \begin{cases} |f| - 6|f|\left(\dfrac{|f|}{f_c}\right)^2\left(1 - \dfrac{|f|}{f_c}\right) & 0 < |f| < \dfrac{f_c}{4} \\ 2|f|\left(1 - \dfrac{|f|}{f_c}\right)^2 & \dfrac{f_c}{4} < |f| < \dfrac{f_c}{2} \\ 0 & |f| \geqslant \dfrac{f_c}{2} \end{cases}$$



(a) (b)

Figure 2.4: (a) Filter $h[j\tau]$ given in (2.48) used to filter projections in real domain. (b) DFT of $h[j\tau]$ used to filter projections in Fourier domain.

where f_c is the *cutoff-frequency*, i.e. the high frequencies corresponding to $|f| > f_c$ are suppressed. Fig. 2.5, where the four window functions are compared in the interval [0,1], shows that high frequencies are suppressed the most by the Parzen window. Since high frequencies correspond either to noise or highly varying signal features, the conclusion is that FBP reconstructions with the ramp filter will have the best spatial resolution and the highest amount of noise, whereas those with the ramp filter combined to the Parzen window will have the worst spatial resolution and the lowest amount of noise. The filter function plays a crucial role in determining the reconstruction quality in terms of signal-to-noise ratio/spatial resolution trade-off. For this reason, it is always mandatory, especially in published works showing analytical tomographic reconstructions, to specify what kind of filter has been used for FBP.

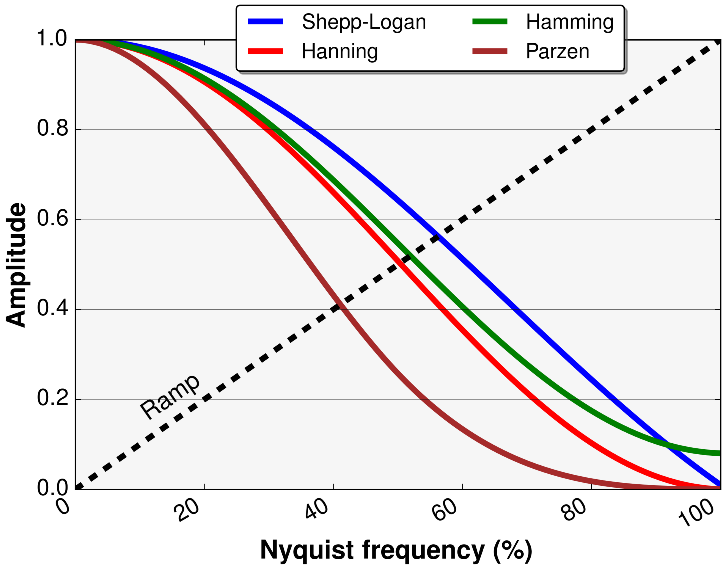


Figure 2.5: Shepp-Logan, Hanning, Hamming and Parzen windows at comparison.

2.4.4 Implementation aspects of the discrete backprojection

The discrete backprojection formula (2.44) is schematically depicted in Fig. 2.6: consider a single data point belonging to a filtered projection, $p^{(f)}[\theta, t]$; trace the X-ray line corresponding to that point, i.e. $r : x_1 \cos\theta + x_2 \sin\theta = t$; assign the (scaled) data point to each image pixel traversed by r; do the same operation for the other data points; repeat for all projections in $[0, \pi)$. All contributions from data points of different projections to the same image pixel are summed up. The backprojection of all data points leads to an approximated version of the object $\mathbf{f}$.

Backprojection $\mathcal{B}$ represents the *adjoint operator* of the Radon transform $\mathcal{R}$, i.e. $\mathcal{B} = \mathcal{R}^*$. Given

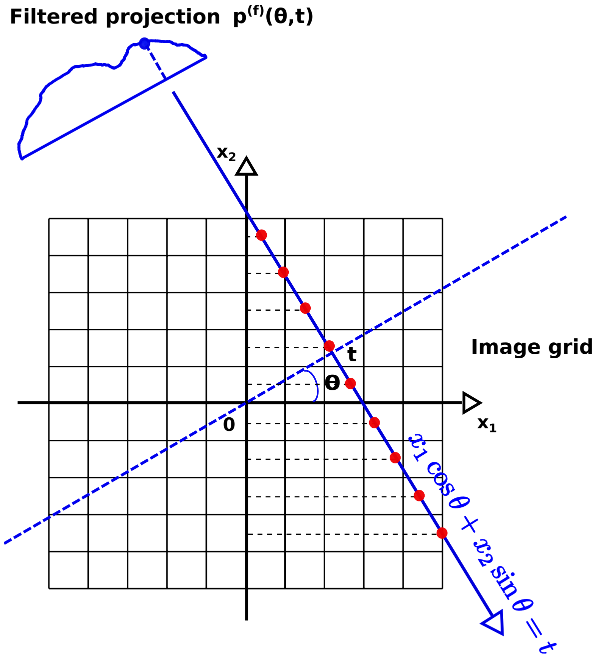


Figure 2.6: Sketch of the backprojection operation.

a generic linear operator $\mathcal{A} : \mathrm{U} \Longrightarrow \mathrm{V}$, with U, V being vector spaces, the adjoint operator, $\mathcal{A}^* : \mathrm{V} \Longrightarrow \mathrm{U}$, satisfies:

$$\Big\langle y \,, \mathcal{A}(x) \Big\rangle = \Big\langle \mathcal{A}^*(y) \,, x \Big\rangle \qquad \forall x \in \mathrm{U} \quad \forall y \in \mathrm{V} \quad . \tag{2.49}$$

From a computational point of view, the implementation of $\mathcal{A}^*$ consists of the same (all linear) operations involved by $\mathcal{A}$, but in reverse order and with input/output arrays exchanged (if $\mathcal{A}$ is complex, then, each adjoint calculation undergoes additional complex conjugation). This means that every implementation scheme described in 2.3.2 can be used to build a corresponding backprojector. The complexity of all these algorithms is, at least, $\mathcal{O}\left(N^3\right)$ [93–95, 97, 99, 102]. The computational hurdle of the backprojection operation can be greatly reduced through parallelized implementations on Graphics Processing Units (GPUs). Two well-known examples of Open Source softwares offering very fast GPU-based tomographic projectors have been published in [107, 108].

An algorithm characterized by much lower complexity is the *hierarchical backprojector* [109]. The idea is to split the reconstruction of an image $f^{(N)} \in \mathbb{R}^{N \times N}$ from a sinogram consisting, for example, of N views × N pixels into the reconstruction of 4 images with half number of pixels, namely, $f^{(N/2)} \in \mathbb{R}^{N/2 \times N/2}$. A fairly accurate reconstruction of each $f^{(N/2)}$ requires only $N/2$ views, since the pixels are halved in both dimensions. Backprojection for the 4 images $f^{(N/2)}$ requires approximately $4 \cdot (N/2)^3 = N^3/2$ floating operations. The recursive usage of this strategy of halving the image grid and the sinogram leads to an algorithm with complexity $\mathcal{O}\left(N^2 \log_2 N\right)$ [109], making the hierarchical backprojector much faster than any of the previously mentioned FBP implementations.

2.5 Iterative reconstruction

2.5.1 Underconstrained datasets

FBP has represented the standard CT reconstruction algorithm for decades, since its introduction. This method is easy to understand, works fast with a parallel implementation and provides high-quality reconstructions, when the sinogram is "properly" sampled.

Practical rules of thumb discussed in [12] suggest that the reconstruction of a $N \times N$ grid requires $M \approx N\pi/2$ projections homogeneously acquired in $[0, \pi)$, each sampled with $K \approx N$ data points. When $M \approx N\pi/2$ and $K \ll N$, the reconstruction is affected by thick streak artifacts due to the aliasing of single projections. Thin streaks appear in case $M \ll N\pi/2$ and $K \approx N$, giving rise to a characteristic star-shaped artifact [12]. In both scenarios, a sinogram in parallel beam geometry is regarded as *undersampled* for FBP.

Two sources of noise usually affect projection data: additive noise due to roundoff errors/electrical noise and shot noise, which is signal-dependent and follows a Poissonian statistics. Assuming the additive noise a stationary zero-mean random process uncorrelated for different projections, it is possible to show that [12]:

$$\sigma_r^2 = \sigma_p^2 \, \pi \int_{-\infty}^{+\infty} d\omega \, |\omega|^2 \, |w(\omega)|^2 \tag{2.50}$$

where σ_r is the standard deviation (std) of the FBP reconstruction, σ_p is the std of the projection data and $w(\omega)$ is the additional window function for the filtering step. Relation (2.50) shows that the smaller the area under $|\omega|^2 \, |w(\omega)|^2$, the smaller the error affecting the final reconstruction, but the bigger the distortions due to edge-blurring, as already stated in 2.4.3. An expression similar to (2.50) can be derived for the case of shot noise [12].

Although the previous arguments are helpful in understanding the FBP performance under data insufficiency or noise, the reconstruction algorithm needs always to be tested on experimental data acquired with the tomographic setup under study. Effects like the non-zero size of the detector aperture function or some kind of structured noise characterizing the data can highly deviate the performance of FBP (and any other reconstruction algorithm) from what theoretically expected. For example, a finite detector aperture function acts roughly as a low-pass filter, that can suppress the potential aliasing of the projection data [12].

In this work, the term *underconstrained* is used to generally identify tomographic datasets, for which analytical algorithms like FBP are unable to provide high quality reconstructions. In the experimental practice, underconstrained datasets result from low-dose fast scans or from tomographic setups, where projections cannot be acquired homogeneously in $[0, \pi)$. The study conducted in this thesis focuses on the first case. Low-dose scans enable in-vivo studies that are particularly relevant in biomedical research; fast scans allow tomography of quickly varying specimens of any kind. To decrease total dose and scan time for a given CT setup, two possibilities are available: acquiring a little number of projections, which leads to undersampled sinograms ($M \ll N\pi/2$)[2]; decreasing the exposure time per projection, that yields very noisy data. Another current challenge in SRXTM is the reconstruction of

---

[2] The case $K \ll N$ is of no interest for SRXTM, due to the high-resolution detectors employed for these kind of tomographic scans.

interior tomography data, arising when the irradiated object does not fit inside the CT field-of-view, leading to the acquisition of truncated projections. Undersampled, noisy and truncated datasets (or a combination of these issues), whose reconstruction cannot be "successfully" tackled by means of FBP, are all here referred as underconstrained.

2.5.2 Iterative methods

The following paragraph does not want to be an extensive review of iterative methods for tomographic reconstruction. Some key aspects and the various "families" of iterative algorithms are here listed.

Iterative algorithms are non-linear methods designed to provide better reconstruction accuracy than FBP when dealing with underconstrained datasets. Reconstructions are calculated through minimization of a cost function embedding a fidelity term, that steers the solution to fit the input data, and a regularization term, that pushes the solution to fulfill the expectations or a-priori knowledge of the object under study. Customary components are: a forward and a backprojection operator, an iterative solver, a regularization scheme, constraints, a stopping criterion and parameters (supervised or unsupervised) weighting the influence of each component. Iterative algorithms require more computations than FBP and the computational bottleneck usually lies in the few calls per iteration of the forward and backprojector. 3D compressed sensing based regularizations can easily represent a computational bottleneck as well (more on this aspect in 5.2.7).

Algebraic methods were the first iterative algorithms to have ever been introduced. They include the algebraic reconstruction technique (ART) [13], the simultaneous iterative reconstruction algorithm (SIRT) [14] and the simultaneous algebraic reconstruction technique (SART) [15]. Algebraic methods treat the tomographic problem as a system of equations, iteratively solved by the Kaczmarz method [16].

Bayesian methods incorporate the statistical model (usually a Poissonian or a compound of Poissonians) ruling the signal formation at the detector. The maximum likelihood expectation maximization (MLEM) [110], the penalized weighted least square method (PWLS) [17, 18, 111] belong to this group of algorithms.

Modern techniques for convex optimization like the split Bregman method [112] and the alternate direction method of multipliers (ADMM) [113] have recently been applied to tomographic reconstruction [114–116].

The projection-onto-convex-sets (POCS) method [117] is designed to find the intersection area of closed convex sets and has been mainly used to address the interior tomography problem combined to differentiated backprojection [118].

Discrete tomography algorithms [119] deal with the reconstruction of images whose domain is a discrete set. Some studies focus on the reconstruction of binary images [120, 121], others contemplate reconstruction problems involving any small number of grey levels [122, 123]. The discrete algebraic reconstruction technique (DART) [124] is a recent approach designed for discrete tomography, that combines both non-discrete and discrete update steps at each iteration.

Iterative methods incorporate regularization schemes to stabilize the inversion problem and to improve the reconstruction accuracy by exploiting some kind of a-priori knowledge regarding the object under study. Examples of "classical" regularization schemes are Tikhonov [125] and Huber [126] penalties. Regularization schemes based on compressive sensing [127] like total variation (TV) [128] are described in 5.2.7.

Chapter 3

Simulation framework

This chapter describes the simulation framework used in this work to test the performance of reconstruction algorithms. In particular, the following topics are analyzed: (i) how to choose simulated objects; (ii) how to compute underconstrained sinograms; (iii) how to properly assess the reconstruction accuracy.

Devising a reliable simulation framework represents the first non-trivial task of any research in the field of image processing/retrieval. The simulation framework is essentially a "measurement instrument", which is necessary to confirm theoretical expectations or to shed light when such expectations cannot be so easily formulated in advance. It also plays a fundamental role in comparing the accuracy of different methods designed to perform the same operation.

Despite its importance, very little attention has been dedicated to the problem of designing some kind of "universal" simulation framework, shared by all people working and publishing in the field. This may explain, why, in some cases, promising reconstruction algorithms have not received the necessary attention and others have, despite all their shortcomings.

It is not easy to offer a universal solution to this problem. Nevertheless, this chapter discusses crucial aspects that should be considered to get meaningful results when assessing the accuracy of a reconstruction algorithm.

3.1 Creating tomographic datasets

3.1.1 Phantoms

The term *phantom* is a synonym for "simulated object or image".

The Shepp-Logan (SL) phantom [129] simulates the map of attenuation coefficients of a head section (Fig. 3.1) and has been used in countless publications dedicated to reconstruction algorithms. Since the SL consists of the superposition of different roto-translated ellipses, its Radon transform can be computed analytically [12]. This is a very useful feature to test forward projection operators (the analytical sinogram works as a "ground truth") or to create sinograms without errors generated by a specific implementation of the Radon transform. The original phantom (SLO) (Fig. 3.1a) has

(a) Original Shepp-Logan (b) Modified Shepp-Logan

Figure 3.1: Original and modified Shepp-Logan (SL) phantoms.

very poor contrast[3], since internal structures can be barely discerned from each other. For some experiments, it is preferable to work with a modified version of the SL (SLM) (Fig. 3.1b) which is characterized by higher contrast and is better suited, e.g., to recognize the presence of reconstruction artifacts.

Unless the design of the algorithm is driven by a well-defined application, it is always preferable to test a reconstruction algorithm on a tomographic dataset composed of multiple phantoms rather than a single one. Following this strategy ensures that results obtained for a single phantom are not biased by its specific features. For example, an algorithm that performs well on the piece-wise-constant SL with curved corners will not necessarily perform satisfactorily for piece-wise-continuous objects with sharp corners.

Moreover, a set with different phantoms should be organized according to a criterion or feature for which the reconstruction algorithm has to be "measured", like, e.g., contrast or spatial information. Fig. 3.2 shows an example of dataset with varying contrast, obtained by multiplying the original thorax section for a monotonically decreasing sequence of numbers $\in (0, 1)$.

Given an image $\mathbf{f} \in \mathbb{R}^{M \times N}$, the spatial information (SI) is defined as follows [130]:

$$\mathrm{SI}_{ij} := \sqrt{(\mathbf{f} * \mathbf{h}_1)^2_{ij} + (\mathbf{f} * \mathbf{h}_2)^2_{ij}} \tag{3.1}$$

where $\mathbf{h}_1$ and $\mathbf{h}_2$ are the convolution kernels of the Sobel filter:

$$\mathbf{h}_1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad \mathbf{h}_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & +2 & 1 \end{bmatrix} \quad . \tag{3.2}$$

---

[3] The *contrast-to-noise ratio* (CNR) is defined and discussed in Section 3.2

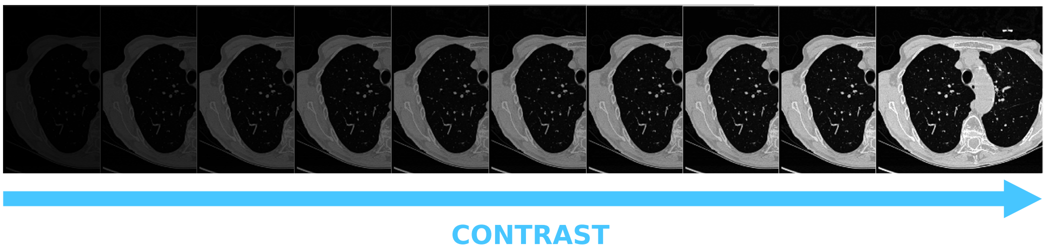


Figure 3.2: Illustrative example of dataset with varying contrast.

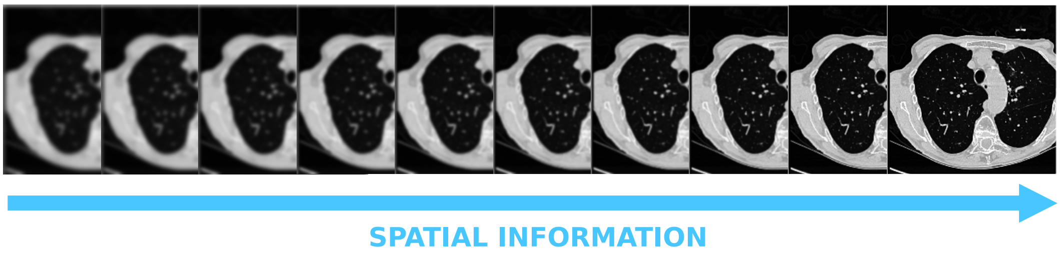


Figure 3.3: Illustrative example of dataset with varying spatial information.

SI measures the edge energy and can be also used to build a metric that estimates the image complexity [130]:

$$\mathrm{SI}_{\mathrm{mean}} = \frac{1}{NM}\sum_{i=0}^{N}\sum_{j=0}^{M}\mathrm{SI}_{ij} \tag{3.3}$$

$$\mathrm{SI}_{\mathrm{rms}} = \sqrt{\frac{1}{NM}\sum_{i=0}^{N}\sum_{j=0}^{M}\mathrm{SI}_{ij}^{2}} \tag{3.4}$$

$$\mathrm{SI}_{\mathrm{std}} = \sqrt{\frac{1}{NM}\sum_{i=0}^{N}\sum_{j=0}^{M}\mathrm{SI}_{ij}^{2} - \mathrm{SI}_{\mathrm{mean}}^{2}} \quad . \tag{3.5}$$

Figure 3.3 shows an example of dataset with varying spatial information, created by blurring the original phantom with a Gaussian kernel of increasing radius.

3.1.2 Sinograms

Simulated sinograms are computed by an implementation of the Radon transform (see 2.3.2), unless the object belongs to a family of functions that admits analytical forward projection.
The performance of a reconstruction algorithm can be properly assessed only if the error/deformation induced by the forward projector is negligible. Tests conducted on phantoms with analytical Radon transform can single out the best available forward operator. To further reduce the discretization error for a given projector, a possible strategy is to work on a k-times upsampled version of the phantom $\mathbf{f}^{(\mathrm{up})} \in \mathbf{R}^{kN\times kN}$ ($k \in \mathbb{N}_0$) to compute the sinogram $\mathbf{b}^{(\mathrm{up})} \in \mathbf{R}^{M\times kN}$; the target sino-

(a) Result of standard procedure. Mean diff. with SL $= -4.452\,10^{-4} \pm 0.147$

(b) Result of upsampled procedure. Mean diff. with SL $= 4.4139\,10^{-4} \pm 0.117$

Figure 3.4: Difference between the SL (512 × 512 pixels) and FBP reconstructions of two sinograms with 800 × 512 pixels. The first sinogram has been created by slant stacking with linear interpolation. The second sinogram has been created, first, by upsampling to 2048 × 2048 pixels with bicubic interpolation the SL, applying slant stacking with linear interpolation and, finally, downsampling the number of pixels by a factor 4 with no interpolation. The mean difference and standard deviation of the difference images are reported in their captions.

gram $\mathbf{b} \in \mathbf{R}^{M \times N}$ is, then, obtained by downsampling[4] $\mathbf{b}^{(\mathrm{up})}$ along the channel direction. Fig. 3.4 shows an experiment employing the slant-stacking forward projector with linear interpolation (formula (2.31)), where the upsampling strategy is compared to the case in which no upsampling is used: the mean difference between the original phantom and the FBP reconstruction of the sinogram created by the upsampling strategy has a standard deviation decreased by 20%. However, in reconstruction experiments involving underconstrained datasets, the absolute difference in accuracy between the reconstruction of a sinogram created with the upsampling strategy and a reconstruction of a standard sinogram usually does not play an effective role in biasing the benchmark of the algorithm under study. Therefore, for computational efficiency, it is preferable to avoid this trick.

Another important aspect is the degree of *coupling* between the implementation of the forward projector $\mathcal{R}$ and backprojector $\mathcal{B}$. When the actual implementation of $\mathcal{B}$ corresponds to the actual "adjoint" implementation of $\mathcal{R}$, the tomographic operators are strongly coupled from a computational point of view. The coupling effect is here shown with an experiment involving analytical reconstructions: consider a SL phantom with 512 pixels; compute well-sampled sinograms (805 views × 512 pixels) with the pixel- (PD), ray- (RD) and distance-driven (DD) forward projectors (see 2.3.2); reconstruct each sinogram with FBP consisting of ramp-filtering followed by the PD, RD and DD backprojector; calculate the average absolute difference between analytical reconstructions and SL phantom. Re-

---

[4] The procedure works fine if no interpolation is used to downsample the sinogram, otherwise artifacts may arise in the reconstruction.

	$\mathcal{R}$-PD	$\mathcal{R}$-RD	$\mathcal{R}$-DD
$\mathcal{B}$-PD	7.12e-2	9.90e-2	7.25e-2
$\mathcal{B}$-RD	1.00e-1	6.67e-2	9.74e-2
$\mathcal{B}$-DD	7.31e-2	9.65e-2	5.11e-2

Table 1: Average absolute difference between the analytical reconstructions and the SL phantom for the experiment showing the coupling forward-adjoint operator. PD, RD and DD are the abbreviations for pixel-, ray- and distance-driven approaches for either the forward projector, $\mathcal{R}$, or the backprojector, $\mathcal{B}$. The yellow color highlights the lowest value along each column. Ramp-filtering has been used for the reconstructions.

sults displayed in Tab. 1 show that, for a given forward projector, the highest accuracy, i.e. the lowest absolute difference along a column of the table (highlighted in yellow) is obtained when the reconstruction is performed by the corresponding backprojector. Moreover, the coupling effect plays a more decisive role than the accuracy of the standalone operators. For instance, [94, 98] report that the DD algorithm is more accurate than the classic PD. Nonetheless, the best reconstruction of the PD-sinogram is achieved by the ramp-filtered PD-backprojector. When testing the performance of a reconstruction algorithm on simulated data, it is, therefore, fundamental to compute the forward projection either analytically, if possible, or with a forward projector not coupled to the operators implemented inside the reconstruction algorithm itself[5]. In this way, the bias of accuracy due to the coupling effect is eliminated and the overall simulation will better describe the performance of the reconstruction algorithm on experimental data, where the "forward projector", being the X-ray setup, is surely not correlated to any PC-implementable tomographic operator. More details on the topic can be found in 4.5 or [30].

Underconstrained tomographic datasets are generally created by choosing a number of views $M \ll N \cdot \pi/2$ and by adding either Poisson or Gaussian noise with a certain variance σ. In theory, a Poisson model is suited to represent the noise affecting projection data that have been acquired with a photon-counting device [131]. If the CT detector is based on an energy integration system, a Poisson model can still represent a good approximation for the noise characterizing the sinogram. A compound of Poisson distributions can account both for a polychromatic X-ray beam and a camera with an energy integration system [132], but, due to its complex likelihood expression, it has been often approximated with either a Poisson or a Gaussian model [133]. Systematic experimental studies on the noise properties of X-ray CT data have shown that the effective sinogram noise follows a much more complex distribution than a Poisson or a Gaussian distribution, but, if a simple model is required, a Gaussian distribution is a better approximation than a Poisson distribution [134]. This and the fact that several works on phase contrast projections used Gaussian noise to conduct studies on simulated datasets (e.g. [135]) lead us to use a Gaussian and not a Poisson model for the experiments to follow. In any case, all these models neglect other sources of signal corruption like roundoff errors and electrical noise (see also 2.5.1). Projections are always homogeneously acquired in $[0, \pi)$, unless it is specified otherwise. The "limited-view" problem occurs when the experimental setup does not

---

[5] This would not represent, however, a perfect solution, as the forward projector, not coupled to any of the backprojectors, could be more coupled to one backprojector than to the other one, still creating a "dangerous" bias.

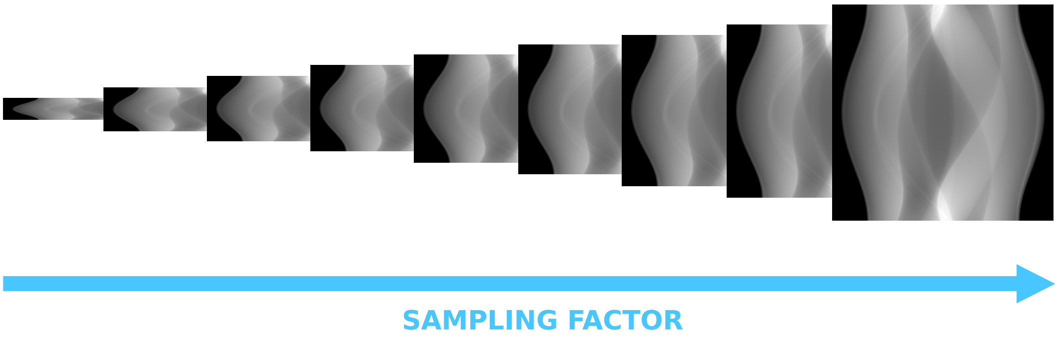


Figure 3.5: Illustrative example of tomographic dataset with varying sampling factor.

allow to cover the entire interval $[0, \pi)$. This kind of underconstrained dataset is simulated by leaving empty an entire "angular cone", e.g., centered around $\pi/2$. This work focuses on the reconstruction of underconstrained datasets with projections homogeneously distributed over the entire straight angle. M/N and σ are selected according to the phantom under study: if the phantom is characterized by sufficient structural complexity, a sinogram constructed without too small M/N and/or too large σ will already represent a tough challenge for FBP and, therefore, will be an interesting case of study for iterative reconstruction algorithms or any other approach alternative to FBP. The *sampling factor (SF)* is a quantity introduced here to describe the sampling level of a sinogram independently from its number of pixels. SF is defined as:

$$\text{SF} = \frac{M}{M_{\text{opt}}} \cdot 100\% \qquad , \quad \text{where} \quad M_{\text{opt}} := N\frac{\pi}{2} \quad . \tag{3.6}$$

Equation (3.6) shows that the SF corresponds to the percentage of projections required to achieve a FBP reconstructions free of undersampling artifacts. Figure 3.5 shows an example of dataset with varying SF.

3.2 Image quality metrics

Metrics for image quality are tools designed to measure the distortion of an image caused by the imaging system and to translate the visual perception of such degradation into numbers. Since images are rather complex objects to analyze, no single metric is able to effectively describe the overall quality. A complete analysis always requires the computation of multiple metrics focusing on different aspects of the image.

In the following, the metrics for image quality used throughout this work are introduced and described. Particular emphasis is put in explaining how to properly make use of such metrics, especially in the field of tomographic reconstruction.

3.2.1 SNR and CNR

The signal-to-noise ratio (SNR) [136] estimates the amount of noise affecting the image, whereas the contrast-to-noise ratio (CNR) [136] describes how well a certain material (tissue, phase or attenuation coefficient) can be visually discerned from the other ones. The computation of both figures of merit requires the identification of one (for the SNR) or two (for the CNR) regions of interest (ROIs), i.e. areas of the image characterized by the same material. Some a-priori knowledge about the investigated object is always required in order to draw the necessary ROIs.

Given an image $\mathbf{f}$, the SNR corresponds to the ratio between the mean value and standard deviation of a selected ROI R_1:

$$\text{SNR} = \frac{\bar{f}_{R_1}}{\sigma_{R_1}} \qquad \text{where} \quad \bar{f}_{R_1} = \frac{1}{N_{R_1}} \sum_{i \in R_1} f_i \quad \text{and} \quad \sigma_{R_1} = \sqrt{\frac{1}{N_{R_1}} \sum_{i \in R_1} \left(f_i - \bar{f}_{R_1}\right)^2} \quad . \tag{3.7}$$

Since $\bar{f}_{R_1}$ is fixed, the higher σ_{R_1}, i.e. the ROI noise, the lower the SNR.

The CNR corresponds to the ratio between the absolute difference of mean values and the sum of standard deviations of two different neighboring ROIs R_1 and R_2:

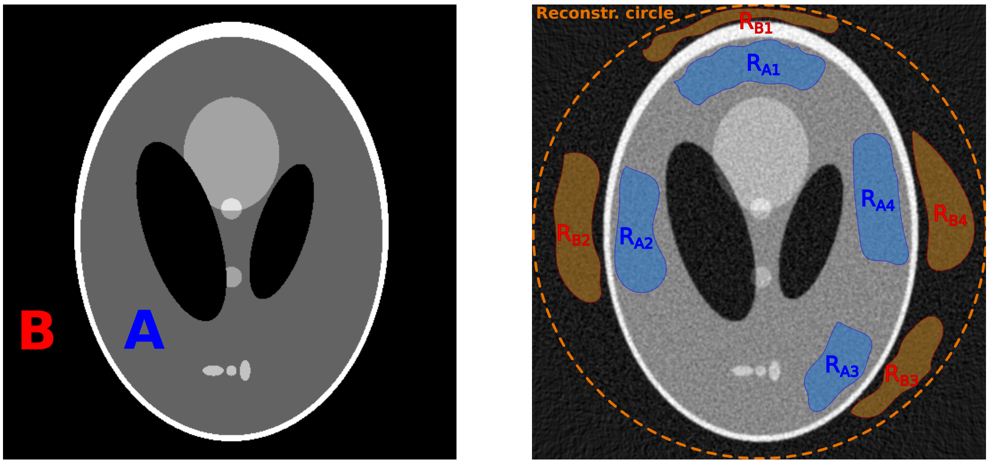
$$\text{CNR} = \frac{\left|\bar{f}_{R_1} - \bar{f}_{R_2}\right|}{\sigma_{R_1} + \sigma_{R_2}} \qquad \text{where} \quad \bar{f}_{R_k} = \frac{1}{N_{R_k}} \sum_{i \in R_k} f_i \quad \text{and} \quad \sigma_{R_k} = \sqrt{\frac{1}{N_{R_k}} \sum_{i \in R_k} \left(f_i - \bar{f}_{R_k}\right)^2} \quad . \tag{3.8}$$

When the attenuation coefficients of the materials characterizing R_1 and R_2 are similar, a lower amount of fluctuations in both ROIs suffices to suppress the CNR and vice versa.

In this work, the SNR and CNR assigned to an image correspond to averages of SNR and CNR values computed for different ROIs and couples of ROIs, respectively. For tomographic reconstructions, it is important to select ROIs at various distances from the center of the reconstruction circle, since the reconstruction quality is generally best at the center and worsens towards the limit of the FOV. Moreover, it is recommendable to select each time the largest possible ROIs to improve the statistics.

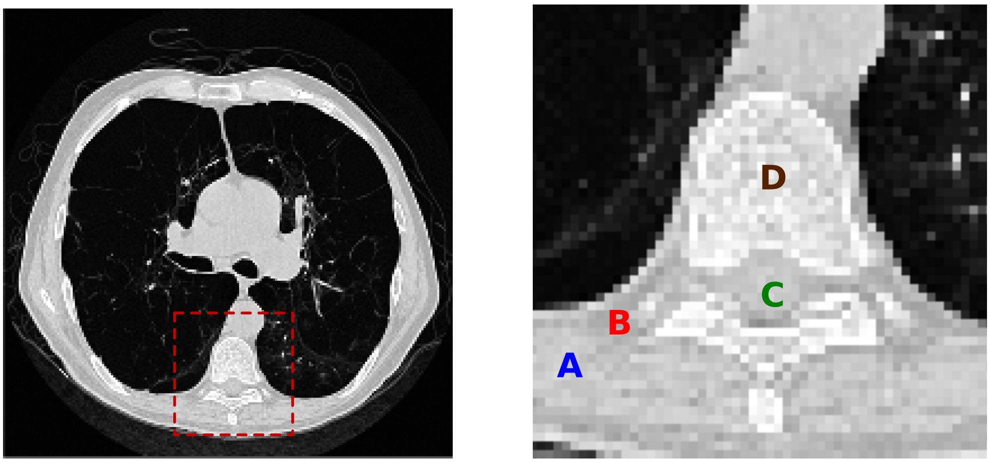
Figure 3.6 shows an example of SNR and CNR calculation for a simulated tomographic reconstruction of a SL phantom. The gray material indicated with "A" and the background indicated with "B" are targeted for this analysis (Fig. 3.6a). In Fig. 3.6b, eligible ROIs are overlaid to a FBP reconstruction of the phantom. Notice that all ROIs lie within the reconstruction circle (shown in Fig. 3.6b). In this example, $\text{SNR}_A = 1/4\,(\text{SNR}_{A1} + \text{SNR}_{A2} + \text{SNR}_{A3} + \text{SNR}_{A4})$; analogously one proceeds for SNR_B; $\text{CNR}_{A-B} = 1/4\,(\text{CNR}_{A1-B1} + \text{CNR}_{A2-B2} + \text{CNR}_{A3-B3} + \text{CNR}_{A4-B4})$. The SL phantom is well suited for the selection of ROIs to estimate SNR and CNR, because it is piece-wise-constant and each material occupies relatively large areas.

An example of phantom not suited to SNR and CNR calculation is displayed in Fig. 3.7a, which is a simulated tomographic slice of a human thorax. The phantom detail in Fig. 3.7b shows how variegated it is, for example, the white material on the bottom: subregions with supposedly different tissues (labeled with "A", "B", "C" and "D") can be roughly identified, even though it is hard to define their contours. If the SNR-CNR analysis is run on ill-defined ROIs, the standard deviation(s) required by formula (3.7) and (3.8) will not only represent the noise affecting the signal, but also the characteristic heterogeneity of the object under study. Thus, the conduction of a reliable SNR-CNR



(a) SL phantom (b) FBP reconstr. with ROIs

Figure 3.6: Example of selection of ROIs for SNR and CNR estimation. (a) SL phantom with indicated the materials, A (intermediate grey) and B (black, background), selected for the analysis. (b) FBP reconstruction of the SL phantom with 4 (blue) ROIs for A and 4 (red) ROIs for B used to compute SNR and CNR. The orange dashed curve delimits the reconstruction circle: ROIs have to be selected within this circle.



(a) Thorax phantom (b) Detail

Figure 3.7: Example of a phantom not suited to the selection of ROIs for SNR and CNR estimation. (a) Thorax phantom; the dashed square delimits the detail zoomed on the right. (b) Zoomed detail of the "white material" on the bottom.

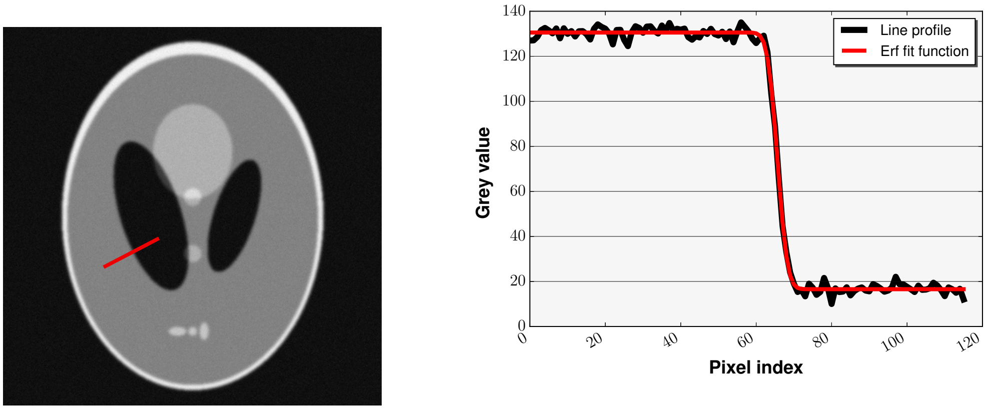
analysis requires the investigated phantom/specimen to be as much piece-wise-constant as possible and, in general, precise a-priori knowledge regarding the composing materials is strongly recommended. Since the user has to carefully draw the ROIs by hand, SNR and CNR can be computed only for a limited amount of slices and regions within each slice. Altogether, these metrics can be applied only to specific cases and can only assess the quality of a selected subset of the reconstructed volume, but, differently from metrics requiring a reference image, they are always applicable to real data.

3.2.2 Spatial resolution

The spatial resolution quantifies the sharpness of an image. A metric for spatial resolution is supposed, for example, to compute a value in pixel units for each image of Fig. 3.8, concluding that the SL phantom on the left is sharper than the one on the right. The lower the calculated value in pixel units, the higher the spatial resolution of the image under study. The spatial resolution is a feature strongly connected to the point-spread-function (PSF) of a linear imaging system. In the same way, for a fixed tomographic setup, reconstruction algorithms can be thought as either linear (analytical algorithms) or non linear (iterative algorithms) "imaging systems" and assessing the spatial resolution of the reconstructed slices can provide an estimate of the intrinsic algorithm "PSF".
A simple method to estimate the spatial resolution is the analysis of the edge slope [136]. An ideal linear imaging system has point-wise PSF everywhere and preserves step-like edges [72]. A real imaging system has PSF of finite size changing from region to region of the image and step-like edges are mapped into uphills/downhills with a certain slope. Fitting a line crossing an edge perpendicularly with an error function (erf) of the form $y_i = a_1 + a_2 \,\text{erf}[(x_i - a_3)/a_4]$ for $i = 0, ..., n-1$ (example in Fig. 3.9), where $\mathbf{a} = \{a_i\}$ are the fit parameters and (x_i, y_i) the data points, can provide a measure of the edge slope and, therefore, of the spatial resolution at that specific location. The parameter values used here to initialize the least square fit are $\mathbf{a}^{(0)} = \{\min(y_i), y_{n-1} - y_0, 1/2\,(x_{n-1} - x_0), 1.0\}$. The erf derivative is a bell extending from the start, x_s, to the end point, x_e of the erf slope. Once the fit function $y^{\text{fit}}(x, \mathbf{a}^{\text{fit}})$ is obtained, the first and last abscissa where $dy^{\text{fit}}/dx \neq 0$ correspond to x_s and x_e. The erf slope is finally computed as $(y^{\text{fit}}(x_e) - y^{\text{fit}}(x_s))/(x_e - x_s)$. For a tomographic reconstruction, this procedure should be repeated on several edge profiles at different distances from the image center and the single results are averaged. Analogously to what discussed for SNR and CNR calculation, the described method has the drawback of not being unsupervised and the analysis


(a) (b)

Figure 3.8: SL phantom with 512 × 512 pixels convolved to a Gaussian kernel of radius 2 (a) and 6 (b).



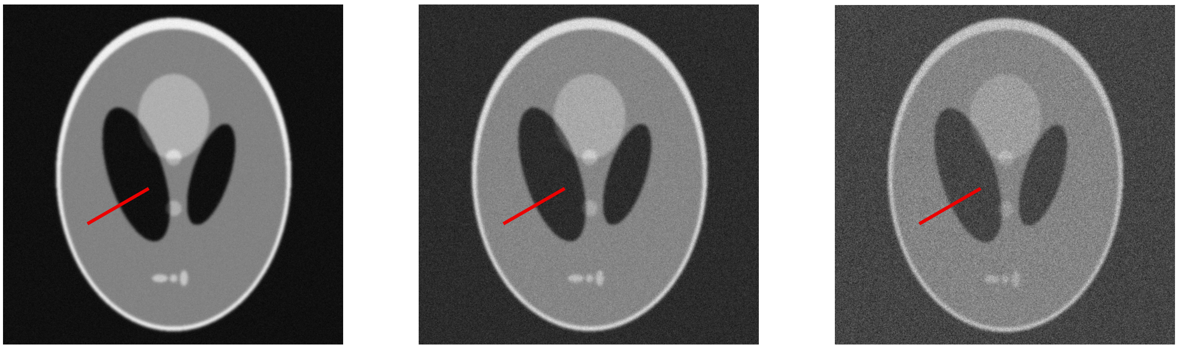
(a) Blurred and noisy SL phantom

(b) Erf fit of line profile

Figure 3.9: Example of estimation of spatial resolution through erf fit of a line profile. (a) SL phantom with 512 × 512 pixels is convolved to a Gaussian kernel of radius 2 and Gaussian noise has been added. The red segment indicates the position of the line profile used for the resolution analysis. (b) Line profile (in black); fit by erf (in red).

can practically be conducted only over a limited portion of the reconstructed volume.

A metric for spatial resolution should be sensitive to the level of blurring and the amount of noise. One of the major issues with the edge-slope-fitting procedure is that a very low SNR can impair the estimation of the spatial resolution: the fit is sensitive to the choice of $\mathbf{a}^{(0)}$ and for a noisy edge profile $a_0^{(0)} = \min(y_i)$ will not necessarily represent the right translation factor along y or



(a) SR=3.2 pix.

(b) SR=2.4 pix.

(c) SR=2.5 pix.

Figure 3.10: Simulation showing the dependence of the edge profile fitting method on the amount of noise affecting the image. All SL phantoms with 512 × 512 pixels were convolved to a Gaussian of radius=2. SR stands for spatial resolution, measured in pix.= pixel units. (a) Phantom plus additional Gaussian noise with $\sigma = 3\%\,\bar{p}$, $\bar{p}$ = average SL value. (b) Phantom plus additional Gaussian noise with $\sigma = 10\%\,\bar{p}$. (c) Phantom plus additional Gaussian noise with $\sigma = 20\%\,\bar{p}$.

$a_1^{(0)} = y_{n-1} - y_0$ will not precisely correspond to the slope height. Figure 3.10 shows three SL phantoms characterized by the same level of blurring and different amounts of Gaussian noise. The edge-slope-fitting procedure should return increasing values of spatial resolution going from the right to the left; instead the image in Fig. 3.10a has the worst result.

A different approach to estimate the spatial resolution is the *Fourier ring correlation (FRC)* [137, 138]. The FRC works on single images, its extension to volumes is called Fourier shell correlation (FSC) [139, 140]. The FRC estimates the spatial resolution by measuring in Fourier space the correlation between two scans of the sample [137]. It has been mainly applied in single particle analysis for transmission electron microscopy and cryo-electron microscopy. A perfect microscope would be able to generate identical projections of the sample at different instants; a real microscope, instead, provides an uncorrelated pattern, due to distortions and noise, that changes the sample projections at every instant. The FRC procedure is depicted in Fig. 3.11: consider two independently acquired images of the same object, I_1 and I_2; compute their FFT-2D (the DC term is placed at the grid center), $F_i = \text{fftshift}(\text{fft2}(I_i))$ and $i = 1, 2$; start from the center and at each time construct rings of fixed width δr at increasing distances r from the center; consider all Fourier components of F_1 and F_2 inside each ring to compute a single point of the FRC curve according to [137]:

$$\text{FRC}(r) = \frac{\sum\limits_{r_i \in R(r,\Delta r)} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum\limits_{r_i \in R(r,\Delta r)} |F_1(r_i)|^2 \cdot \sum\limits_{r_i \in R(r,\Delta r)} |F_2(r_i)|^2}} \,, \tag{3.9}$$

where the notation $r_i \in R(r, \Delta r)$ indicates all pixel elements r_i within the ring R, placed at distance r, with thickness Δr. The FRC curve starts from 1.0 at $r = 0$ (the DC components of the two images should coincide) and goes to zero as r approaches the Nyquist frequency, due to the fact that the uncorrelated components of the projections mainly affect the high frequencies.

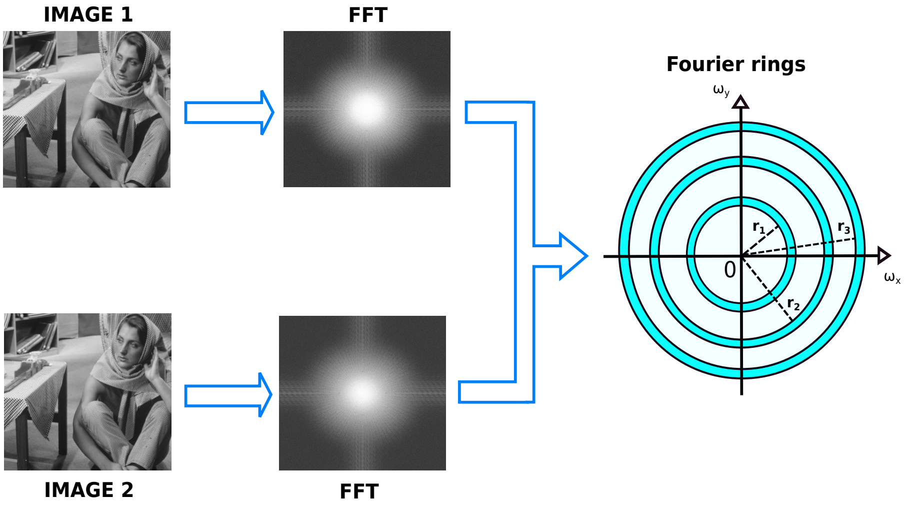


Figure 3.11: General scheme of FRC.

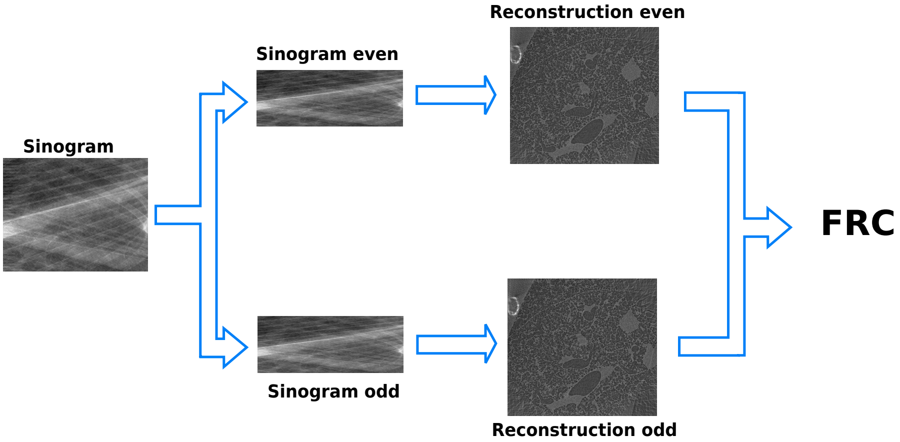


Figure 3.12: Data splitting strategy to compute FRC (FSC) for tomography.

To obtain a value of spatial resolution from the FRC curve, different criteria can be applied. There are criteria based on fixed-value thresholds like the "0.5-threshold" criterion [140], that estimates the spatial resolution from the abscissa corresponding to FRC=0.5. The "σ-curve" criterion computes the intersection between the FRC curve and $\sigma(r) = 3/\sqrt{n(r)/2}$, where $n(r)$ is the number of pixels of $R(r, \Delta r)$; the curve $\sigma(r)$ describes the expected theoretical pure-noise behavior of the Fourier cross-correlation coefficient [137, 141]. The "one-bit" and "half-bit" criteria compute the intersection of the FRC curve with $T_{1\,\text{bit}}(r) = (0.5 + 2.41/\sqrt{n(r)})/(1.5 + 1.41/\sqrt{n(r)})$ and $T_{1/2\,\text{bit}} = (0.21 + 1.91/\sqrt{n(r)})/(1.21 + 0.91/\sqrt{n(r)})$, respectively, [141]. When the FRC curve meets $T_{1\,\text{bit}}(r)$, the SNR of the combined images corresponds to unity and the average information per pixel is 1 bit. The "half-bit" criterion is just a less stringent version of the "one-bit" criterion [141]. Differently from what occurs in microscopy, it is usually not possible to acquire two distinct tomographic datasets of the same sample, due to time and dose constraints. A solution (shown in Fig. 3.12) is to split the sinogram into the set of even and odd projections and to reconstruct each set to obtain two independent representations of the sample [142, 143]. This strategy ensures that each sinogram is homogeneously sampled in $[0, \pi)$. When the number of projections is the resolution-limiting factor, the resolution of each sub-tomogram is expected to be half of the full-tomogram resolution [142]. The presence of noise in the data is accounted through a correction based on SNR considerations [142]:

$$\text{FRC}^*(r) = \frac{2\text{FRC}(r)}{\text{FRC}(r) + 1} \quad . \tag{3.10}$$

This adaptation of the FRC (FSC) analysis to tomographic datasets works if the spatial resolution SR satisfies the following relationship [142]:

$$\text{SR} > \text{SR}_{\text{crit}} = 2\,\text{SR}_{\text{ang}} \approx d\theta \cdot D \quad , \tag{3.11}$$

where $d\theta$ is the angular increment between consecutive projections and D is the diameter of the object under study. This condition can be also explained from another point of view: tomographic reconstructions of strongly undersampled datasets are affected by aliasing artifacts, equally affecting

both images submitted to the FRC analysis, and these artifacts give rise to an erroneous high correlation.
In Fig. 3.11, as example, the FRC analysis is used to compare the spatial resolution of FBP reconstructions with ramp filter + Hanning window using ray-driven (FBP-RD) and distance-driven (FBP-DD) backprojector. The sinogram of SL, used for this experiment, has 804 views × 512 pixels and additional Gaussian noise with $\sigma = 2.0\%$ average of the noiseless sinogram. According to the described procedure, the sinogram is split into distinct datasets, one collecting even-indexed projections, the other odd-indexed projections; both sub-sinograms are, then, reconstructed with FBP-RD and FBP-DD. Figure 3.13c shows a comparison between the FRC curve of the reconstructions with FBP-RD and FBP-DD: for each radius r, the higher the curve the better the performance of the algorithm in terms of spatial resolution. In this case, FBP-DD performs better than FBP-RD starting from $r \approx 0.3$. In Fig. 3.13d, 3.12e and 3.12f, the "0.5-threshold", "one-bit" and "half-bit" criterion are applied to the FRC curve of FBP-RD to determine a value for the spatial resolution, being 3.3 pixels, 2.8 pixels and 2.1 pixels, respectively. To compute the intersection with any criterion-curve, the FRC data points are fitted with a polynomial of degree 10. In this example, the FRC analysis can be considered reliable because the number of projections is sufficiently high to satisfy (3.11). If the task is to measure the spatial resolution of reconstructions of strongly undersampled datasets, FRC may not provide reliable results.
Compared to the edge-fitting method, the FRC (FSC) analysis has the upside of being unsupervised, i.e. no ROI or line has to be selected by the user. On the other hand, it is not clear how the FRC can provide an estimate for the spatial resolution of a tomographic reconstruction and, in particular, which criterion should be adopted to obtain a result that has a physical meaning. In practice, the FRC (FSC) can be used only for relative assessments, as in Fig. 3.11, Furthermore, a caveat to bear in mind is that the FRC (FSC) analysis works properly if the distortions characterizing the two distinct input datasets can be considered uncorrelated, otherwise a higher spatial resolution will be wrongly measured. Considering its clear connection to the spatial resolution, the edge-fitting method and not the FRC is used to analyze the reconstructions presented in the next chapters, when needed.

3.2.3 Full reference metrics

The term "full reference" identifies a family of unsupervised metrics, that measure the difference between the image (volume) under study and a ground-truth. The full reference metrics presented here and utilized for experiments in the next chapters are the mean squared error (MSE), the peak signal-to-noise ratio (PSNR) [144], the structural similarity index (SSIM) [145] and the normalized mutual information (NMI). In the following, $\mathbf{f} \in \mathbb{R}^{M \times N}$ indicates the image to analyze and $\mathbf{r} \in \mathbb{R}^{M \times N}$ is the oracle, ground-truth or reference image.

The MSE, probably the most known figure of merit in the field of signal processing, is defined as:

$$\text{MSE} = \frac{1}{MN} \sum_{i=0}^{M} \sum_{j=0}^{N} (f[i,j] - r[i,j])^2 \quad , \tag{3.12}$$

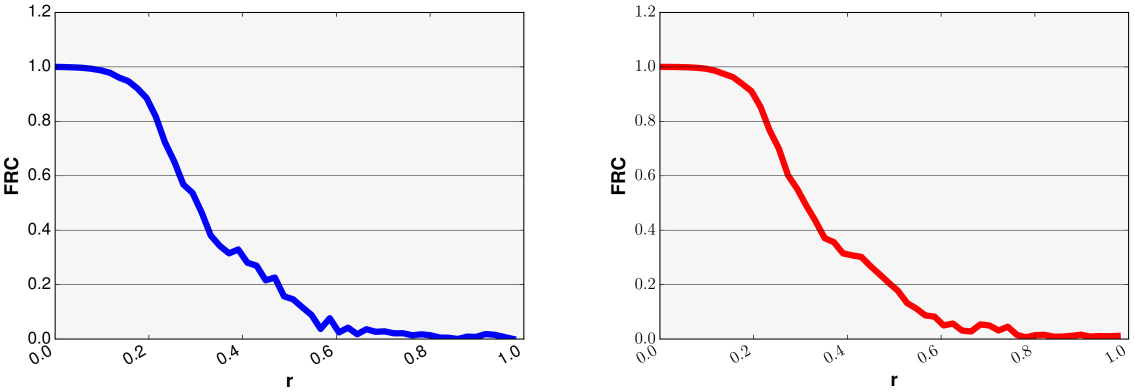
and the root-mean-square error (RMSE) is:

$$\mathsf{RMSE} = \sqrt{\mathsf{MSE}} = \sqrt{\frac{1}{MN}\sum_{i=0}^{M}\sum_{i=0}^{N}\left(f[i,j]-r[i,j]\right)^2} \quad . \tag{3.13}$$

The PSNR is a modified version of the MSE/RMSE measured in decibels (dB) [144]:

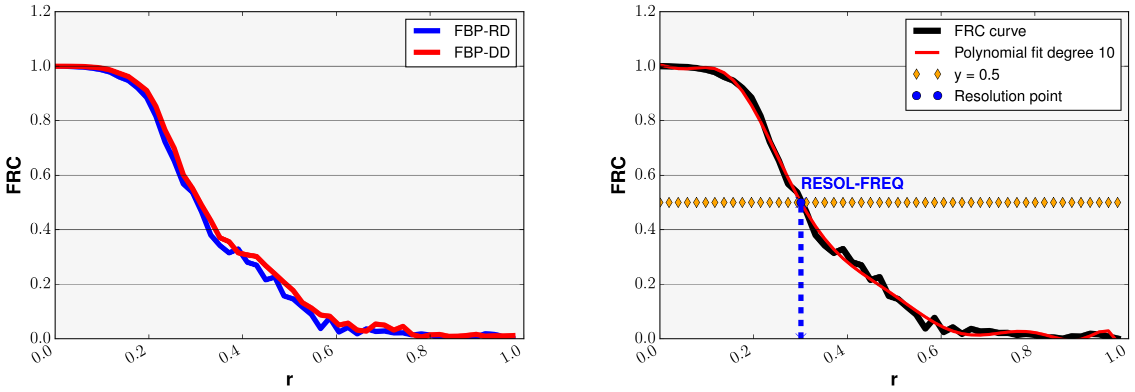
$$\mathsf{PSNR} = 10\,\log_{10}\left(\frac{\max\{\mathbf{r}\}^2}{\mathsf{MSE}}\right) = 20\,\log_{10}\left(\frac{|\max\{\mathbf{r}\}|}{\sqrt{\mathsf{MSE}}}\right) \quad . \tag{3.14}$$

The higher the MSE/RMSE, the lower the accuracy of $\mathbf{f}$ with respect to $\mathbf{r}$; the opposite holds for the PSNR. Placing the MSE term at the denominator enables the PSNR to be generally more sensitive to differences between $\mathbf{f}$ and $\mathbf{r}$ than the MSE and RMSE themselves.



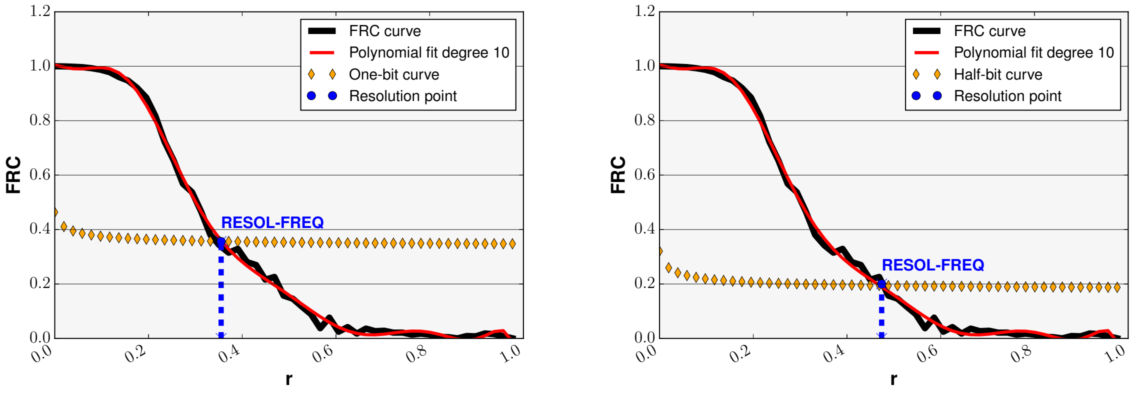
(a) FRC curve of FBP-RD

(b) FRC curve of FBP-DD



(c) Both FRC curves

(d) 0.5 crit. for FBP-RD



(e) 1 bit crit. for FBP-RD

(f) 1/2 bit crit. for FBP-RD

Figure 3.11: Example of FRC analysis applied to measure the spatial resolution of FBP reconstructions with ray-driven (FBP-RD) and distance-driven (FBP-DD) backprojector using Hanning filter. (a-b) Single FRC curves for each algorithm. (c) Comparison between the two FRC curves. (d-f) Resolution point computed on the FRC curve of FBP-RD with three different criteria: 0.5, 1 bit and 1/2 bit criterion, respectively.

The SSIM synthesizes comparisons of luminance, contrast and structural information between $\mathbf{f}$ and $\mathbf{r}$, that have been conducted separately in a patch-wise fashion [145]. The result of this analysis is a map of SSIM values, one for each pixel, ranging from 0 (lowest similarity) to 1 (highest similarity). Working on patches, rather than on the entire image at once, allows to better account for local changes of the three aforementioned quantities. The SSIM map is useful to know which areas of $\mathbf{f}$ differ the most from the reference. To convert the SSIM map into a single number, the mean value or MSSIM can be computed. Considered the patch $\mathbf{f}_{pq}$ ($\mathbf{r}_{pq}$) of $\mathbf{f}$ ($\mathbf{r}$), with $m \times n$ pixels, centered around the pixel (p, q), the SSIM value is given by [145]:

$$\begin{aligned} \text{SSIM}\left(\mathbf{f}_{pq}, \mathbf{r}_{pq}\right) &= l\left(\mathbf{f}_{pq}, \mathbf{r}_{pq}\right) \cdot c\left(\mathbf{f}_{pq}, \mathbf{r}_{pq}\right) \cdot s\left(\mathbf{f}_{pq}, \mathbf{r}_{pq}\right) \\ &= \frac{\left(2\mu(\mathbf{f}_{pq})\mu(\mathbf{r}_{pq}) + C_1\right)\left(2\sigma(\mathbf{f}_{pq}, \mathbf{r}_{pq}) + C_2\right)}{\left(\mu(\mathbf{f}_{pq})^2 + \mu(\mathbf{r}_{pq})^2 + C_1\right)\left(\sigma(\mathbf{f}_{ij})^2 + \sigma(\mathbf{r}_{pq})^2 + C_2\right)}, \end{aligned} \tag{3.15}$$

where l, c and s stand for luminance, contrast and structural information, respectively. Here, given generic images $\mathbf{g}, \mathbf{h} \in \mathbf{R}^{m \times n}$, $\mu(\mathbf{g})$, $\sigma(\mathbf{g})$ and $\sigma(\mathbf{g}, \mathbf{h})$ are defined as follows:

$$\mu(\mathbf{g}) = \frac{1}{mn} \sum_{i=0}^{m} \sum_{j=0}^{n} g[i, j] \qquad \sigma(\mathbf{g}) = \left(\frac{1}{(m-1)(n-1)} \sum_{i=0}^{m} \sum_{j=0}^{n} \left(g[i, j] - \mu(\mathbf{g})\right)^2 \right)^{\frac{1}{2}}$$

(a) Original image (b) MSE=107.01 | MSSIM=0.76 (c) MSE=107.01 | MSSIM=0.03

Figure 3.12: Simulation readapted from [146]. (a) Original image. (b) Image summed to a positive constant, t. (c) Image summed to $t \cdot \mathrm{rand}\{-1, 1\}$. At visual inspection, the degradation of (c) is considerably more severe than that of (b). The MSE fails at identifying the difference between (b) and (c). The MSSIM, instead, succeeds.

$$\sigma(\mathbf{g}, \mathbf{h}) = \left(\frac{1}{(m-1)(n-1)} \sum_{i=0}^{m} \sum_{j=0}^{n} (g[i,j] - \mu(\mathbf{g}))\,(h[i,j] - \mu(\mathbf{h})) \right)^{\frac{1}{2}} \quad .$$

C_1 and C_2 in (3.15) are constants chosen on the basis of the dynamic range of the data.
The SSIM was devised to replace the MSE (and similar figures of merit) as a metric better capable to mirror the perceived image quality. Several simulations reported in [146] prove that the MSE can fail at mimicking the human vision, when trying to assess the degradation of an image. One of such simulations is reproposed in Fig. 3.12. The original pepper image (Fig. 3.12a) is altered in two different ways: the version in Fig. 3.12b is obtained by summing a constant positive step, t, to the original image; the version in Fig. 3.12c is computed by summing t multiplied by either +1 or -1, which is randomly chosen pixel-wise. At visual inspection, Fig. 3.12c is clearly more degraded than Fig. 3.12b with respect to the original image. This difference is identified by the MSSIM, but not by the MSE. The distorted images of Fig. 3.12 are said to belong to an "equal MSE hypersphere".
Despite the initial enthusiasm of the imaging community towards this metric, several studies have questioned the effective superiority of the SSIM with respect to more classical figures of merit. [145] and [147] show that the PSNR and SSIM perform very similarly the same in identifying image distortions. [148] proves that the structural term of the SSIM, namely s in (3.15), does not represent the perceived structural content of the image. [149] finds no evidence of the superior performance of the SSIM when assessing video quality. [150] not only shows that it is possible to create an "equal SSIM hypersphere" through simulations, where the roles of MSE and SSIM in the example of Fig. 3.12 are reversed, but also demonstrates the existence of a direct mathematical link between MSE and SSIM, reaching the conclusion that the two metrics are essentially "cut from the same cloth".
Putting this controversy aside for the moment, it is important to point out that MSE, RMSE and PSNR are fast to calculate and parameterless, whereas the SSIM requires longer run times and features parameters, that can highly affect the accuracy of the metric: the constants C_1 and C_2; the patch

size; the radius of the Gaussian blurring kernel applied to each patch before computing the SSIM [145]. Moreover, the strategy of working patch-wise to produce an accuracy map can, in principle, be easily extended to MSE, RMSE and PSNR, if needed.

The NMI, mainly used as similarity measure for image registration [151, 152], is given by [153]:

$$\text{NMI} = \frac{I(\mathbf{r}, \mathbf{f})}{H(\mathbf{r})} \quad , \tag{3.16}$$

where H, the (Shannon) entropy, and I, the mutual information, are defined as follows:

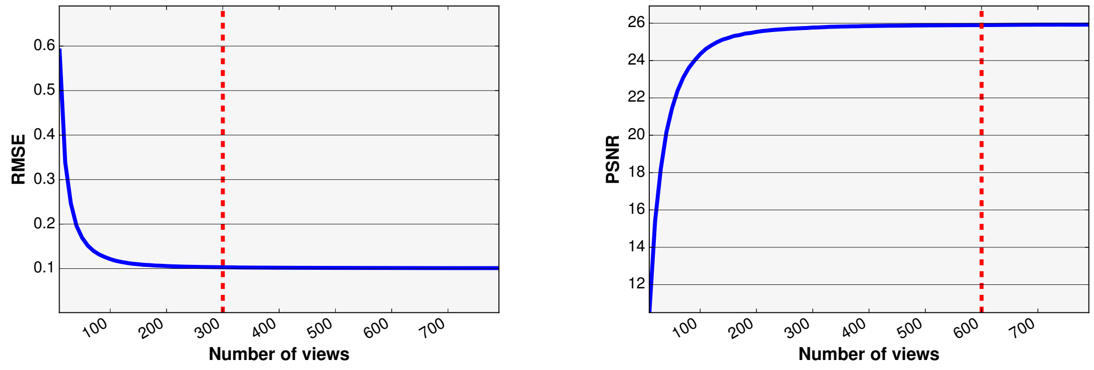
$$H(X) = -\sum_{x} p(x) \log_2(p(x)) \qquad I(X, Y) = H(X) - H(X|Y) = \sum_{x,y} p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad . \tag{3.17}$$

X and Y are generic variables and $p(x)$ is the probability that X assumes the value/state x. If $\mathbf{f}$ (and $\mathbf{r}$) is quantized in n bins, the i–th bin corresponds to the value $f^{(i)} = \min\{\mathbf{f}\} + i\,(\max\{\mathbf{f}\} - \min\{\mathbf{f}\})/n$ and $p(f^{(i)})$ is the number of counts inside the bin at $x = f^{(i)}$ of the normalized histogram. $p(r^{(i)}, f^{(j)})$ is the joint probability distribution or joint histogram of the two images. The NMI, analogously to the SSIM, ranges from 0, no mutual information between $\mathbf{f}$ and $\mathbf{r}$, to 1, maximal mutual information due to $\mathbf{f} \equiv \mathbf{r}$. This metric has the appealing feature of depending only on one parameter, the number of bins n, and to offer a similarity measure based on principles of information theory, differently from MSE and SSIM.

The aforentioned trap of the "equal hyperspheres" can occur only when dealing with images affected by several different kinds of distortions. When the accuracy of a tomographic reconstruction algorithm is under study, for example, the possible distortions affecting the retrieved image/volume are known: fluctuations in piece-wise constant ROIs, poor contrast, blurred edges or disappereance of small features.

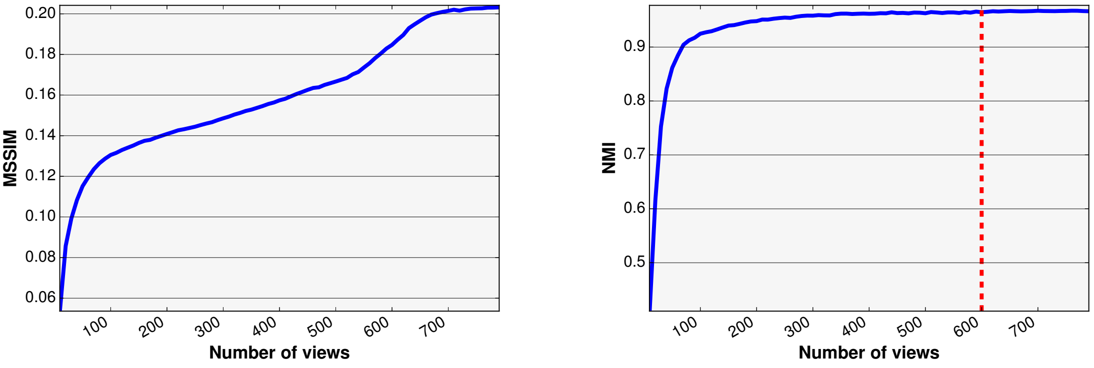
In principle, all the presented full reference metrics can be valid tools to benchmark a reconstruction algorithm. A good strategy is to devise ad-hoc experiments to know which metrics are reliable and sensitive enough for the study that we have in mind.

Consider, for example, to study the accuracy of different backprojectors for analytical reconstruction of undersampled sinograms. Which full reference metrics should be selected for this task? To answer this question, the following experiment is conducted: noiseless sinograms with varying numbers of projections, created by the slant stacking algorithm, are reconstructed by means of FBP with ramp filtering and pixel-driven backprojector (linear interpolation). The FBP reconstructions are analyzed with MSE, PSNR, MSSIM and NMI and results are reported in Fig. 3.13. The ideal metric should feature either a strictly decreasing (in case of RMSE) or increasing (PSNR, MSSIM, NMI) plot. Results show that all metrics are reliable, i.e the quality of a reconstruction with less projections is never superior than that of a reconstruction with more projections. However, the sensitivity is rather different: the RMSE saturates starting from around 300 projections (red dashed line in Fig. 3.13a); the PSNR and the NMI saturate starting from around 600 projections (red dashed line in Fig. 3.13b) and 3.13d); the MSSIM curve does not feature any saddle nor saturation point (Fig. 3.13c). In this



(a) RMSE results

(b) PSNR results



(c) MSSIM results

(d) NMI results

Figure 3.13: Example of experiment to test full reference metrics. Sinograms of SL with 512 pixels and number of views going from 10 to 800 (at steps of 10) are reconstructed with FBP using ramp filter and pixel-driven backprojector. Reconstructions are compared to the original SL phantom through RMSE, PSNR, MSSIM and NMI. The red dashed lines indicate when the metric saturates, i.e., the point from where it stops identifying differences with the previous reconstruction.

case, the MSSIM is the best suited metric to assess analytical reconstructions of undersampled data. For other types of experiments, the MSSIM may not represent the most reliable and sensitive tool to adopt, as it depends from case to case.

3.2.4 A final comment

Although metrics for image quality are a fundamental tool to conduct research in the field of image processing, a lot of work is still ahead and no infallible or generally applicable analysis framework has been devised yet. The sole reasonable approach is to handle every case separately to find the set of metrics, that can measure at best the image distortions occurring in the considered experiment.

The experimental parts presented in the following chapters are characterized by rather heterogeneous analysis frameworks. Metrics were chosen according to the aforementioned criterion or to simply meet the requests of reviewers, particularly affectioned to one or the other figure of merit.

Chapter 4

Gridding projectors

The *gridding projectors* are Fourier-based implementations of $\mathcal{R}$ and $\mathcal{B}$ with complexity $O(N^2 \log_2 N)$. These operators are based on the gridding method, a very old technique in signal processing dating back to the late 60s.

The chapter opens with the mathematical formulation of the gridding method. The gridding projectors and their usage for iterative reconstruction are, then, described. The following section discusses future perspectives of the gridding operators in the context of the current research. The last section is dedicated to the experimental study on the coupling effect between forward and backprojector.

4.1 The gridding method

4.1.1 Mathematical formulation

The gridding method (GM) is a signal processing technique designed to retrieve a signal from samples of its Fourier transform located on a non-Cartesian lattice [154]. The term "gridding" indicates per se the operation of mapping samples from an input grid onto a target grid by means of an interpolation scheme. The peculiarity of the GM is that data-gridding and the Fourier transform are both involved in the procedure.

The GM has been mainly applied to radio-astronomy [155–157] (starting from the late 60s), magnetic resonance imaging (MRI) [154, 158–161] (starting from the early 90s), CT analytical reconstruction [162–169] (starting from the mid 80s) and CT iterative reconstruction [28, 170, 171] (starting from 2004).

Consider a 2D real function $f(x_1, x_2)$ with (continuous) Fourier transform $F(u_1, u_2)$ and a sampling function $S(u_1, u_2) = \sum_i \delta(u_1 - u_1^{(i)}, u_2 - u_2^{(i)})$, where the set of points $(u_1^{(i)}, u_2^{(i)})$ form a generic grid in Fourier space. In the GM, the sampled Fourier transform, $F^{(s)} = F \cdot S$, is convolved with a function $W(u_1, u_2)$ (also called "kernel") and sampled again onto a Cartesian grid, described by the shah or comb function $\text{III}(u_1, u_2) = \sum_i \sum_j \delta(u_1 - i, u_2 - j)$ [158, 163]:

$$F^{(scs)} = \left[F^{(s)} * W \right] \cdot \text{III} = \left[(F \cdot S) * W \right] \cdot \text{III} \quad , \tag{4.1}$$

where the superscript "scs" summarizes the three steps: (input F) sampled (by S), convolved (with W), sampled (by III). By applying the inverse Fourier transform to the first and third member and considering that $\mathcal{F}^{-1}\{\text{III}(u_1, u_2)\} = \text{III}(x_1, x_2)$ (for a unit-spaced grid), Eq.(4.1) becomes [158]:

$$f^{(scs)} = [(f * s) \cdot w] * \text{III} \quad , \tag{4.2}$$

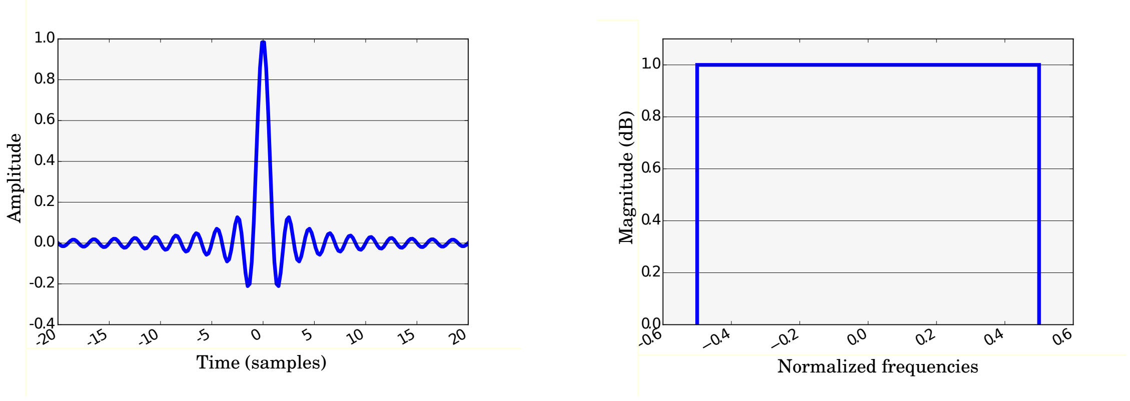
where s and w are the inverse Fourier transforms of S and W. If $f^{(scs)}$ is supposed to approximate f, two aspects have to be amended in Eq.(4.2): (i) $f^{(scs)}$ needs to be corrected for the uneven sampling in Fourier space [156]; (ii) the contribution of the window w has to be removed from $f^{(scs)}$ [163]. The uneven sampling in Fourier space is accounted by re-weighting S in Eq.(4.1) with the so called *density compensation factors (DCFs)* [158]:

$$F^{swcs} = \left[\left(F \cdot \left(\frac{S}{S * W} \right) \right) * W \right] \cdot \text{III} \quad . \tag{4.3}$$

The additional "w" in the superscript stands for "weighted" and $\rho := S/(S * W)$ are the sought DCFs. By computing, once again, the inverse Fourier transform, Eq.(4.3) becomes [158]:

$$f^{(swcsd)}(x, y) = \left\{ \left[f * \left(s *^{-1} (s \cdot w) \right) \cdot w \right] * \text{III} \right\} \cdot \frac{\Theta(x, y)}{w(x, y)} \quad , \tag{4.4}$$

where $*^{-1}$ is the symbol for deconvolution and $\Theta(x, y)$ is the box function used to crop the reconstruction ROI, i.e. $\Theta(x, y) = 1$ for $|x| < 1/2$, $|y| < 1/2$ and $\Theta(x, y) = 0$ otherwise. The removal of the window function contribution from the reconstruction (division per w) is also called *deapodization*. (4.4) represents the GM equation.



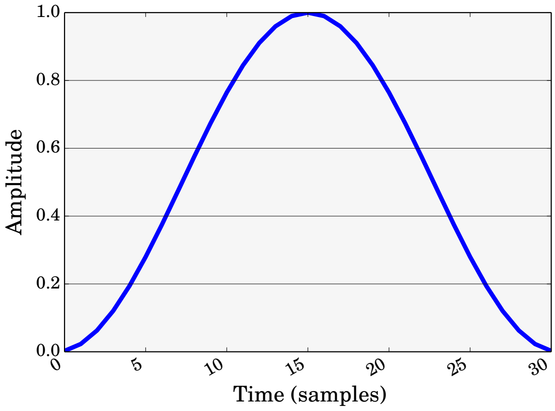
(a) Normalized sinc function

(b) $\mathcal{F}^{-1}$ of sinc function

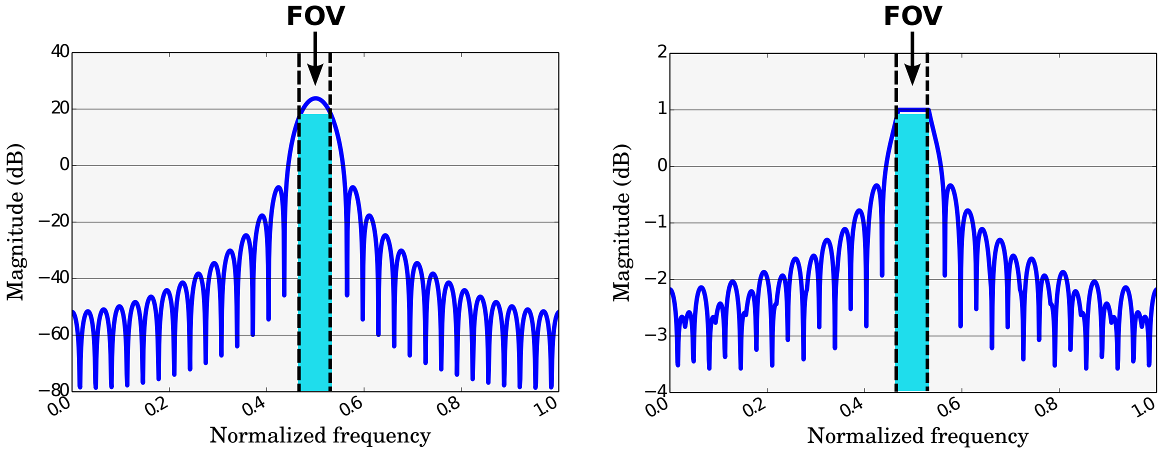
Figure 4.1: Normalized sinc function (a) and its Fourier transform (b) corresponding to the box function $\Theta(x, y)$. Since the box function is constant, the GM with a sinc kernel does not require deapodization. The sinc function has no compact support and, therefore, all available samples in Fourier space have to be used to compute each interpolated point. This makes the GM with a sinc kernel computationally not practical.

According to the Nyquist-Shannon theorem, the optimal kernel W for a properly bandlimited interpolation problem is the sinc function (Fig. 4.1a), that does not require deapodization, because its Fourier transform corresponds to the box function (Fig. 4.1b). Since the sinc kernel has no compact support, all samples available in Fourier domain contribute to the calculation of the new interpolated point [163]. As a consequence, the convolution results extremely slow. If the kernel has compact support, only few samples are needed to compute the interpolated ones and the procedure is computationally efficient.

The inverse Fourier transform of a compact kernel is not compact, features a non-constant main



(a) Hanning window



(b) Spectrum with FOV

(c) Deapodized spectrum

Figure 4.2: The Hanning function is used here as example of a compact kernel eligible for the GM. (a) Hanning kernel. (b) Kernel spectrum with FOV placed inside the main lobe and sidelobes aliasing back into the FOV. (c) Kernel spectrum after deapodization: the central region of the main lobe is now constantly equal to 1.0.

lobe, that has to be corrected through deapodization, and is characterized by sidelobes aliasing back into the FOV. An example of compact kernel is the Hanning function [172], displayed in Fig. 4.2a. Figure 4.2b shows the Hanning spectrum, the main lobe containing the FOV and the sidelobes. The deapodization removes the kernel contribution by flattening the main lobe to 1.0 (Fig. 4.2c), but in doing so it enhances the height of the sidelobes, as the comparison of the y-axis scales in Fig. 4.2b and Fig. 4.2c shows. The stratagem to avoid such artifacts is to perform the convolution in Fourier space with an oversampled grid [158, 163]. Oversampling by a factor $\alpha > 1.0$ increases the main lobe (and FOV) width and reduces the height of the sidelobes, as more energy will be concentrated inside the main lobe. Once the deapodization is applied, the enlarged FOV still suffers from aliasing artifacts, but the target reconstruction is placed in a central artifact-free area corresponding to the original FOV, which is finally cropped from the rest.

The recommended oversampling ratio for the GM is $\alpha = 2.0$ [158, 163]. This value comes from simple arguments of band-limited interpolation theory [163]. Consider an object $f(x_1, x_2)$ with compact support, such that $f(x_1, x_2) = 0$ for $|x_1| > \bar{x}_1$ or $|x_2| > \bar{x}_2$. Its continuous Fourier transform, $F(u_1, u_2)$, is measured on a Cartesian grid with spacings $\bar{u}_1$, $\bar{u}_2$. The sampled Fourier transform is, therefore, $F^{(s)}(u_1, u_2) = F(u_1, u_2) \cdot \text{Ш}(u_1/\bar{u}_1, u_2/\bar{u}_2)$. It follows that [163]:

$$\begin{aligned} f^{(s)}(x_1, x_2) &= \mathcal{F}^{-1}\left\{F^{(s)}(u_1, u_2)\right\} = \mathcal{F}^{-1}\left\{F(u_1, u_2) \cdot \text{Ш}(u_1/\bar{u}_1, u_2/\bar{u}_2)\right\} \\ &= \bar{u}_1 \bar{u}_2 \cdot f(x_1, x_2) * \text{Ш}(x_1 \bar{u}_1, x_2 \bar{u}_2) = \sum_i \sum_j f\left(x_1 - \frac{i}{\bar{u}_1}, x_2 - \frac{j}{\bar{u}_2}\right) \quad . \end{aligned} \tag{4.5}$$

Equation (4.5) shows that $f^{(s)}$ corresponds to a combination of infinite translated copies of the original function. To ensure that the cropped region $|x_1| < \bar{x}_1 \cup |x_2| < \bar{x}_2$ of $f^{(s)}$ equals f, no overlap should occur among the different copies. This condition is fulfilled only if each copy extends maximum to half area $\bar{u}_1 \cdot \bar{u}_2$, i.e. $\bar{x}_1 <= 1/2\bar{u}_1$ and $\bar{x}_2 <= 1/2\bar{u}_2$ as depicted in Fig. 4.3. Setting $\alpha = 2$ allows to sample more densely the Fourier domain and to work with an oversampling grid whose spacings are $\bar{u}_1 = 1/2\bar{x}_1$ and $\bar{u}_2 = 1/2\bar{x}_2$.

4.1.2 Optimal kernels

The GM performance is entirely determined by the kernel characteristics: if W is compact enough, the convolution in Fourier space is computationally efficient; if the kernel inverse Fourier transform, w, is well concentrated under the main lobe, the aliasing artifacts arising from the sidelobes are less severe and the GM yields a more accurate reconstruction. Of course, it is not possible to simultaneously require compactness in real and Fourier domain, but a compact kernel, whose energy in dual space is maximally concentrated inside a reasonably small region, can still be found. A metric designed to measure and compare the accuracy of different GM kernels is the *aliasing error* defined as [158]:

$$\epsilon(i) = \sqrt{\frac{\sum_{p \neq 0} [w(i + Gp)]^2}{[w(i)]^2}} \quad , \tag{4.6}$$

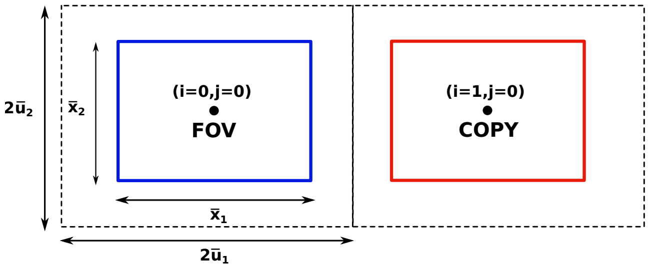
where G is the dimension of the oversampled grid used in Fourier space.
The following discussion is taken from Section IIB and IIC of [28].
A kernel of *prolate spheroidal wavefunctions (PSWFs)*, studied by Slepian in the early 60s [173], is by construction the most compact in real and Fourier domain. Consider the operators P_T and P_Ω, that create respectively time- and band-limited functions:

$$P_T f(t) = \Theta_T f(t) \qquad P_\Omega f(t) = \mathcal{F}_\omega^{-1}\left[\Theta_\Omega \mathcal{F}_t(f(t))\right]$$
$$\Theta_T(t) = \begin{cases} 1 & \text{if } |t| \leqslant T/2 \\ 0 & \text{if } |t| > T/2 \end{cases} \qquad \Theta_\Omega(\omega) = \begin{cases} 1 & \text{if } |\omega| \leqslant \Omega/2 \\ 0 & \text{if } |\omega| > \Omega/2 \end{cases} \tag{4.7}$$

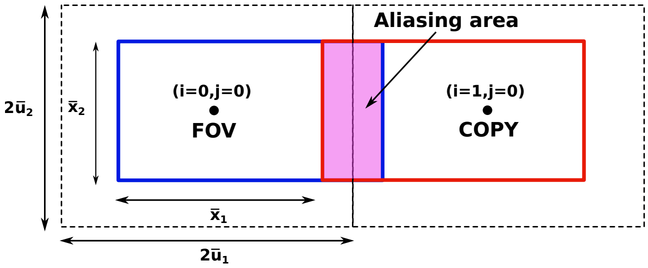
The prolate spheroidal wavefunctions (PSWFs) are the eigenfunctions of the self-adjoint operator $P_\Omega P_T$ [173]:

$$P_\Omega P_T \psi_i(t) = \int_{-T/2}^{T/2} dt' \, \frac{\sin(\Omega(t-t'))}{\pi(t-t')} \psi_i(t') = \lambda_i \psi_i(t) \quad . \tag{4.8}$$

If ψ_i's are bandlimited, the PSWF of 0-th order, ψ_0, with the largest eigenvalue λ_0, has the highest amount of energy $E_0 = \|P_T\psi_0\|^2$ in the interval $[-T/2, T/2]$ [173]. Although the PSWFs represent



(a) Antialiasing condition fulfilled



(b) Antialiasing condition not fulfilled

Figure 4.3: Equation (4.5) shows that $f^{(s)}$ is given by a sum over infinite translated copies of f. To ensure that $f^{(s)}$ within the FOV equals f, the condition $\bar{x}_1 < 1/2\bar{u}_1 \cup \bar{x}_2 < 1/2\bar{u}_2$ has to be fulfilled, otherwise aliasing artifacts will arise and $f^{(s)}_{FOV} \neq f$. (a) Case where the condition is fulfilled. (b) case where the condition is not fulfilled along the x_1-axis.

the optimal kernel for GM, the definition in Eq.(4.8) does not offer a closed-form expression, i.e., a mathematical expression that can be evaluated in a finite number of operations. The difficulty to compute the PSWFs was known also to the early literature of GM [163]. An iterative method utilizing the operators P_T and P_Ω was proposed as a viable non-efficient solution in [158]. However, depending on the magnitude of the parameter $c = T\Omega/2$ [173], (4.8) can be approximated to find a GM-specific formula for the PSWFs. Since T should not be larger than ± 1 to guarantee an accurate deapodization correction, this window function is best approximated using spherical Bessel functions, expressed as a normalized linear combination of even Legendre polynomials in Fourier space [173]:

$$\hat{\psi}_0^{(a)}(k) = \frac{\sum\limits_{n=0}^{N_a} c_{2n} L_{2n}(k)}{\sum\limits_{n=0}^{N_a} c_{2n} L_{2n}(0)} \quad , \tag{4.9}$$

where the superscript $^{(a)}$ stands for "approximated", L_n is the Legendre polynomial of degree n and $\{c_{2n}\}$ are the expansion coefficients. The inverse Fourier transform of a series of even Legendre polynomials corresponds to a series of even spherical Bessel functions, that can themselves be expanded in the interval $[-1, 1]$ in terms of even Legendre polynomials [174, 175]. According to this argument, $\psi_0^{(a)}(x)$ represents a scaled version of $\hat{\psi}_0^{(a)}(k)$:

$$\psi_0^{(a)}(x) = C \cdot \hat{\psi}_0^{(a)}\left(x\frac{S}{G}\right) \quad , \tag{4.10}$$

where S is the density of the convolution LUT and $C \in \mathbb{R}_0$ is a constant.
A handy alternative to the PSWFs are the Kaiser Bessels (KBs) [101]. The KBs are very compact and the associated aliasing error is inferior to that of many tested kernels [158], although they are slightly sub-optimal compared to the PSWFs. The closed-form expression of the KBs allows to optimize of their shape as a function of the oversampling [168]. The formula for the KB window and its inverse Fourier transform are [101]:

$$\hat{h}(k_x) = \frac{G}{s_w} I_0\left(\beta\sqrt{1-\left(\frac{2Gk_x}{s_w}\right)^2}\right) \quad \text{for } |k_x| \leqslant \frac{s_w}{2G} \ , \tag{4.11}$$

$$h(x) = \frac{\sin\sqrt{\left(\frac{\pi s_w x}{G}\right)^2 - \beta^2}}{\sqrt{\left(\frac{\pi s_w x}{G}\right)^2 - \beta^2}} \quad , \tag{4.12}$$

where I_0 is the zero-th modified Bessel function, s_w is the size of the convolving kernel and β is the tapering parameter, that determines how fast the KB drops to zero. Working with $\alpha < 2$ yields a substantial decrease in run time and memory required for the reconstruction compared to the case of the GM with standard oversampling. The shape optimization of the KB kernel as a function of the oversampling gives the following results [168]:

$$\beta = \pi \sqrt{\left(\frac{S_W}{\alpha}\right)^2 \cdot \left(\alpha - \frac{1}{2}\right)^2 - 0.8} \quad , \tag{4.13}$$

$$S = \frac{0.91}{\alpha \cdot \epsilon^{(s)}_{max}} \text{ for NN} \qquad S = \frac{1}{\alpha}\sqrt{\frac{0.37}{\epsilon^{(s)}_{max}}} \text{ for LIN} , \tag{4.14}$$

where NN and LIN refer, respectively, to the nearest neighbor and linear interpolation scheme to sample the convolution LUT and $\epsilon^{(s)}_{max}$ is the maximum allowed sampling error (4.6).

4.1.3 Gridding method for CT

This paragraph is partially taken from Section 3.1 of [29].
The connection between the GM and CT reconstruction is provided by the Fourier slice theorem (FST). The discretized version of the FST states that the fast 1D Fourier transform (FFT) of a projection line at angle θ yields equidistant samples of $\hat{f}$ along the line $u_2 = u_1 \tan\theta$. The samples collected from multiple projections form a polar lattice in Fourier space. To retrieve f through IFFT-2D, these polar samples need to be interpolated at Cartesian locations, an operation that can be performed by GM. CT reconstruction through a filtered gridding backprojector consists in the following steps [163, 166, 167, 169]:

1. FFT1 of the sinogram along the channel direction $\longleftarrow \hat{P}_{\{\theta\}}$;
2. ramp filtering of the sinogram $\longleftarrow \hat{P}^{(f)}_{\{\theta\}}$;
3. convolution with kernel $\longleftarrow \hat{f}^{(d)} = \hat{W} * \hat{P}^{(f)}_{\{\theta\}}$;
4. IFFT2 of $\hat{f}^{(d)} \longleftarrow f^{(d)}$;
5. deapodization $\longleftarrow f = f^{(d)}/w$.

When the GM is used for tomographic reconstruction, the compensation factors accounting for the uneven sampling density in Fourier space (DCFs) are simply provided by the ramp filter.

4.2 Gridding projectors for iterative reconstruction of absorption data

The following section represents a reprinted manuscript published as: F. Arcadu, M. Nilchian, A. Studer, M. Stampanoni, and F. Marone, *"A forward regridding method with minimal oversampling for accurate and efficient iterative tomographic algorithms"*, IEEE Transactions on Image Processing, 24(13), 14748-14764, 2016 [28].

4.2.1 Abstract

The reconstruction of underconstrained tomographic datasets still represents a major challenge. In fact, standard analytical techniques mostly give unsatisfactory results due to the lack of sufficient information. In the last decades, several iterative algorithms, which can easily integrate a priori knowledge, have been developed to tackle this problem. Most of them are based on some implementation of the Radon transform, that acts as forward projector. This operator and its adjoint, the backprojector, are typically called few times per iteration and represent the computational bottleneck of the reconstruction process. In this work, we present a Fourier-based forward projector, founded on the regridding method with minimal oversampling. We show that this implementation of the Radon transform significantly outperforms in efficiency other state-of-the-art operators with $O(N^2 \log_2 N)$ complexity. Despite its reduced computational cost, this method provides comparable accuracy to more sophisticated projectors and can therefore be exploited in iterative algorithms to substantially decrease the time required for the reconstruction of underconstrained tomographic datasets without loss in the quality of the results.

4.2.2 Introduction

The word *tomography* refers to the method of reconstructing virtual cross-sections of an object using its projections acquired at different angles. The filtered backprojection (FBP) algorithm, introduced in the 70's [176, 177], is an analytical method, that still represents the standard tomographic reconstruction technique. It offers, in fact, a good trade-off between computational efficiency and accuracy. FBP provides high quality reconstructions, when a sufficient amount of projections is acquired and the noise affecting the experimental data is low with respect to the object signal [12].

In many tomographic applications, one or both of these conditions are not met due to experimental constraints. This occurs, in particular, when the total scan time needs to stay under a certain threshold due to dose limitations or to the evolution speed of the sample.

Many iterative algorithms have been designed to outperform FBP in the reconstruction of underconstrained datasets. These methods can easily incorporate a-priori knowledge about the object or the image formation process. For this reason, they usually provide better reconstructions than FBP, but at a much higher computational cost. Most iterative reconstruction algorithms are based on a *forward projector* and its adjoint (the backprojector), typically called few times per iteration. The forward projector usually corresponds to a discrete implementation of the Radon transform on a specific basis. The efficiency of this operator strongly determines the performance of the entire iterative procedure. Standard tomographic forward projectors require $O(N^3)$ operations for an image of $N \times N$ pixels [93, 97, 98, 102, 178–182] and its computational cost remains substantial, even for highly optimized implementations on graphic processor units (GPUs) [108]. More recently, algorithms with $O(N^2 \log_2 N)$ complexity have been proposed to improve the computational efficiency. They are based on three different strategies. The first group of projectors exploits the Fourier slice theorem [12]. Samples in the Fourier space are directly interpolated from a Cartesian to a polar grid [183]. These methods are characterized by poor accuracy and were never considered for iterative applications. The second category uses a hierarchical decomposition of either the line integrals [184–189] or the image domain [109, 190, 191]. In particular, the backprojector of [109, 191] has shown great performance in terms of accuracy and speed. This latter work is, however, focused on analytical problems and an application to iterative algorithms has never been presented so far. The third class is based on the non-uniform fast Fourier transform with min-max interpolation [192]. This accurate and fast method has been exploited for iterative reconstruction techniques [170].

In this paper, we present a novel application of the *regridding method* (in a particular cost-efficient form) as tomographic forward projector with $O(N^2 \log_2 N)$ complexity. For the first time, we compare this method with state-of-the art implementations of the Radon transform. In this work, we do not consider only the forward projector itself, but also study its role in iterative procedures and, in particular, its influence on the quality of the results.

The regridding method, originally devised for image reconstruction in radio astronomy [155, 157], has been applied to a broad range of reconstructive imaging methods like magnetic resonance imaging (MRI) [154, 168, 193] and as backprojector in computer tomography (CT) [163, 166, 169]. This method is designed to retrieve a 2D or a 3D signal from samples of its Fourier transform located on a non-Cartesian lattice: the Fourier samples are, first, convolved with a window function onto a Cartesian grid, the IFFT is, then, applied and the signal in the real space is, finally, corrected from the windowing contribution. The speed and accuracy of the method depend entirely on the choice of the convolving window. There are, in particular, three milestones in the rather extensive literature about the regridding method: the prolate-spheroidal wavefunctions (PSWFs) and the kaiser-bessels (KBs) are recognized as high performing convolving windows [163]; an oversampling factor of 2, defined as ratio between the number of Fourier samples convolved onto the Cartesian grid and the number of pixels to be retrieved in the image space, is necessary for accurate reconstructions [158]; an efficient KB-based regridding method for MRI designed to work with oversampling < 2 is introduced in [168]. In this work, we integrate these three previous results and propose an efficient forward projector tailored especially for iterative reconstruction algorithms. The original contribution of this manuscript can be summarized as follows.

- Novel application of the regridding method using PSWF and KB kernels as forward projector in iterative algorithms. In CT, regridding was used exclusively as backprojector for analytical reconstruction.

- Utilization of the minimal oversampling strategy (so far used only in the context of MRI) for such forward projector to further improve its computational efficiency.

- Systematic comparison with state-of-the-art implementations of the Radon transform, clearly demonstrating the advantages of the proposed projector for iterative reconstruction.

This paper is structured as follows. In Section II, we present the formulation and implementation details of the proposed forward regridding projector (FRP) with standard and minimal oversampling. In Section III, we benchmark the efficiency and accuracy of the FRP against state-of-the-art space- and Fourier-based forward operators. Section IV deals with the reconstruction of underconstrained simulated and experimental tomographic datasets with the alternate direction method of multipliers (ADMM) implemented with the FRP and the most accurate among the projectors used in Section III. We conclude the paper in Section V with a summary and some final remarks.

4.2.3 Proposed method

Formulation

In this paper, we focus on the tomographic reconstruction of a single 2D slice from line projections of an object in parallel beam geometry. However, Fourier methods have been previously applied also to fan-beam [164, 171] and cone-beam [165] geometry and the method proposed here can analogously be extended beyond this simple 2D case.

The set of line projections for one tomographic slice is called sinogram. The object to be reconstructed is a finite integrable real function $f = f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ with bounded support. The symbol ˆ corresponds to the Fourier transform; $P_\theta(t)$ is the Radon transform of f for an angle θ and a distance t from the origin, as defined in [12]. We also use round and square brackets when referring respectively to continuous and discrete functions.

We start from the Fourier slice theorem [12]:

$$\hat{P}_\theta(\omega) = \hat{f}(\omega \cos\theta, \omega \sin\theta) \quad . \tag{4.15}$$

Thanks to (4.15), the problem of reconstructing f from its projections $P_\theta(t)$ would be immediately solved, if the samples $\hat{f}[\omega_i \cos\theta_j, \omega_i \sin\theta_j]$ were located on a Cartesian grid. This is, however, not the case and the samples on the Cartesian grid need to be estimated from those available on a polar grid. Simple interpolation schemes (nearest neighbor, linear or bilinear interpolation), used for this approximation, lead to inaccurate reconstructions, since the localized interpolation error in Fourier space smears back onto the entire image, once the IFFT is computed.

In the regridding method (RM) [154, 158], the interpolation from the $[\omega_i, \theta_j]$ to the $[u_m, v_n]$ grid is, instead, performed as a convolution with a smooth window function $\hat{h}$, yielding more accurate results, depending on the chosen kernel. The standard RM for analytical CT reconstruction consists of five steps [163, 166, 169]:

1. FFT-1D of each projection $\longleftarrow \hat{P}_\theta$;
2. ramp filtering of the projections $\longleftarrow \hat{P}_\theta^{(f)}$;
3. convolution with window function $\longleftarrow \hat{f}^{(m)} = \hat{h} * \hat{P}_\theta^{(f)}$;
4. IFFT-2D of $\hat{f}^{(m)} \longleftarrow f^{(m)}$;
5. removal of the window function $\longleftarrow f = f^{(m)}/h$.

The last step, known as *deapodization*, is necessary, because computing $\hat{h} * \hat{f}$ in Fourier space is equivalent to calculating $h \cdot f$ in real space, so that removal of the convolving function is required if f is sought. Accordingly, the forward regridding projector (FRP) is defined by the same steps listed above (ramp filtering excluded) in the reverse order:

1. pre-deapodization $\longleftarrow f^{(m)} = f/h$;
2. FFT-2D of $f^{(m)} \longleftarrow \hat{f}^{(m)}$;
3. convolution with window function $\longrightarrow \hat{P}_\theta = \hat{h} * \hat{f}^{(m)}$;

4. IFFT-1D of each polar slice of $\hat{P}_\theta \longleftarrow P_\theta$.

The FRP does not require density compensation factors, because, differently from the regridding method used for backprojection, the input samples are uniformly distributed in Fourier space. To simplify the algorithm, the convolving function $\hat{h} = \hat{h}(u,v)$ is chosen to be separable in u, v, implying that also $h = h(x,y)$ will be separable in x, y:

$$\hat{h} = \hat{h}(u,v) = \hat{\psi}(u) \cdot \hat{\psi}(v) \implies h(x,y) = \psi(x) \cdot \psi(y) \tag{4.16}$$

Indicating with $\mathcal{F}_a$ the (discrete) Fourier transform with respect to the variable a and with $\{\cdots\}$ a collection of discrete values, the formula for the FRP results:

$$\begin{aligned} P_{\theta_i}[t_l] &= \mathcal{F}_\omega^{-1}\left\{\hat{P}_{\theta_i}[\omega_j]\right\}_l = \mathcal{F}_\omega^{-1}\left\{\left(\hat{h} * \hat{f}^{(m)}\right)[\omega_j, \theta_i]\right\}_l = \\ &= \mathcal{F}_\omega^{-1}\left\{\sum_{\substack{u_n \in U_{ij} \\ v_m \in V_{ij}}} \underbrace{\hat{\psi}[g_1(\omega_{j,u} - u_n)]\hat{\psi}[g_1(\omega_{j,v} - v_m)]}_{C_{i,j,n,m}} \hat{f}^{(m)}[u_n, v_m]\right\}_l \\ &= \mathcal{F}_\omega^{-1}\left\{\sum_{\substack{u_n \in U_{ij} \\ v_m \in V_{ij}}} C_{i,j,n,m}\ \mathcal{F}_{x,y}\left\{\frac{f[x_p, y_q]}{\psi[g_2 x_p]\psi[g_2 y_q]}\right\}_{u_n, v_m}\right\}_l \end{aligned} \tag{4.17}$$

where $\omega_{j,u} = \omega_j \cos\theta_i$, $\omega_{j,v} = \omega_j \sin\theta_i$, U_{ij} and V_{ij} are respectively the sets of values of u_n and v_m inside the interpolation support, g_1 and g_2 are factors to map distances in the polar and Cartesian grid into look-up tables (LUTs) indices.

The accuracy and efficiency of the RM (and, therefore, of the FRP) lie entirely in the choice of the convolving function $\hat{h}$. In terms of accuracy, the optimal kernel is an infinite sinc function [163], which is, however, computationally impractical. For finite kernels and, therefore, superior efficiency, the smaller the supports of h and $\hat{h}$ are, the better the performance of the method [163]. If h has compact support in real space, the interpolation error is still smeared out on the whole image, but it can be expected to remain as local as possible and the reconstruction will be accurate [158]; if $\hat{h}$ has compact support in Fourier space, only few datapoints in the support region will contribute to the convolution and the method will be computationally efficient. The accuracy of a convolution kernel $\hat{h}$ is strongly determined by the shape (in particular, the rolloff) of the central lobe and the amplitude of the aliasing sidelobes characterizing h. The deapodization removes the rolloff induced by this central lobe, especially at the borders of the output image, but, at the same time, it amplifies the aliasing sidelobes [158]. Aliasing contamination from sidelobes can be reduced by *oversampling* [158, 163], i.e. artificially extending the image field-of-view (FOV) and ignoring the outer portion of the image after reconstruction. In this way, smaller sidelobes will alias back into a wider central lobe, whose tails are neglected, once the output array is cropped. The extended central lobe also tapers less steeply, so that less rolloff correction is required. For a more detailed discussion on the optimal characteristics of convolution kernels, the reader is referred to the abovementioned references.

An oversampling $\alpha = G/N = 2.0$, with G the size of the grid, where the interpolation takes place, and N the size of the original input grid, guarantees accurate reconstructions with the RM [158, 163]. The aliasing error is a metric to compare the performance of different convolving functions and has the following pixel-based form [158, 168]:

$$\epsilon(i) = \sqrt{\frac{\sum\limits_{p \neq 0} [h(i+Gp)]^2}{[h(i)]^2}} \quad . \tag{4.18}$$

FRP with prolate spheroidal wavefunctions

As outlined above, the *energy* of the ideal convolution kernel should be maximally concentrated in a compact support both in real and Fourier space. We introduce the operators P_T and P_Ω, that create respectively time- and band-limited functions:

$$\begin{aligned} P_T f(t) &= \Theta_T f(t) \qquad P_\Omega f(t) = \mathcal{F}_\omega^{-1}\left[\Theta_\Omega\, \mathcal{F}_t(f(t))\right] \\ \Theta_T(t) &= \begin{cases} 1 & \text{if } |t| \leqslant T/2 \\ 0 & \text{if } |t| > T/2 \end{cases} \qquad \Theta_\Omega(\omega) = \begin{cases} 1 & \text{if } |\omega| \leqslant \Omega/2 \\ 0 & \text{if } |\omega| > \Omega/2 \end{cases} \end{aligned} \tag{4.19}$$

The prolate spheroidal wavefunctions (PSWFs) are the eigenfunctions of the self-adjoint operator $P_\Omega P_T$ [173]:

$$P_\Omega P_T \psi_i(t) = \int_{-T/2}^{T/2} dt' \, \frac{\sin(\Omega(t-t'))}{\pi(t-t')} \psi_i(t') = \lambda_i \psi_i(t) \quad . \tag{4.20}$$

If ψ_i's are bandlimited, the PSWF of 0-th order, ψ_0, with the largest eigenvalue λ_0, has the highest amount of energy $E_0 = \|P_T\psi_0\|^2$ in the interval $[-T/2, T/2]$ [173]. From the eigenvalue equation (4.20), no analytical expression for the PSWFs can be obtained; however, several approximations exist, depending on the magnitude of the parameter $c = T\Omega/2$ [173]. Since T should not be larger than ± 1 to guarantee an accurate deapodization correction [169], this window function is best approximated using spherical Bessel functions, expressed as a normalized linear combination of even Legendre polynomials in Fourier space [173]:

$$\hat{\psi}_0^{(a)}(k) = \frac{\sum\limits_{n=0}^{N_a} c_{2n} L_{2n}(k)}{\sum\limits_{n=0}^{N_a} c_{2n} L_{2n}(0)} \quad , \tag{4.21}$$

where the superscript $^{(a)}$ stands for "approximated", L_n is the Legendre polynomial of degree n and $\{c_{2n}\}$ are the expansion coefficients. The inverse Fourier transform of a series of even Legendre polynomials corresponds to a series of even spherical Bessel functions, that can themselves be expanded in the interval $[-1, 1]$ in terms of even Legendre polynomials [174],[175]. According to this argument, $\psi_0^{(a)}(x)$ represents a scaled version of $\hat{\psi}_0^{(a)}(k)$:

$$\psi_0^{(a)}(x) = C \cdot \hat{\psi}_0^{(a)}\left(x\frac{S}{G}\right) \quad , \tag{4.22}$$

where S is the density of the convolution LUT and $C \in \mathbb{R}_0$ is a constant.
An analytical regridding algorithm, based on this approximation of the PSWFs of 0-th order, has shown to be about 20 times faster than FBP, while keeping the same image quality [169]. The FRP has been implemented analogously with separable convolving kernel and deapodizer given respectively by (4.21) and (4.22). The additional parameters were set to: $\alpha = 2.0$, $N_a = 10$ and $S = 2048$. This implementation is abbreviated with FRP-PSWF. Since $\alpha = 2$ guarantees accurate results [158], we consider this oversampling as “standard”.

FRP with minimal oversampling

Due to the lack of an analytical expression for the PSWFs, it is difficult, in this case, to optimize the oversampling and other parameters as a function of the aliasing error (4.18). Kaiser Bessels (KBs) kernels have shown great accuracy compared to other convolving windows like two- and three-term cosine, Gaussian and B-splines [158], although their aliasing error is slightly higher than for the PSWFs approximation in (4.21). The simple analytical expression of KBs kernels enables, however, an optimization of their shape as a function of the oversampling [168]. The formula for the KB window and its inverse Fourier transform are [101]:

$$\hat{h}(k_x) = \frac{G}{W} I_0\left(\beta\sqrt{1-\left(\frac{2Gk_x}{W}\right)^2}\right) \quad \text{for } |k_x| \leqslant \frac{W}{2G} \ , \tag{4.23}$$

$$h(x) = \frac{\sin\sqrt{\left(\frac{\pi W x}{G}\right)^2 - \beta^2}}{\sqrt{\left(\frac{\pi W x}{G}\right)^2 - \beta^2}} \quad , \tag{4.24}$$

where I_0 is the zero-th modified Bessel function, W is the size of the convolving kernel and β is the tapering parameter, that determines how fast the KB drops to zero. Working with $\alpha < 2$ yields a substantial decrease in run time and memory required for the reconstruction compared to the case of the RM with standard oversampling. The shape optimization of the KB kernel as a function of the oversampling gives the following results [168]:

$$\beta = \pi\sqrt{\left(\frac{W}{\alpha}\right)^2 \cdot \left(\alpha - \frac{1}{2}\right)^2 - 0.8} \quad , \tag{4.25}$$

$$S = \frac{0.91}{\alpha \cdot \epsilon_{max}} \ \text{for NN} \qquad S = \frac{1}{\alpha}\sqrt{\frac{0.37}{\epsilon_{max}}} \ \text{for LIN} \ , \tag{4.26}$$

where NN and LIN refer, respectively, to the nearest neighbor and linear interpolation scheme to sample the convolution LUT and ϵ_{max} is the maximum allowed aliasing error (4.18). The FRP with minimal oversampling has been implemented with a KB kernel based on (4.25) and (4.26), setting $\epsilon_{max} = 0.01$, $\alpha = 1.125$ and choosing a linear interpolation scheme for the LUT. The FRP with this setting of parameters, abbreviated as FRP-KB, computes artifact-free forward projections in the shortest time and with the smallest memory allocation.

Further details

The kernel width is set to $W = 14.0/\pi$ for both FRP-PSWF and FRP-KB. Experiments with the filtered adjoint operators, conducted on the Shepp-Logan (SL) phantom [129], show that for $W \in [18/\pi, 20/\pi]$ the reconstruction is affected by a "blackening" artifact at the sides of the image, because the intepolation is not local anymore, whereas for $W \in [4/\pi, 10/\pi]$ the interpolation support is too small and the reconstruction is consequently affected by bright symmetric curves. Reconstructions look reasonable, by visual inspection, for $W \in (10/\pi, 18/\pi)$; in our implementation, $W = 14.0/\pi$ provides the most accurate results.

Moreover, we recall that the backprojector is the adjoint (not the inverse!) operator of a forward projector. From a computational point of view, the backprojector performs all operations of the forward projector, but in reverse order. Density compensation factors (DCFs) are not needed for the backprojector, when coupled to the forward operator inside an iterative scheme. The DCFs come into play only when the backprojector is used for analytical tomographic reconstruction. In such case, they correspond to a discrete version of the ramp filter, analogously to FBP.

The FRP operators are implemented in C with the fast FFTW library [194] and the pseudocode is shown in Appendix. The 1D functions $\hat{\psi}$ and ψ are stored in precomputed LUTs to allow for fast computations.

Algorithm complexity

The cost of the FRP lies in the convolution and the call of FFT/IFFT. Given an input image of $N \times N$ pixels, an oversampling ratio α, a kernel width W and a number of views M, the convolution amounts to approximately WMN operations, whereas the overall call of FFT/IFFT corresponds to $\alpha^2 N^2 \log_2(\alpha N) + \alpha M N \log_2(\alpha N)$ (1-time FFT-2D and M-times IFFT-1D) floating operations, that represents the leading cost term when N is big enough.

4.2.4 Accuracy assessment

Benchmark procedure

The accuracy of the proposed forward projectors has been assessed with a phantom for which the analytical Radon transform can be computed. The SL phantom [129] consists of 10 roto-translated ellipses and, is, therefore, suited for this task [12]. We use here a modified version of the original SL featuring higher contrast to improve visual perception.

The FRP-PSWF and FRP-KB are benchmarked against state-of-the-art implementations of the Radon transform, listed in the following.

The ray-driven (RT-RAY) [181] and distance-driven (RT-DIST) [98] forward projectors are widely used and are characterized by $\mathrm{O}(\mathrm{N}^3)$ complexity.
The Radon transform based on a cubic B-spline basis (RT-BSPLINE) [102] is a very accurate projector, superior, for instance, to implementations on pixel [93] and Kaiser-Bessel basis [182]. It is characterized by $\mathrm{O}(\mathrm{N}^3)$ complexity. For its implementation, a LUT of projected B-spline tensor products is precomputed for each view ($S = 2048$) and used inside a standard pixel-based forward projector.
The hierarchical filtered backprojection [109] treats the reconstruction of an $\mathrm{N} \times \mathrm{N}$ image as the sum of shifted reconstructions of smaller images centered at the origin. While the image grid is recursively halved, the input sinogram is halved in angles and pixels, since each sub-image requires fewer projections for the reconstruction. Thanks to this decomposition scheme, the backprojection is performed with $\mathrm{O}(\mathrm{N}^2 \log_2 \mathrm{N})$ operations. We used the implementation described in [109, 191] to build the corresponding forward projector (FHFP). The parameters ruling the image/sinogram decomposition and the choice of the interpolation schemes can strongly influence the accuracy and efficiency of this projector. Using the performance study in [191] as guide, we selected the best parameters for accurate results, when studying the operator accuracy. For the computational cost evaluation, parameters providing fastest artifact-free (at visual inspection) forward/backprojection have instead been used.
The non-uniform fast Fourier transform allows to Fourier transform finite-length signals to a non-uniformly sampled frequency space. Similarly to the regridding method, it can be used to reconstruct a signal from a non-Cartesian collection of samples in Fourier space. Several versions exist ([195–198]): the non-uniform fast Fourier transform using min-max interpolation represents, so far, the most accurate implementation of this method [192]. The forward projector based on this technique (NUFFT) requires $\mathrm{O}(\mathrm{N}^2 \log_2 \mathrm{N})$ operations, but also a significant amount of additional calculations and memory to precompute and store shift-invariant interpolators. For accurate results, NUFFT necessitates an oversampling of 2.

Analysis tools

The accuracy of sinograms and reconstructed images obtained with different forward projectors is assessed using the following standard metrics: root mean square error (RMSE), peak-signal-to-noise-ratio (PSNR) [144] and mean structural similarity index (MSSIM) [145]. For the accuracy assessment of the projectors in iterative schemes (Section 4.2.5), a linearly regressed version $\hat{\mathrm{I}}_{\mathrm{regr}}$ of the reconstructed slice (I) is used for the metrics computation. In this way, the obtained score is less biased by the different gray scales, characterizing I and the reference image, O, guaranteeing fair comparisons. Reconstruction methods exploiting total variation can, in fact, lead to slight shrinkage of the sparse coefficients [199]. $\hat{\mathrm{I}}_{\mathrm{regr}}$ is computed in the following way:

$$\hat{\mathrm{I}}_{\mathrm{regr}} = \underset{\mathrm{I}_{\mathrm{regr}}}{\operatorname{argmin}} \|\mathrm{I}_{\mathrm{regr}} - \mathrm{O}\|_2^2 \quad \text{s.t.} \quad \mathrm{I}_{\mathrm{regr}} = a \cdot \mathrm{I} + b \tag{4.27}$$

where $a, b \in \mathbb{R}$.

Comparison of forward projectors

The accuracy assessment of the two proposed forward projectors (FRP-PSWF , FRP-KB) is assessed against the five operators introduced in 4.2.4 (RT-RAY, RT-DIST, FHFP, NUFFT, RT-BSPLINE).
The results of this section are focused on the forward projection and subsequent analytical reconstruction of a SL phantom with 512×512 pixels. These experiments were also performed on SL phantoms with other resolutions (128×128, 256×256, 1024×1024, 2048×2048) and different number of views. The results show that the observed trends are independent from the phantom size and the angular sampling.
Sinograms with 805 views × 512 pixels have been computed using the seven above-mentioned forward projectors and are compared with the analytical Radon transform of the SL phantom by means of the RMSE, PSNR and MSSIM values (Tab. 2). The figures of merit clearly show that the FRPs perform the worst, while RT-BSPLINE provides the most accurate sinogram. The other operators (NUFFT, RT-RAY, RT-DIST and FHFP) are characterized by an intermediate accuracy.
Each sinogram is, then, filtered with a standard ramp kernel in Fourier space and reconstructed using the corresponding backprojector. The chosen number of views $805 \simeq 512 * \pi/2$ guarantees that these analytical reconstructions are not affected by aliasing artifacts [12], ensuring that we are really quantifying the accuracy of the entire projection-reconstruction process for each operator. The results in Tab. 3 show metric scores very close to each other, contrary to the trend seen in Tab. 2. RT-BSPLINE confirms its highest accuracy, but NUFFT, RT-RAY and RT-DIST perform in a very similar way to the FRP operators. In this case, FHFP shows the lowest accuracy. This second experiment indicates that either the coupling forward-adjoint operator and/or the effect of a simple ramp filter may reduce or even cancel the accuracy superiority of a well-performing projector with respect to others of inferior quality. Computed sinograms and reconstructed slices are not shown because differences are not detectable by visual inspection.
Tab. 4, presents the computational efficiency of the different projectors. It reports the time needed and memory allocated to create a SL sinogram with 2048 pixels and 800, 1600 and 3200 views. The used hardware is an Intel(R) Core(TM) i7-3520M CPU 2.90GHz. FRP-KB is, on average, faster by a factor of 4.4, 11.0, 21.3 and 928 compared to FRP-PSWF, NUFFT, FHFP and RT-BSPLINE. The computational cost for all projectors, except RT-BSPLINE, scales similarly with the number of views in particular for 800 and 1600 projections. Among the Fourier-based methods, FRP-KB allocates, on average, 64% and 70% less memory than FRP-PSWF and NUFFT, respectively. RT-BSPLINE requires, instead, more memory than a standard pixel-based forward projector because the precomputed LUT has to be finely sampled to guarantee high accuracy. RT-RAY and RT-DIST are not listed in these tables, since the implementations used for this work are based on GPUs. However, extrapolations suggest that a single core implementation of RT-DIST could require approximately the same memory, but larger computational cost than FHFP. On the other hand, RT-RAY may require more memory than RT-DIST and longer times than RT-BSPLINE.

4.2.5 Application to iterative reconstruction algorithms

In this section, we investigate the usage of the proposed forward projection operator inside iterative algorithms. The study is aimed at experimentally assessing the connection between the accuracy of

	RMSE	PSNR	MSSIM
FRP-PSWF	4.14	43.99	0.82
FRP-KB	4.25	42.75	0.80
RT-RAY	3.34	44.84	0.90
RT-DIST	3.34	44.85	0.91
FHFP	3.51	44.42	0.88
NUFFT	3.28	45.01	0.90
RT-BSPLINE	1.24	53.44	0.89

Table 2: RMSE, PSNR and MSSIM values for SL sinograms with 805 views × 512 pixels generated by the algorithms listed in the first column. The reference sinogram is represented by the analytical Radon transform of SL.

	RMSE	PSNR	MSSIM
FRP-PSWF	0.09	26.00	0.24
FRP-KB	0.10	25.90	0.21
RT-RAY	0.10	25.80	0.23
RT-DIST	0.10	25.80	0.24
FHFP	0.11	24.95	0.13
NUFFT	0.11	25.45	0.24
RT-BSPLINE	0.09	27.42	0.37

Table 3: RMSE, PSNR and MSSIM values for the reconstructions of the SL sinograms with 805 views × 512 pixels generated by the algorithms listed in the first column. The recostructions were performed by means of the ramp-filtered adjoint operator corresponding to each sinogram. The reference image is the original SL phantom. The scores were computed inside the reconstruction circle.

the forward projector and the quality of the iterative reconstruction. Two completely different iterative methods are used in the following: the separable paraboloidal surrogate and the alternate direction method of multipliers.

	TIME (s)		
	800 views	1600 views	3200 views
FRP-PSWF	3.22	4.60	5.13
FRP-KB	0.68	0.94	1.47
FHFP	11.24	15.88	44.88
NUFFT	7.83	10.12	16.01
RT-BSPLINE	438.60	875.71	1775.04

Table 4: Run time (in s) required to generate a sinogram with 2048 pixels and 800, 1600 and 3200 views for the algorithms listed in the first column. The used hardware is an Intel(R) Core(TM) i7-3520M CPU 2.90GHz.

	MEMORY (MB)		
	800 views	1600 views	3200 views
FRP-PSWF	207.93	214.49	227.61
FRP-KB	70.28	76.84	89.96
FHFP	26.05	35.33	53.88
NUFFT	234.13	266.91	332.46
RT-BSPLINE	29.89	43.00	69.23

Table 5: Memory (in MB) required to generate a sinogram with 2048 pixels and respectively 800, 1600 and 3200 views for the algorithms listed in the first column.

Separable paraboloidal surrogate

The separable paraboloidal surrogate (SPS) [17] is an algorithm for penalized likelihood based on Poisson statistics, modeling the measurements at the detector. The SPS globally converges to a unique minimizer and easily allows to enforce non-negativity constraints and to exploit non-quadratic but convex penalty functions. SPS is implemented with the edge-preserving non quadratic cost introduced in [17] and [200]. For further details about this well-established iterative method for tomographic reconstruction, please refer to [17].

Alternate direction method of multipliers

The alternate direction method of multipliers (ADMM) [27] is an iterative scheme suitable to minimize a large variety of functionals, including those with L1-norm terms. Thus, the ADMM can be used to solve a tomographic reconstruction problem with total variation regularization:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \mathrm{TV}_\Omega(\mathbf{x}) \right] \quad . \tag{4.28}$$

$\mathbf{A}$ is the matrix representation of the forward projector, $\mathbf{b}$ is the sinogram, $\mathbf{x}$ is the unknown object and $\lambda \in \mathbb{R}$ determines the weight of the regularization; TV stands for total variation, defined as [128]:

$$\mathrm{TV}_\Omega(f) \; = \; \int_\Omega |\nabla f(\mathbf{x})| \, d\mathbf{x} \; = \; \|\nabla f(\mathbf{x})\|_1 \quad . \tag{4.29}$$

The problem (4.28) can be mapped into the minimization of the following augmented Lagrangian [103]:

$$\mathcal{L}_\mu(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \sum_{\mathbf{k}} \|\mathbf{u}_\mathbf{k}\|_1 + \boldsymbol{\alpha}^T(\mathbf{L}\mathbf{x} - \mathbf{u}) + \frac{\mu}{2}\|\mathbf{L}\mathbf{x} - \mathbf{u}\|_2^2 \tag{4.30}$$

where $\boldsymbol{\alpha}$ are the Lagrangian multipliers, $\mathbf{L}$ is the gradient operator and $\mathbf{u}$ is an auxiliary variable. The ADMM iteratively minimizes $\mathcal{L}_\mu(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha})$ by sequentially solving smaller problems; each iteration involves two sub-optimizations with respect to $\mathbf{x}$ and to $\mathbf{u}$, followed by the update of $\boldsymbol{\alpha}$:

1. $\mathbf{x}_{k+1} \longleftarrow \underset{\mathbf{x}}{\operatorname{argmin}} \, \mathcal{L}_\mu(\mathbf{x}, \mathbf{u}_k, \boldsymbol{\alpha}_k)$
2. $\mathbf{u}_{k+1} \longleftarrow \underset{\mathbf{u}}{\operatorname{argmin}} \, \mathcal{L}_\mu(\mathbf{x}_{k+1}, \mathbf{u}, \boldsymbol{\alpha}_k)$
3. $\boldsymbol{\alpha}_{k+1} \longleftarrow \boldsymbol{\alpha}_k + \mu(\mathbf{L}\mathbf{x}_{k+1} - \mathbf{u}_{k+1}) \quad .$

In step 1), the conjugate gradient (CG) method [201] is applied to the following linear system:

$$(\mathbf{A}^\dagger \mathbf{A} + \mu \mathbf{L}^\dagger \mathbf{L})\mathbf{x} \; = \; \mathbf{A}^\dagger \mathbf{b} + \mu \mathbf{L}^\dagger \left(\mathbf{u}_k - \frac{\boldsymbol{\alpha}_k}{\mu} \right) \quad . \tag{4.31}$$

The subproblem of step 2) is solved through a shrinkage operation:

$$\mathbf{u}_{k+1} = \max\left\{ \left| \mathbf{L}\mathbf{x}_{k+1} + \frac{\boldsymbol{\alpha}_k}{\mu} \right| - \frac{\lambda}{\mu}, 0 \right\} \operatorname{sgn}\left(\mathbf{L}\mathbf{x}_{k+1} + \frac{\boldsymbol{\alpha}_k}{\mu} \right) \quad . \tag{4.32}$$

Our implementation of the ADMM performs 4 of CG (4.31) for the $\mathbf{x}$-subproblem, whereas μ and λ are set to 1.0. This setting provides accurate reconstructions when dealing with the group of simulated and real data used in the following Section. Nevertheless, when performing experiments with different number of iterations for the CG and/or different values for μ and λ, neither trends regarding the accuracy nor the convergence of the ADMM implementations change.

Experiments

SPS and ADMM have been implemented with the proposed projectors (SPS / ADMM-PSWF, SPS / ADMM-KB) and with RT-BSPLINE (SPS / ADMM-BSPLINE), since this latter operator has shown the best accuracy in the experiments presented in 4.2.4.

The convergence of the iterative algorithms has been studied through the tomographic reconstruction of a SL sinogram with 200 views × 128 pixels. In this experiment, the number of iterations was 200. The cost function plots in Fig. 4.4, 4.5 show that, for both SPS and ADMM, the convergence is not significantly affected by the choice of the forward operator. In fact, the three implementations of each method show a slight mismatch only for the first 10 iterations (Fig. 4.4, 4.5). The following reconstructions of simulated and real data are computing by stopping the iterative algorithm when the L_2-norm of the relative difference between reconstructions of subsequent iterations is smaller than a threshold $\epsilon = 0.1$.

In the first experiment, an underconstrained SL sinogram with 50 views × 512 pixels has been used. Gaussian noise with $\sigma = 2.4\%$ of the sinogram mean has been added. This σ fairly reproduces the noise level affecting the real data used in the following experiments. The reconstructed slices in Fig. 4.6, 4.7 show no difference at visual inspection and the metric scores (Tab. 6, 7) confirm that the image quality is comparable for all SPS and ADMM implementations.

For the next two experiments, real data acquired at the TOMCAT beamline of the Swiss Light Source at the Paul Scherrer Institut have been used. In the first case, a sinogram with 1441 views × 2560 pixels of a modern seed acquired in full tomography (i.e., the sample support is entirely in the FOV) has been downsampled to 300 views to create an underconstrained version of the original dataset. Fig. 4.8, 4.9 show the FBP reconstruction with Parzen filter of the original sinogram (FBP-FULL), followed by either the SPS or the ADMM reconstructions of the underconstrained dataset. FBP-FULL is used as reference image to compute the scores in Tab.9. The iterative reconstructions show again no evident differences at visual inspection except a slight blurring in the SPS-BSPLINE and ADMM-BSPLINE case. This image quality degradation is also reflected in the higher RMSE and lower MSSIM values for the implementations with RT-BSPLINE (Tab. 8, 9).

In the second case, mouse lung tissue data acquired in interior tomography configuration (i.e. the sample support extends beyond the FOV) have been chosen. The original sinogram, consisting of 901 views × 2016 pixels, has been downsampled to 272 views. Metric scores are computed against FBP-FULL as in the previous case. Once again, the SPS and ADMM reconstructions present no visible difference and the image quality assessed through the figures of merit (Tab. 10, 11) lead to the same conclusion.

4.2.6 Discussion and conclusion

In this paper, we have studied a novel application of the regridding method as forward projector with $O(N^2 \log_2 N)$ complexity for CT reconstruction. More specifically, we have designed a forward regridding projector (FRP) with standard ($\alpha = 2.0$, FRP-PSWF) and minimal oversampling ($\alpha = 1.125$, FRP-KB).

The computational efficiency and accuracy of the proposed operators have been systematically com-

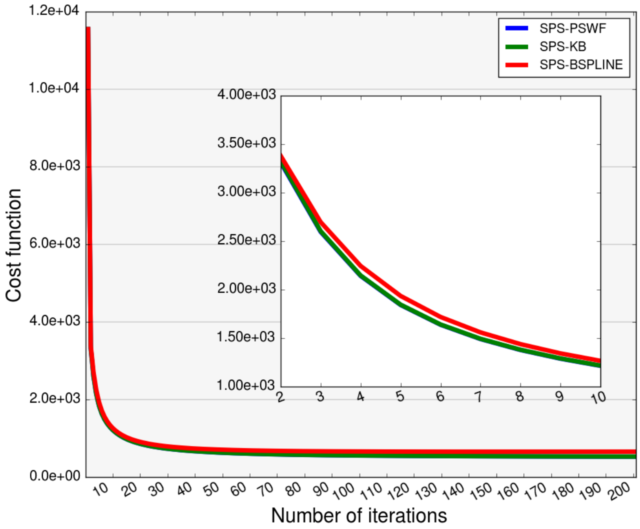


Figure 4.4: Cost function of the SPS against the number of iterations for the three SPS implementations. The plot corresponds to the reconstruction of a noiseless SL sinogram with 128 views × 200 pixels.

	RMSE	PSNR	MSSIM
SPS-PSWF	0.14	22.86	0.11
SPS-KB	0.14	22.86	0.11
SPS-BSPLINE	0.16	21.67	0.12

Table 6: RMSE, PSNR and MSSIM values for the SPS reconstructions of the SL sinogram shown in Fig4.6. The scores are computed inside the reconstruction circle.

pared with other state-of-the-art implementations of the Radon transform: a ray- and distance-driven projector (RT-RAY, RT-DIST), a Radon transform based on a cubic B-spline basis (RT-BSPLINE), a fast hierarchical forward projector (FHFP) and a Radon transform based on the non-uniform fast Fourier transform with min-max interpolation (NUFFT). Among the algorithms with $O(N^2 \log_2 N)$ complexity, FRP-KB is on average faster by a factor of 4, 11 and 21 with respect to FRP-PSWF, NUFFT and FHFP. Compared to pixel-based projectors like RT-BSPLINE, the speed increase achieved with FRP-KB is significantly larger (up to 3 orders of magnitude). FRP-KB also allows to reduce memory allocation by 64% and 70% compared to the Fourier-based projectors FRP-PSWF and NUFFT. The accuracy of the FRP operators is slightly inferior compared to the other implementations of the Radon transform, although differences in the computed sinograms are barely visible at visual in-

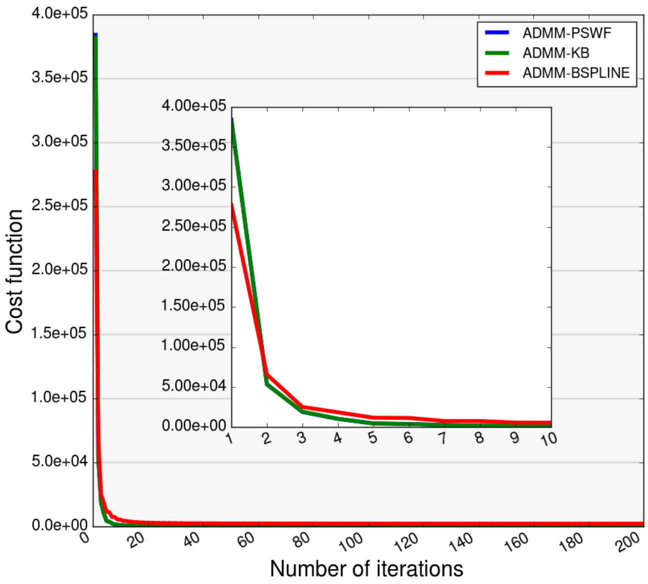


Figure 4.5: Cost function of the ADMM (4.28) against the number of iterations for the three ADMM implementations. The plot corresponds to the reconstruction of a noiseless SL sinogram with 128 views × 200 pixels.

	RMSE	PSNR	MSSIM
ADMM-PSWF	0.15	22.44	0.13
ADMM-KB	0.15	22.47	0.13
ADMM-BSPLINE	0.15	22.62	0.12

Table 7: RMSE, PSNR and MSSIM values for the ADMM reconstructions of the SL sinogram shown in Fig4.7. The scores are computed inside the reconstruction circle.

spection. We have shown that this lower accuracy becomes, however, negligible already when each sinogram is reconstructed with the related ramp-filtered backprojector, indicating that other aspects involved in the reconstruction process (e.g. filtering) have a stronger influence on the accuracy of the obtained results.

In a second step, the proposed forward regridding operators (FRP-PSWF and FRP-KB) have been integrated inside two different iterative schemes, (the separable paraboloidal surrogate, SPS, and the alternate direction method of multipliers, ADMM) to asses their performance for the reconstruction of simulated and real underconstrained tomographic datasets. We benchmarked these iterative algo-

	RMSE	PSNR	MSSIM
SPS-PSWF	0.06	31.33	0.14
SPS-KB	0.06	31.33	0.14
SPS-BSPLINE	0.07	30.56	0.10

Table 8: RMSE, PSNR and MSSIM values for the SPS reconstructions of the sinogram of the modern seed specimen shown in Fig4.8. The scores are computed inside the reconstruction circle.

	RMSE	PSNR	MSSIM
ADMM-PSWF	8.10	29.96	0.38
ADMM-KB	8.08	29.98	0.38
ADMM-BSPLINE	7.83	30.25	0.13

Table 9: RMSE, PSNR and MSSIM values for the ADMM reconstructions of the sinogram of the modern seed specimen shown in Fig4.9. The scores are computed inside the reconstruction circle.

	RMSE	PSNR	MSSIM
SPS-PSWF	4.23e-6	18.44	0.48
SPS-KB	4.23e-6	18.44	0.48
SPS-BSPLINE	4.87e-6	17.21	0.35

Table 10: RMSE, PSNR and MSSIM values for the SPS reconstructions of the sinogram of the mouse lung tissue shown in Fig4.10. The scores are computed inside the reconstruction circle.

	RMSE	PSNR	MSSIM
ADMM-PSWF	3.38e-6	20.40	0.61
ADMM-KB	3.47e-6	20.14	0.60
ADMM-BSPLINE	3.22e-6	20.81	0.61

Table 11: RMSE, PSNR and MSSIM values for the ADMM reconstructions of the sinogram of the mouse lung tissue shown in Fig4.11. The scores are computed inside the reconstruction circle.

rithms against the same implemented with RT-BSPLINE, the most accurate projector among those tested in this study. The results show that, despite the inferior accuracy of the fast operators based on the regridding method, both SPS/ADMM-PSWF and SPS/ADMM-KB have a similar convergence rate and provide comparable image quality as SPS/ADMM-BSPLINE. This result suggests that the minimization of the cost function in the tomographic reconstruction process is not strongly influenced by the chosen forward projector.

To conclude, we show that the proposed forward regridding projector with minimal oversampling (FRP-KB), thanks to its high computational efficiency, is an interesting new operator able to substantially speed up any iterative tomographic reconstruction algorithm, while preserving the results accuracy achieved with more sophisticated operators. Its low computational cost and reduced memory requirements make the proposed projector, so far developed for the parallel beam case, very appealing for addressing problems characterized by more complex geometries, usually very demanding especially from the memory point of view.

4.2.7 Acknowledgment

This study was supported by the Swiss National Science Foundation, Grant Nr. CR23I2-135550.

4.2.8 Appendix: C-like pseudocode of the FRP

The following pseudocode refers to an implementation of FRP, where the convolution LUT is sampled with nearest neighbor scheme. The notation is the same as introduced in section 4.2.3. In addition, $l = W/2$ is the half width of the convolving kernel. The subscript p stands for "padded". The parentheses $\lfloor \cdots \rfloor$ correspond to the floor operator, $\lceil \cdots \rceil$ to the ceil operator and $\lfloor \cdots \rceil$ to the round operator. Although the pseudocode refers to our C implementation of the FRP, there are a couple of functions (the slicing operator ":" and the memory allocating function "zeros"), that are characteristic of high-level programming languages like Matlab or Python, and are used to express some passages in a more concise form.

```
1. compute $\mathbf{I}_p \longleftarrow$ zeropad($\mathbf{I}$) up to G rows and columns

2. pre-removal of the interpolation kernel
   for $i = 0, \dots, G-1$ do:
     for $j = 0, \dots, G-1$ do:
       $\mathbf{I}_p^{(m)}[i,j] \longleftarrow \dfrac{\mathbf{I}_p[i,j]}{\psi[g_2 i]\psi[g_2 j]}$ ;
     endfor
   endfor

3. compute $\hat{\mathbf{I}}_p^{(m)} \longleftarrow$ FFT2$\left(\mathbf{I}_p^{(m)}\right)$ ;

4. initialize $\hat{\mathbf{S}}_p$ = zeros(M, G);
```

```
5. do interpolation in the Fourier space:
   for $i = 0, \dots, M-1$ do:
     for $k = 0, \dots, G/2-1$ do:
       $k_x = k\cos(\theta[i]) + G/2$
       $k_y = k\sin(\theta[i]) + G/2$
       $a_{k_x} = \lceil k_x - l \rceil$ , $b_{k_x} = \lfloor k_x + l \rfloor$
       $a_{k_y} = \lceil k_y - l \rceil$ , $b_{k_y} = \lfloor k_y + l \rfloor$
       for $k1 = a_{k_x}, \dots, b_{k_x}$ do:
         $\hat{\psi}_x = \hat{\psi}\,[g_1 \cdot \lfloor |k_x - k_1| \rceil]$
         for $k2 = a_{k_y}, \dots, b_{k_y}$ do:
           $\hat{\psi}_y = \hat{\psi}\,[g_1 \cdot \lfloor |k_y - k_2| \rceil]$
           $\hat{\psi}_{x,y} = \hat{\psi}_x \cdot \hat{\psi}_y$
           $\hat{\mathbf{S}}_p[i,k] += \hat{\mathbf{I}}_p^{(m)}[k1,k2] \cdot \hat{\psi}_{x,y}$
           $\hat{\mathbf{S}}_p[i,G-k] += \hat{\mathbf{I}}_p^{(m)}[G-k1,G-k2] \cdot \hat{\psi}_{x,y}$
         end for on Cartesian index $k_2$
       end for on Cartesian index $k_1$
       $\hat{\mathbf{S}}_p[i,k] = \hat{\mathbf{S}}_p[i,k] \cdot e^{-\pi i k G}$
       $\hat{\mathbf{S}}_p[i,G-k] = \hat{\mathbf{S}}_p[i,G-k] \cdot e^{\pi i k G}$
     end for loop on radial variable $k$
     $\mathbf{S}_p[i,:] = \mathrm{IFFT1}(\hat{\mathbf{S}}_p[i,:])$
   end for loop on angular index $i$

6. crop padded sinogram: $\mathbf{S} = \mathbf{S}_p[:, i1 : i2]$
   where $i1 = \lfloor (G-N)/2 \rceil$   $i2 = i1 + N$

7. return $\mathbf{S}$ .
```

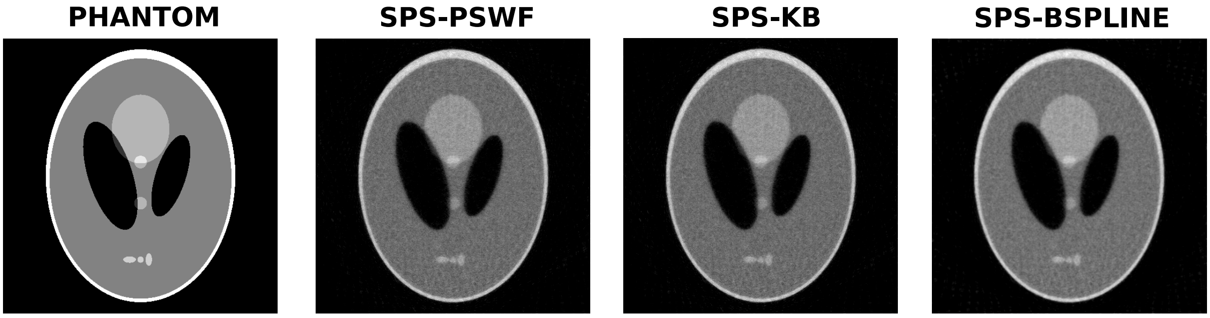


Figure 4.6: SPS reconstructions of a SL sinogram with 50 views × 512 pixels and additional Gaussian noise with $\sigma = 2.4\%$ of the sinogram mean.

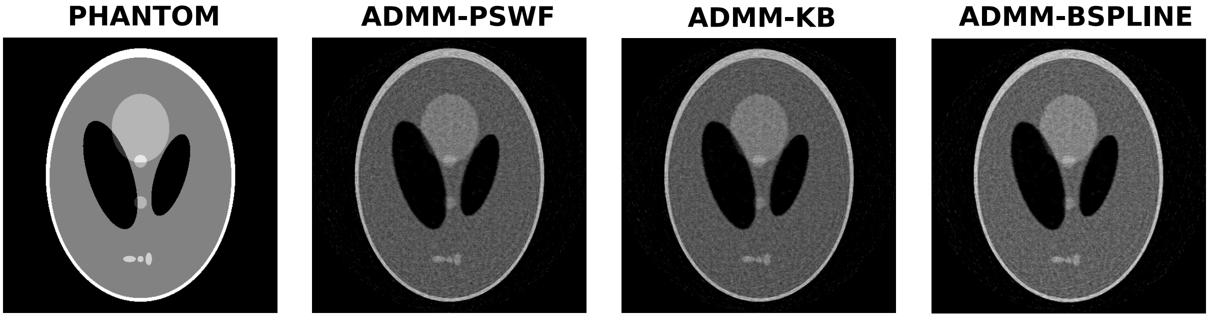


Figure 4.7: ADMM reconstructions of a SL sinogram with 50 views × 512 pixels and additional Gaussian noise with $\sigma = 2.4\%$ of the sinogram mean.

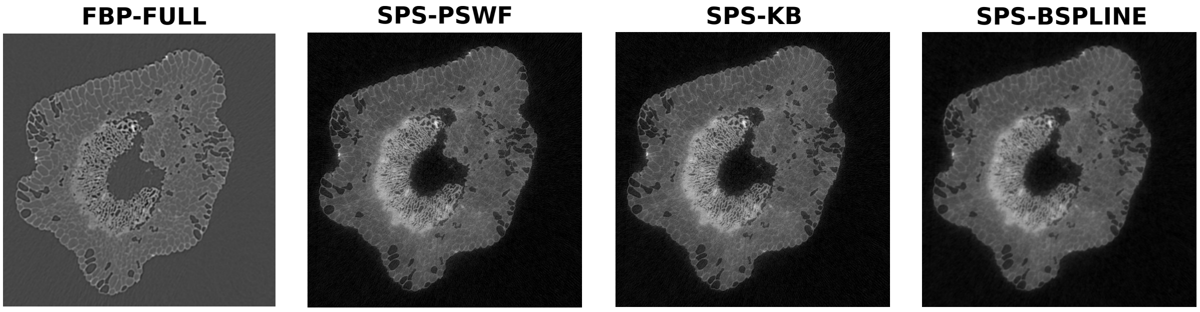


Figure 4.8: SPS reconstructions of the sinogram (300 views × 2560 pixels) of a modern seed specimen. FBP-FULL is the FBP reconstruction of the fully sampled sinogram with 1441 views × 2560 pixels Sample courtesy: S. Smith, University of Michigan.

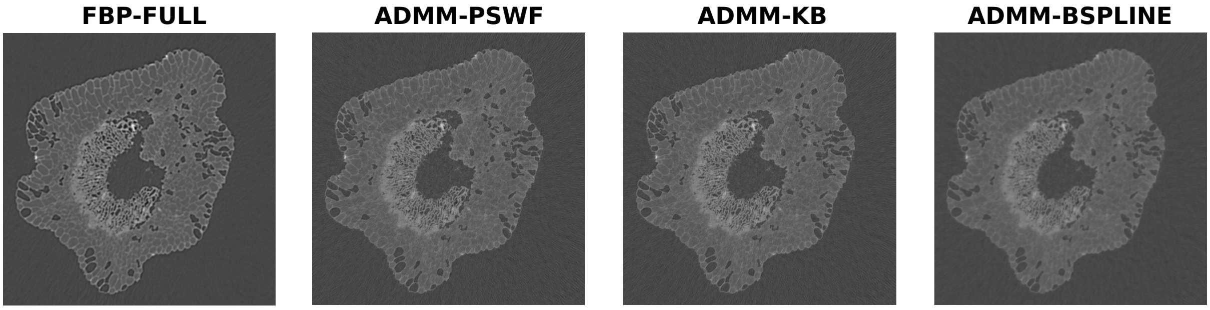


Figure 4.9: ADMM reconstructions of the sinogram (300 views × 2560 pixels) of a modern seed specimen. FBP-FULL is the FBP reconstruction of the fully sampled sinogram with 1441 views × 2560 pixels Sample courtesy: S. Smith, University of Michigan.

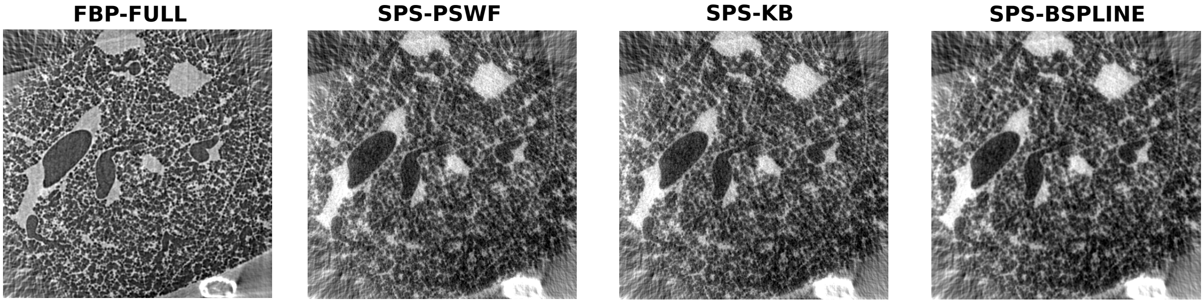


Figure 4.10: SPS reconstructions of the sinogram (272 views × 2016 pixels) of mouse lung tissue. FBP-FULL is the FBP reconstruction of the fully sampled sinogram with 901 views × 2016 pixels. Sample courtesy: G. Lovric [1].

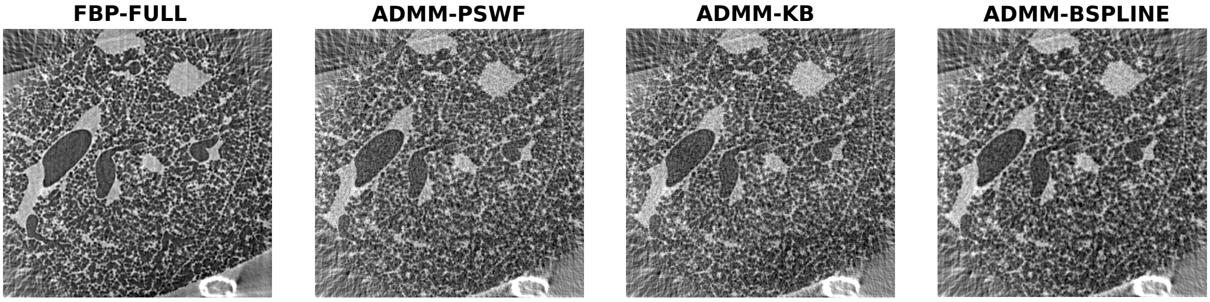


Figure 4.11: ADMM reconstructions of the sinogram (272 views × 2016 pixels) of mouse lung tissue. FBP-FULL is the FBP reconstruction of the fully sampled sinogram with 901 views × 2016 pixels. Sample courtesy: G. Lovric [1].

4.3 Gridding projectors for iterative reconstruction of DPC data

The following section represents a reprinted manuscript published as: F. Arcadu, M. Nilchian, A. Studer, M. Stampanoni, and F. Marone, *"Fast gridding projectors for analytical and iterative tomographic reconstruction of differential phase contrast data"*, Optics Express, 24(13), pp. 14749-64, 2016 [29].

4.3.1 Abstract

This paper introduces new gridding projectors designed to efficiently perform analytical and iterative tomographic reconstruction, when the forward model is represented by the derivative of the Radon transform. This inverse problem is tightly connected with an emerging X-ray tube- and synchrotron-based imaging technique: differential phase contrast based on a grating interferometer. This study shows, that the proposed projectors, compared to space-based implementations of the same operators, yield high quality analytical and iterative reconstructions, while improving the computational efficiency by few orders of magnitude.

4.3.2 Introduction

The *gridding method* is a technique designed to retrieve a 2D or 3D signal from samples of its Fourier transform located on a non-Cartesian lattice [154]. This method has been applied for decades to image reconstruction problems in radio-astronomy [155, 157], magnetic resonance imaging [154, 168, 193] and absorption computer tomography (CT) [28, 158, 163–166, 169, 170, 170, 192, 202].

A tomographic forward or backprojector based on the gridding method has $O(N^2 \log_2 N)$ complexity (N is the number of rows/columns of an image supposed to be square), leading to a significantly higher computational efficiency than a space-based implementation of the same operator ($O(N^3)$ complexity) [93]. Several studies have also shown that CT analytical [169, 170, 202] and iterative [28, 170] algorithms, adopting gridding projectors, can provide high quality reconstructions, while substantially reducing the calculation times (up to 3 orders of magnitude).

This work introduces a gridding forward and backprojector designed to perform analytical and iterative reconstruction of tomographic datasets, where the forward model is represented by the derivative of the Radon transform. These operators are, in particular, applied to the reconstruction of differential phase contrast (DPC) tomographic data acquired with a grating interferometer.

A grating interferometer is an imaging setup yielding three complementary types of information [79, 203, 204]: attenuation, differential phase and dark field signals. DPC focuses exclusively on the phase information related to the real part of the object refractive index and provides a high sensitivity to electron-density variations, down to $0.18\,e/\mathrm{nm}^3$ [205]. For this reason, DPC based on a grating interferometer is particularly suitable for the visualization of soft-tissue specimens.

DPC data can be analytically reconstructed by means of filtered backprojection without need for integration of the projection information, if an imaginary filter function corresponding to a Hilbert transform in the image space (HFBP) is used instead of the usual Ram-Lak filter [204]. This approach is however not suited for the reconstruction of DPC underconstrained data because of insufficient accuracy. Recent interest in the biomedical field for grating interferometry, when investigating dose sensitive specimens, has lead to several studies aimed at extending established CT iterative algorithms to the DPC case. In this regard, ad-hoc forward projectors working with a blob [206] and a cubic B-spline [102] basis and iterative schemes relying on different algorithms (the regularized maximum likelihood [206], the separable paraboloidal surrogate [207], the alternate direction method of multipliers [103] and a combination of statistical model and thresholding [208]) have been proposed.

So far, all analytical and iterative reconstruction methods developed for DPC utilize space-based projectors, characterized by a complexity of $O(N^3)$ and, therefore, strongly limiting their computational performance. The projectors proposed here greatly reduce the computational times required for analytical and, in particular, iterative reconstruction, while preserving the accuracy of the results.

Contributions

The contributions of this manuscript are summarized as follows:

- Design of a gridding forward projector acting as derivative of the Radon transform.
- Optimization of the parameters by means of the minimal oversampling strategy [168].

- Comparison of accuracy and efficiency with a standard space-based implementation of the same projector.
- Validation of the proposed operators through analytical and iterative reconstruction of simulated and experimental DPC data.

4.3.3 Notation and preliminaries

This work focuses on parallel beam geometry, even though gridding can be extended to fan- and cone-beam geometries [164, 165, 171].

The study object, i.e., an axial 2D slice of the whole 3D volume, is represented by a finite integrable real function $f = f(\mathbf{x}) = f(x, y) : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ with bounded support $S = \text{supp}(f) \equiv \{\mathbf{x} \in \mathbb{R}^2 \,|\, f(\mathbf{x}) \neq 0\}$. Its CT forward projection or X-ray transform or Radon transform (in 2D, these transforms coincide [92]) is given by:

$$P_\theta(t) := \mathcal{R}\{f(\mathbf{x})\}(\theta, t) := \int_\Omega d\mathbf{x}\, f(\mathbf{x})\, \delta(\mathbf{x} \cdot \mathbf{e}_\theta - t) \quad , \tag{4.33}$$

where $\cdot$ represents the inner product, $\theta \in [0, \pi)$ and $\mathbf{e}_\theta = (\cos\theta, \sin\theta)$. The variables used in (4.33) are displayed in Fig. 4.12(a).

The derivative of the Radon transform is considered with respect to the variable t and is given by:

$$\frac{d}{dt} P_\theta(t) := \mathcal{R}^{(1)}\{f(\mathbf{x})\}(\theta, t) := \frac{d}{dt} \int_\Omega d\mathbf{x}\, f(\mathbf{x})\, \delta(\mathbf{x} \cdot \mathbf{e}_\theta - t) \quad . \tag{4.34}$$

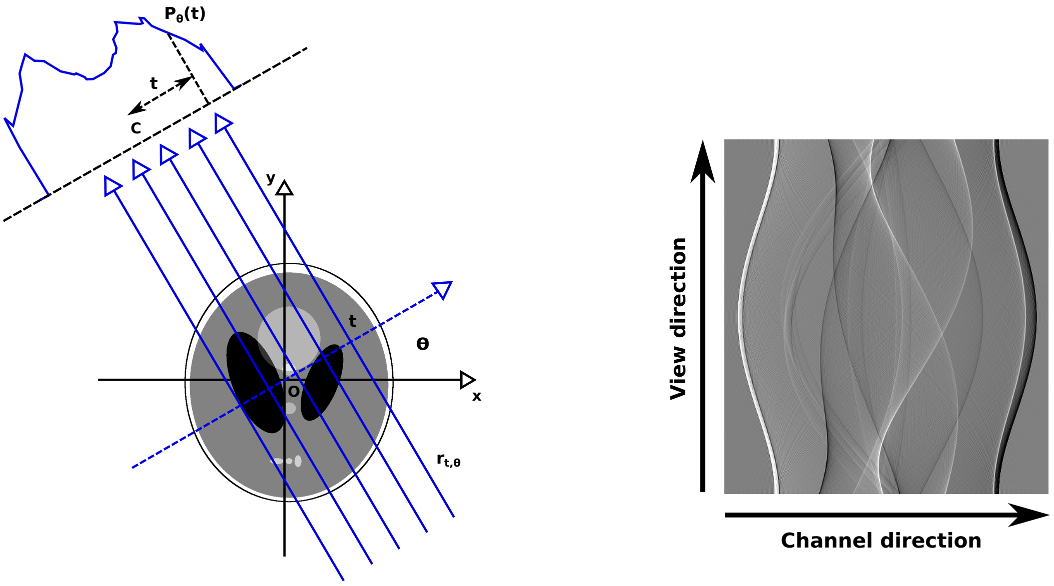
The collection of all projections ($P_{\{\theta\}}$ or $d\,P_{\{\theta\}}/dt$) in parallel beam geometry is also called sinogram. Figure 4.12(b) shows an example of DPC sinogram; the pixels of each projection are placed along the row or channel direction.

The 2D Fourier transform of a generic function $g(a, b)$ is indicated with $\hat{g}$ or $\mathcal{F}_{a,b}\{g\}$; $\mathcal{F}_a\{g\}$ or $\mathcal{F}_b\{g\}$ represents the 1D Fourier transform along the variable a or b, respectively. (x, y), (t, θ), (u, v) and (θ, ω) are the coordinates of the real-space Cartesian, real-space polar, Fourier Cartesian and Fourier polar reference frame, respectively. $(\cdots)$ refers to entries of a continuous function, $[\cdots]$ entries of a discrete function/array and $\{\cdots\}$ either entries of an operator or a collection of discrete values. Vectors and matrices are indicated with lower- and upper-case bold letters, respectively; both † and * correspond to the Hermitian adjoint operator.

4.3.4 Proposed method

Connection between gridding and CT

The gridding method (GM) retrieves a signal from samples of its Fourier transform located on a non Cartesian lattice: the samples are, first, convolved with a smooth and rapidly decaying window function, the inverse fast Fourier transform (IFFT) is computed and, finally, the contribution of the window function is removed from the signal in real space [154].



(a) **(b)**

Figure 4.12: (a) Coordinate system and variables used. (b) DPC sinogram example.

The connection between GM and CT reconstruction is provided by the Fourier slice theorem (FST) [92]:

$$\hat{P}_\theta(\omega) = \hat{f}(\omega\cos\theta, \omega\sin\theta) \quad . \tag{4.35}$$

The discretized version of (4.35) states that the fast Fourier transform (FFT) of a projection at angle θ yields equidistant samples of $\hat{f}$ along the line $v = u\tan\theta$. The samples collected from multiple projections form a polar lattice in Fourier space. To retrieve f through IFFT, these polar samples need to be interpolated at Cartesian locations, an operation that can be performed by the GM. CT reconstruction through a filtered gridding backprojector consists in the following steps [163, 166, 169]:

1. FFT1 of the sinogram along the channel direction $\longleftarrow \hat{P}_{\{\theta\}}$;
2. ramp filtering of the sinogram $\longleftarrow \hat{P}^{(f)}_{\{\theta\}}$;
3. convolution with window function $\longleftarrow \hat{f}^{(d)} = \hat{h} * \hat{P}^{(f)}_{\{\theta\}}$;
4. IFFT2 of $\hat{f}^{(d)} \longleftarrow f^{(d)}$;
5. removal of the window function $\longleftarrow f = f^{(d)}/h$.

Accordingly, the gridding-based implementation of $\mathcal{R}$ or forward gridding projector (FRP) is given by the same steps listed above (ramp filtering excluded) in the reverse order [28, 170]:

1. pre-correction $\longleftarrow f^{(d)} = f/h$;
2. FFT2 of $f^{(d)} \longleftarrow \hat{f}^{(d)}$;
3. convolution with window function $\longrightarrow \hat{P}_{\{\theta\}} = \hat{h} * \hat{f}^{(d)}$;

4. IFFT1 of each polar slice of $\hat{P}_{\{\theta\}} \longleftarrow P_{\{\theta\}}$.

Differentiated forward gridding projector

The formula of the differentiated forward gridding projector (DFRP) is obtained from the FST (4.35), the definitions (4.33), (4.34) and from the property that derivatives in real space are mapped into multiplications in Fourier space.

$$\begin{aligned}\mathcal{R}^{(1)}\{f\}(\theta,t) &= \frac{d}{dt}P_\theta(t) = \frac{d}{dt}\int_\Omega d\omega\, \hat{P}_\theta(\omega)\, e^{-2\pi i\omega t} \\ &= \int_\Omega d\omega\, e^{-2\pi i\omega t}\,(-2\pi i\omega)\cdot\left(\hat{h} * \hat{f}^{(d)}\right)(\theta,\omega) \\ &= \mathcal{F}_\omega^{-1}\left\{(-2\pi i\omega)\cdot\left(\hat{h} * \mathcal{F}_{x,y}\{(f/h)(x,y)\}\right)(\theta,\omega)\right\} \quad ,\end{aligned} \tag{4.36}$$

where i is the imaginary unit ($i^2 = -1$), h and $\hat{h}$ are the gridding interpolation kernel and its Fourier transform, $f^{(d)} = f/h$ is the pre-corrected and oversampled version of f. Choosing the kernel separable, i.e., $h(x,y) = \psi(x)\psi(y) \Longrightarrow \hat{h}(u,v) = \hat{\psi}(u)\hat{\psi}(v)$, simplifies the implementation of the method. Discretizing (4.36) results in:

$$\begin{aligned}\frac{d}{dt}P_{\theta_m}[t_n] &= \frac{d}{dt}\mathcal{F}_{\{\omega\}}^{-1}\left\{\hat{P}_{\theta_m}[\omega_j]\right\}_n = \frac{d}{dt}\mathcal{F}_{\{\omega\}}^{-1}\left\{\left(\hat{h} * \hat{f}^{(d)}\right)[\omega_j,\theta_m]\right\}_n \\ &= \frac{d}{dt}\mathcal{F}_{\{\omega\}}^{-1}\left\{\sum_{\substack{u_r\in U_{mj}\\ v_s\in V_{mj}}} \underbrace{\hat{\psi}[g_1(\omega_{j,u}-u_r)]\hat{\psi}[g_1(\omega_{j,v}-v_s)]}_{C_{m,j,r,s}}\, \hat{f}^{(d)}[u_r,v_s]\right\}_n \\ &= \mathcal{F}_{\{\omega\}}^{-1}\left\{(-2\pi i\omega_j)\sum_{\substack{u_r\in U_{mj}\\ v_s\in V_{mj}}} C_{m,j,r,s}\;\mathcal{F}_{\{x,y\}}\left\{\frac{f^{(o)}[x_p,y_q]}{\psi[g_2x_p]\psi[g_2y_q]}\right\}_{u_r,v_s}\right\}_n\end{aligned} \tag{4.37}$$

where i is the imaginary unit; $\theta_m = m\pi/M \in [0,\pi)$ for $m = 0,1,...,M-1$ and M is the number of views; the image is assumed to be square with N pixels (N is even) and to have unit resolution, therefore, $x_p = -N/2+p$ and $y_q = -N/2+q$ for $p,q = 0,1,...,N-1$ are the image pixel coordinates; $t_n = -N/2+n$ for $n = 0,1,...,N-1$ (N is also the number of detector cells); the Fourier space is sampled at G evenly spaced points in the range $[0.5, 0.5)$, therefore, $u_r = -1/2+r/G$ and $v_s = -1/2+s/G$ for $r,s = 0,1,...,G-1$, $G = \alpha N$ for $\alpha > 1$; α is called *oversampling ratio*; $\omega_{j,u} = \omega_j\cos\theta_m$, $\omega_{j,v} = \omega_j\sin\theta_m$ and $\omega_j = -1/2+j/G$ for $j = 0,1,...,G-1$; U_{mj} (V_{mj}) are the sets of pixel coordinates u_r (v_s) inside the interpolation support; g_1 and g_2 are factors mapping distances from the polar and Cartesian grid into look-up table (LUT) indices and depend on the sampling density of ψ and $\hat{\psi}$. $\mathcal{F}_\omega^{-1}$ and $\mathcal{F}_{\{x,y\}}$ in (4.37) are implemented as IFFT1 along ω and FFT2 along x,y, respectively. $f^{(o)}$ represents a zero-padded or oversampled version of f by a factor $\alpha > 1.0$. Gridding an oversampled grid prevents the appearance of aliasing artifacts due to tails of the reconstruction wrapping back into the field-of-view (FOV) [158, 169]. A general rule of thumb is to set $\alpha = 2.0$.

Since the DFRP is a linear operator, the differentiated gridding backprojector (DBRP) is the Hermitian adjoint operator of (4.37). From a computational point of view, the DBRP performs the same exact operations as the DFRP, but in reverse order.

The accuracy and efficiency of gridding operators depend on the choice of the finite kernel ψ: the smaller the support of ψ, the more accurate the pre-correction in real space; at the same time, the smaller the support of $\hat{\psi}$, the faster the interpolation stage in Fourier space [28, 158, 169]. The accuracy of a convolution kernel is also strongly determined by the shape (in particular, the rolloff) of the central lobe and the amplitude of the aliasing sidelobes characterizing ψ. The pre-correction removes the rolloff induced by this central lobe, but, in doing so, the aliasing sidelobes are amplified [28, 158].

In previous work [158, 163, 193], different compact kernels have been studied for gridding [158]. The conclusion is that a Kaiser-Bessel (KB) [101] kernel has an optimal compact support in real and Fourier space and provides analytical formulas, that enable shape optimization for ψ and $\hat{\psi}$ on the basis of a certain criterion, if available. The formulas for the KB kernel are:

$$\psi_{kb}(x) = \frac{\sin\sqrt{\left(\frac{\pi W x}{G}\right)^2 - \beta^2}}{\sqrt{\left(\frac{\pi W x}{G}\right)^2 - \beta^2}} \quad , \qquad \hat{\psi}_{kb}(k_x) = \frac{G}{W} I_0\left(\beta\sqrt{1 - \left(\frac{2Gk_x}{W}\right)^2}\right) \quad , \tag{4.38}$$

valid for $|k_x| \leqslant W/2G$ and where I_0 is the zero-th order modified Bessel function, W is the size of the convolving kernel and β is the tapering parameter, that determines how fast the KB drops to zero. The accuracy and computational performance of (4.37) with a KB kernel depends on α, W, β and S, the sampling density of $\hat{\psi}$. Working with a pre-sampled LUT enables faster calculations and this approach is, therefore, preferable than invoking Eq. (4.38), every time a point needs to be interpolated in Fourier space.

According to the minimum oversampling criterion, the KB's shape can be optimized by minimizing the maximum pixel aliasing error [168]:

$$\epsilon_{ij} = \sqrt{\frac{\sum_{p \neq 0} (h[i+Gp, j+Gp])^2}{(h[i,j])^2}} = \sqrt{\frac{\sum_{p \neq 0} (\psi[i+Gp]\psi[j+Gp])^2}{(\psi[i]\psi[j])^2}} \quad . \tag{4.39}$$

The optimal kernel ψ^{opt} is found as solution of the minimization problem:

$$\psi^{opt} = \underset{\psi}{\operatorname{argmin}} \max_{i,j} \left(\epsilon_{i,j}\right) \quad , \tag{4.40}$$

which translates in the following results [168]:

$$\beta = \pi\sqrt{\left(\frac{W}{\alpha}\right)^2 \cdot \left(\alpha - \frac{1}{2}\right)^2 - 0.8} \quad , \quad S = \begin{cases} 0.91/(\alpha \cdot \gamma) & \text{for NN} \\ \sqrt{0.37/(\gamma\alpha^2)} & \text{for LIN} \end{cases} \tag{4.41}$$

where NN and LIN refer, respectively, to the nearest neighbor and linear interpolation scheme to sample the convolution LUT and γ is the maximum allowed sampling error for $\hat{\psi}$. The minimal oversampling criterion (4.41) reduces the initial parameter space $\{W, S, \alpha, \beta\}$ to $\{W, \gamma, \alpha\}$.

Algorithm complexity

The cost of the DFRP lies in the convolution and the call of FFT/IFFT. Given an input image of $N \times N$ pixels, an oversampling ratio α, a kernel width W and a number of views M, the convolution amounts to approximately WMN operations, whereas the overall call of FFT/IFFT corresponds to $\alpha^2 N^2 \log_2(\alpha N) + \alpha M N \log_2(\alpha N)$ (1-time FFT-2D and M-times IFFT-1D) floating point operations. This second factor usually represents the leading cost term for real datasets.

4.3.5 Optimization of the gridding parameters

Optimization approach

For the pure Radon transform case, an optimal KB kernel, ψ^{opt}, can be constructed with any setting $\{W, S, \alpha, \beta\}$ fulfilling (4.41). The choice of one specific set among the optimal ones would then be dictated by computational efficiency criteria. However, the minimal oversampling criterion does not account for the influence of the term $-2\pi i \omega_j$ on the accuracy of the actual implementation of the DFRP. To assess which set of theoretically optimal gridding parameters provides the most accurate results for the DFRP, the 3D parameter space $\{W, \gamma, \alpha\}$ needs to be explored in a brute-force manner, through experiments making use of a family of functions, whose $\mathcal{R}^{(1)}$ can be computed analytically. The following family of 2D radially symmetric functions is considered:

$$f^{(n)}(\mathbf{x}) = \begin{cases} (a^2 - ||\mathbf{x}||^2)^n & \text{if} \quad ||\mathbf{x}|| \leqslant a \\ 0 & \text{if} \quad ||\mathbf{x}|| > a \end{cases} \qquad \forall\, n \in \mathbb{N} \quad . \tag{4.42}$$

The analytical differentiated Radon transform is given by:

$$\mathcal{R}^{(1)}\left\{f^{(n)}\right\}(\theta, t) = \begin{cases} -2t\dfrac{(2n)!!}{(2n-1)!!}\left(a^2 - t^2\right)^{n-1/2} & \text{if} \quad n = 2k \\ -\pi t\dfrac{(2n)!!}{(2n-1)!!}\left(a^2 - t^2\right)^{n-1/2} & \text{if} \quad n = 2k+1 \end{cases} \qquad k \in \mathbb{N} \ , \ \forall \theta \in [0, \pi). \tag{4.43}$$

To derive Eq. (4.43), one has to apply definition (4.34) to $f^{(n)}(\mathbf{x})$ and perform a cosinusoidal substitution to obtain an integral of the form $\int_0^\pi d\omega\, (\sin\omega)^m$, whose solution is available in [209].

Optimal gridding parameters are evaluated for analytical and iterative reconstruction methods. For the former case, the Hilbert-filtered DBRP is applied to DPC sinograms, created with (4.43), and the obtained reconstructions are compared with the original phantom (4.42). For the latter case, an approach disregarding the specific choice of iterative algorithm (and its specific parameters) is followed: the forward projection computed with DFRP is compared to the analytical sinogram and is, then, reconstructed with the Hilbert-filtered DBRP. In this way, both the accuracy of the standalone forward projector and of the coupling forward-adjoint operator can be investigated.

The experiments in Sec. 4.3.5 make use of $f^{(n)}(x)$ with $n = \{2, 3\}$, sampled on a grid with 512×512 pixels and its analytical $\mathcal{R}^{(1)}$ is computed for 805 views, homogeneously distributed in $[0, \pi)$. The number of views is enough to prevent undersampling artifacts in the reconstructions, since $805 \simeq 512 * \pi/2$ (according to the FBP Nyquist criterion [92]). Results are not biased by the resolution of the phantoms, because experiments repeated with different grid sizes have yielded the same outcome. The accuracy of the results is measured by the peak signal-to-noise ratio (PSNR) [144], a score quantifying the overall difference between the computed and reference image. The PSNR is defined as:

$$\mathrm{PSNR} = 10 \cdot \log_{10} \left[\frac{\max_{i,j} \left(O[i,j]\right)^2}{\frac{1}{MN} \cdot \sum_{i=0}^{M-1} \sum_{i=0}^{N-1} \left(O[i,j] - I[i,j]\right)^2} \right] \tag{4.44}$$

where O and I are 2D arrays with M rows and N columns, being, respectively, the oracle and the image to analyze. Since in (4.44) the difference is located at the denominator, PSNR results are more sensitive to small variations from the oracle than the mean square error. When analyzing sinograms, the PSNR is computed over the entire image, whereas, for reconstructions, it is computed within the resolution circle.

The mean structural similarity index (MSSIM) [145] has also been employed for the image analysis. Although confirming the trends described by the PSNR, the MSSIM results poorly sensitive for this kind of test. For this reason, only PSNR scores are reported in Sec.4.3.5.

Optimal gridding parameters

To limit the extent of the brute force search of $\{W, \gamma, \alpha\}_{\text{opt}}$, preliminary experiments, studying how the accuracy of the operators depends on just one parameter, while fixing the other two, have been performed. These experiments focus on the forward projection with DFRP and on the reconstruction of analytical sinograms with Hilbert-filtered DBRP. Results, collected in Figs. 4.13(a)-4.13(c), show the following key trends: the choice of α is more critical for the backprojector (Fig. 4.13(b)); the PSNR decreases only when $\gamma > 10^{-4}$ (Fig. 4.13(a)); the DFRP is highly influenced by the choice of W (Fig. 4.13(c)).

The general brute force search of the 3D parameter space has W ranging in [2.7, 17.7] (20 points), γ in $[10^{-8}, 10^{-3}]$ (10 points) and α in [1.05, 3.0] (20 points). Results can be visualized by means of 2D PNSR maps, displaying the accuracy as a function of two parameters while the third one is fixed, as shown in Figs. 4.14(a)-4.14(c). Few relevant trends, mirroring the results of the preliminary experiments, can be recognized. The accuracy of DFRP and DBRP is only weakly influenced by

γ, provided that this parameter is $< 10^{-4}$ (Figs. 4.14(a) and 4.14(b)). DFRP is very sensitive to the choice of W (Fig. 4.14(b)): the PSNR varies up to 5 dB for a difference of 0.8 in kernel size. The reconstruction accuracy of DFRP sinograms (coupling DFRP-DBRP) is sensitive to the choice of α (Fig. 4.14(a)) in a wider interval, $1.05 \leqslant \alpha \leqslant 2.30$, than for the reconstruction of analytical sinograms (DBRP standalone, Fig. 4.14(c)), $1.05 \leqslant \alpha \leqslant 1.70$.

The optimal parameters are found by locating the maximum PSNR values in the 3D volume, created by the brute force search. The PSNR reaches the maximum on a continuous spread area (the intersection of the darkest blue regions in Figs. 4.14(a)-4.14(c)), while no local extrema exist. Since the broad PSNR maximum corresponds to multiple parameter configurations, $\{\alpha, W, \gamma\}_{\text{opt}}$ is chosen as the peripheral point, requiring the smallest computational time. The optimal parameters for analytical reconstruction, as obtained from the top score for the reconstruction with Hilbert-filtered DBRP of analytical sinograms, are $\{W, \gamma, \alpha\}_{\text{opt}_1} = \{4.45,\ 1.7 \cdot 10^{-6},\ 1.75\}$. The optimal parameters for iterative reconstruction, merging the results for the computation of the forward projection with DFRP and the recostructions of these sinograms with Hilbert-filtered DBRP, are $\{W, \gamma, \alpha\}_{\text{opt}_2} = \{6.6,\ 6.0 \cdot 10^{-6},\ 2.38\}$.

The second set of optimal parameters is more conservative and less computationally efficient, because it involves a bigger kernel size and a higher oversampling ratio compared to the first set. The aliasing error introduced by the DFRP can be amplified by the DBRP at each iteration in iterative schemes, therefore a more conservative choice of parameters is mandatory. The results of the brute-force search

Table 12: (a) Reconstruction with Hilbert-filtered DBRP of an analytical SL sinogram (800 views × 512 pixels). (b) Reconstruction with Hilbert-filtered DBRP of a sinogram created by the DFRP (800 views × 512 pixels). The settings $\{W, \gamma, \alpha\}_{\text{nopt}_1}$ and $\{W, \gamma, \alpha\}_{\text{nopt}_2}$ belong to the PSNR maximum surface for analytical and iterative reconstruction, respectively. This is confirmed by the negligible differences in PSNR.

	$\{W, \gamma, \alpha\}_{\text{opt}_1}$	$\{W, \gamma, \alpha\}_{\text{nopt}_1}$
PSNR	24.51	24.14

(a)

	$\{W, \gamma, \alpha\}_{\text{opt}_2}$	$\{W, \gamma, \alpha\}_{\text{nopt}_2}$
PSNR	35.50	35.41

(b)

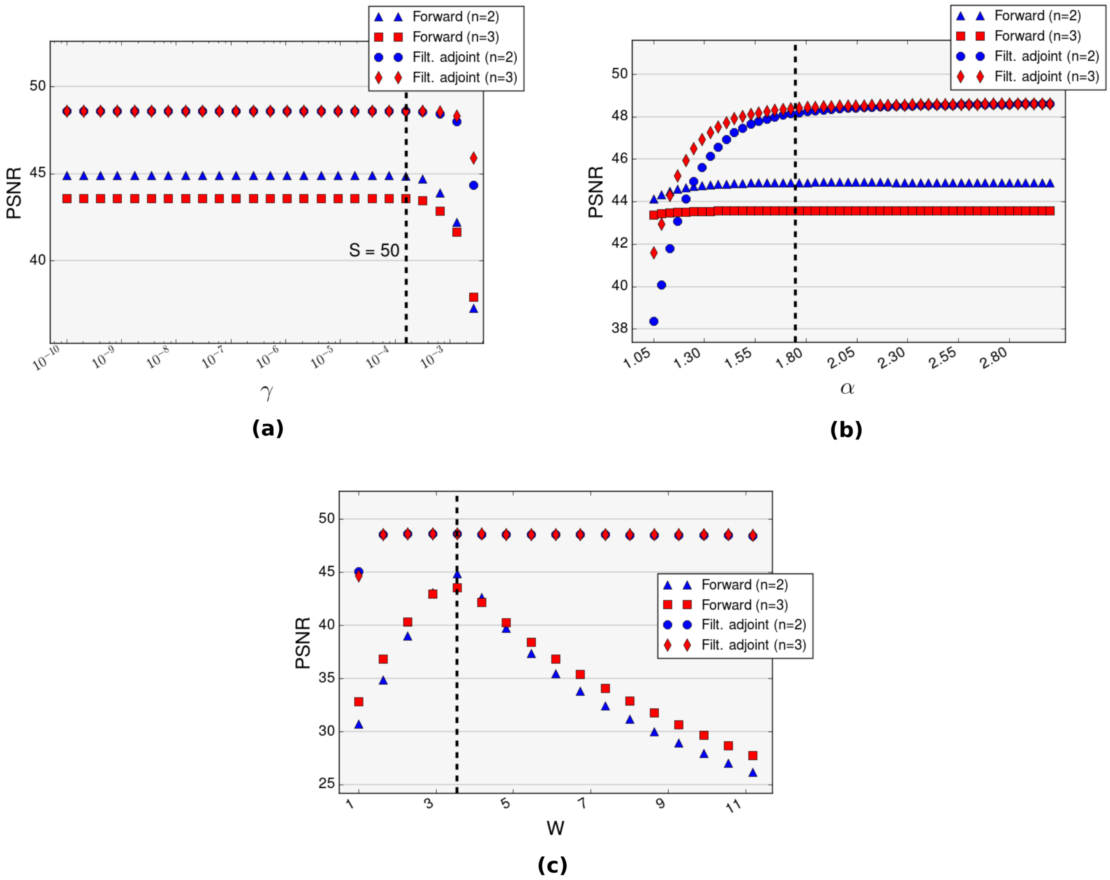
Table 13: (a) Reconstruction with Hilbert-filtered DBRP of an analytical SL sinogram (800 views × 512 pixels). (b) Reconstruction with Hilbert-filtered DBRP of a sinogram created by the DFRP (800 views × 512 pixels). The settings $\{W, \gamma, \alpha\}_{\text{nopt}_3}$ and $\{W, \gamma, \alpha\}_{\text{nopt}_4}$ do not correspond to the PSNR maximum surface for analytical and iterative reconstruction, respectively. The differences in PSNR are not negligible.

	$\{W, \gamma, \alpha\}_{\text{opt}_1}$	$\{W, \gamma, \alpha\}_{\text{nopt}_3}$
PSNR	24.51	21.75

(a)

	$\{W, \gamma, \alpha\}_{\text{opt}_2}$	$\{W, \gamma, \alpha\}_{\text{nopt}_4}$
PSNR	35.50	33.02

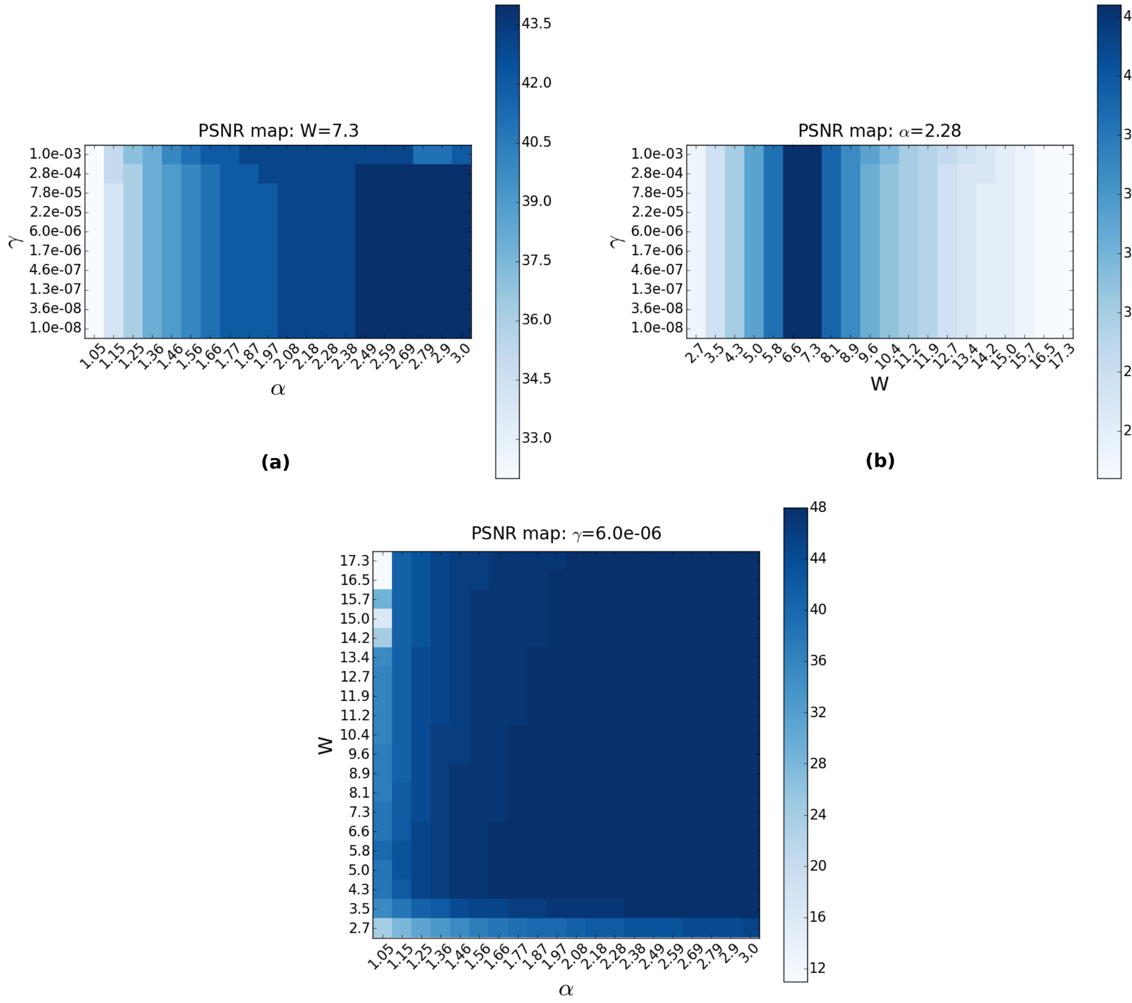
(b)



(a) **(b)**

(c)

Figure 4.13: Preliminary experiments to determine the interesting range where to span each parameter in a more general brute-force search. The DFRP projects $f^{(n)}$ with $n = 2, 3$: the result is compared to the analytical sinogram. The Hilbert-filtered DBRP reconstructs the analytical sinogram: the reconstructed slice is compared to the phantom. (a) Experiments spanning $\gamma \in [10^{-10}, 3.9 \cdot 10^{-3}]$ and setting $W = 4.45$ and $\alpha = 3.0$; all PSNR curves reach the maximum for $\gamma < 10^{-4}$ (corresponding to $S > 50$). (b) Experiments spanning $\alpha \in [1.05, 3]$ and setting $W = 4.45$ and $\gamma = 10^{-10}$; all PSNR curves reach the maximum for $\alpha > 1.8$. (c) Experiments spanning $W \in [1.9, 12.1]$ and setting $\gamma = 10^{-10}$ and $\alpha = 3.0$.



(a) **(b)**

(c)

Figure 4.14: Examples of 2D PSNR maps visualizing the results of the brute-force search of the 3D parameter space $\{W, \gamma, \alpha\}$. These experiments made use of the phantom $f^{(2)}$. (a) Map for the reconstruction with Hilbert-filtered DBRP of DFRP sinograms. (b) Map for the computation of DFRP sinograms. (c) Map for the reconstruction with Hilbert-filtered DBRP of analytical sinograms.

were confirmed, when using $f^{(n)}$ with $n = 1, 4, 5$.
Additional experiments with the Shepp-Logan (SL) [129] phantom have proved that this optimization approach is not biased by the choice of the family function. In the experiment of Tabs. 12(a) and 12(b), the settings $\{W, \gamma, \alpha\}_{\mathrm{nopt}_1}$ and $\{W, \gamma, \alpha\}_{\mathrm{nopt}_2}$, belonging to the PSNR maximum surface for analytical and iterative reconstruction respectively, are tested against the optimal settings. Table 12(a) shows the PSNR score related to the reconstruction with Hilbert-filtered DBRP of the SL analytical sinogram (800 views × 512 pixels) with $\{W, \gamma, \alpha\}_{\mathrm{opt}_1}$ and $\{W, \gamma, \alpha\}_{\mathrm{nopt}_1} = \{7.6,\ 1.7 \cdot 10^{-7},\ 2.1\}$. The PSNR score in Tab. 12(b) refers to the reconstruction with Hilbert-filtered DBRP of the sinogram (800 views × 512 pixels) computed by the DFRP with $\{W, \gamma, \alpha\}_{\mathrm{opt}_2}$ and $\{W, \gamma, \alpha\}_{\mathrm{nopt}_2} = \{7.6,\ 6.0 \cdot 10^{-7}, 2.7\}$. The negligible differences in PSNR confirm, that $\{W, \gamma, \alpha\}_{\mathrm{nopt}_1}$ and $\{W, \gamma, \alpha\}_{\mathrm{nopt}_2}$ correspond indeed to the respective PSNR maximum surfaces. The same approach is followed for the experiment in Tabs. 13(a) and 13(b), where, this time, the settings $\{W, \gamma, \alpha\}_{\mathrm{nopt}_3} = \{11.9,\ 6.0 \cdot 10^{-7}, 1.66\}$ and $\{W, \gamma, \alpha\}_{\mathrm{nopt}_4} = \{8.2,\ 6.0 \cdot 10^{-7}, 2.2\}$ are not located on the maximum surfaces for analytical and iterative reconstruction, respectively. The differences of PSNR in Tabs. 13(a) and 13(b), between optimal and suboptimal settings, are certainly not negligible.

4.3.6 Comparison with a standard implementation of $\mathcal{R}^{(1)}$

Benchmark procedure

To benchmark the accuracy and efficiency of the proposed operators, a standard space-based implementation of $\mathcal{R}^{(1)}$ has been considered: the forward projector introduced by [93] with additional differentiation along the channel direction. Considered a discrete image $f[m, n]$, the original implementation of the Radon Transform of [93], based on linear interpolation, is given by:

$$P_\theta(t) = \begin{cases} \dfrac{1}{\sin\theta} \displaystyle\sum_m \big[(1-\eta) f\left[m, \lfloor am + b \rfloor\right] + \eta f\left[m, \lfloor am + b \rfloor + 1\right]\big] & \text{if} \quad |\sin\theta| > \frac{\sqrt{2}}{2} \\ \dfrac{1}{\cos\theta} \displaystyle\sum_n \big[(1-\eta) f\left[\lfloor \tilde{a}n + \tilde{b} \rfloor, n\right] + \eta f\left[\lfloor \tilde{a}n + \tilde{b} \rfloor + 1, n\right]\big] & \text{if} \quad |\sin\theta| \leqslant \frac{\sqrt{2}}{2} \end{cases} \tag{4.45}$$

where

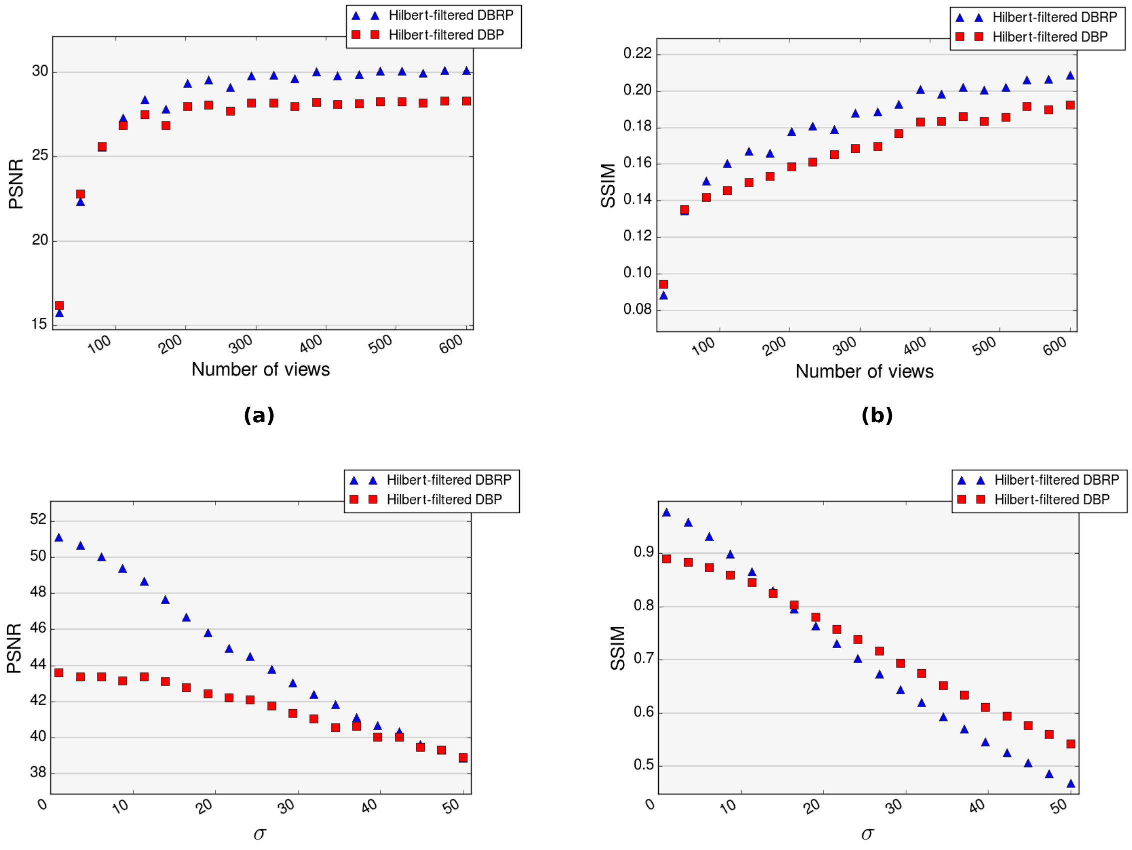
$$\begin{aligned} a &= -\frac{\cos\theta}{\sin\theta} & b &= \frac{t - x_{\min}\cos\theta - y_{\min}\sin\theta}{\sin\theta} \\ \tilde{a} &= \frac{1}{a} & \tilde{b} &= \frac{t - x_{\min}\cos\theta - y_{\min}\sin\theta}{\cos\theta} \\ \eta &= \begin{cases} am + b - \lfloor am + b \rfloor & \text{if} \quad |\sin\theta| > \frac{\sqrt{2}}{2} \\ \tilde{a}n + \tilde{b} - \lfloor \tilde{a}n + \tilde{b} \rfloor & \text{if} \quad |\sin\theta| \leqslant \frac{\sqrt{2}}{2} \end{cases} & & . \end{aligned} \tag{4.46}$$

$\lfloor \ldots \rfloor$ represents the floor operation, $x_{\min}$ and $y_{\min}$ correspond to the smallest abscissa and ordinate of the selected ray within the image grid. $\mathcal{R}^{(1)}$ is computed from finite differences of Eq. (4.45) along the pixel direction, i.e. the t-axis. This operator is named differentiated Radon transform and abbreviated as DRT; its corresponding backprojector is abbreviated as DBP.

Experiments focus on the the impact of undersampling and noise in analytical and iterative reconstructions performed with the proposed operators on one side, DRT and DBP on the other side.
The PSNR and MSSIM are used for the assessment of the reconstruction quality of simulated data. For real data, since a reference image is not available, the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) are averaged for multiple regions-of-interest (ROIs) at different distances from the image center. The SNR is computed within a ROI, supposed to be homogeneous; for the CNR, two different neighboring ROIs are used, each supposed to be homogeneous. SNR and CNR are less reliable metrics than those relying on a reference image, but they can, nevertheless, give an approximate indication of image quality.

Validation of the proposed operators for analytical reconstruction

Analytical reconstructions are performed by the Hilbert-filtered DBRP (HF-DBRP) with optimal parameters $\{W, \gamma, \alpha\}_{\text{opt}_1} = \{4.45, 1.7 \cdot 10^{-6}, 1.75\}$ and Hilbert-filtered DBP (HF-DBP).
The first experiment deals with a simulated undersampled and noisy dataset of $f^{(n)}$ $(n = 1, 2, 3)$ with 512×512 pixels, computed by means of (4.43). A set of noiseless analytical sinograms is created with a number of views ranging from 20 to 600. Another set is, instead, created with an increasing amount of Gaussian noise, with σ ranging from 1% to 50% of the average value, μ, of the sinogram, while the number of views is fixed to 805.
Since reconstructions with HF-DBRP and HF-DBP, for the same condition of either undersampling or noise, are indistinguishable at visual inspection, the analysis of the results relies on the information provided by PSNR and MSSIM (Figs. 4.15(a)-4.15(d)). Figures 4.15(a) and 4.15(b) show that the reconstructions with HF-DBRP have generally superior quality for a number of views $\geqslant$ 12% of what is required by the FBP sampling criterion (in this case, 100 views). For very low number of views, the HF-DBRP scores slightly worse than the space-based counterpart. Figures 4.15(c) and 4.15(d) show that the gridding algorithm appears quite sensitive to the level of noise affecting the data, since the drop of PSNR and SSIM for increasing noise is steeper compared to the other operator. The PSNR scores are everywhere higher for the gridding reconstruction, whereas, according to the MSSIM, for $\sigma > 20\%$, the reconstructions with HF-DBP seem to perform better.
These results indicate that analytical reconstruction with the gridding backprojector is less sensitive to undersampling than to noise. The HF-DBRP may slightly underperform compared to a standard method, when a substantial amount of noise affects the data.
The second experiment tackles the reconstruction of a DPC sinogram of a rat brain, acquired at the TOMCAT beamline of the Swiss Light Source at the Paul Scherrer Institute (Switzerland). The array has 721 views × 1493 pixels. The reconstructions, shown in Figs. 4.16(a) and 4.16(b), are indistinguishable at visual inspection; nevertheless, the reconstruction with HF-DBRP shows improved SNR and CNR values (Tab. 14(a)). When the original sinogram is downsampled to 400 and 300 views, each pair of reconstructions still looks identical and the difference of SNR and CNR becomes rather small (Tab. 14(b) and 14(c)).
The computational efficiency represents the very attractive feature of the proposed operators. Table 15 reports the time elapsed (in seconds) for the reconstruction of DPC sinograms of various sizes. The used hardware is an Intel(R) Core(TM) i7-3520M CPU 2.90GHz. The HF-DBRP can speed-up



(a) **(b)**

(c) **(d)**

Figure 4.15: PSNR and SSIM results for the comparison between the Hilbert-filtered DBRP and DBP for the analytical reconstruction of DPC undersampled and noisy sinograms. (a,b) Scores for the "undersampling-test" (for $f^{(1)}$). (c,d) Scores for the "noise-test" (for $f^{(3)}$); σ on the $x-$axis is the variance of the Gaussian noise corresponding to a percentage of the average value of the noiseless original sinogram.

calculations by two orders of magnitude for small datasets, up to three orders for bigger datasets with respect to a standard space-based operator.

Validation of the proposed operators for iterative reconstruction

Iterative reconstructions are performed by means of the alternate direction method of multipliers (ADMM) [27], solving a standard LASSO (Least Absolute Selection and Shrinkage Operator) problem. This method has been applied to DPC data in [103] with a more sophisticate formulation and a forward projector based on a cubic B-spline basis [102]. Details regarding the ADMM, its usage for tomographic reconstruction and the choice of the parameters are given in the Appendix.

Two implementations of the ADMM are tested in this section and the difference lies in the forward and backprojector used: the DFRP and its Hermitian adjoint operator (ADMM-DFRP) versus the DRT and its Hermitian adjoint operator (ADMM-DRT).

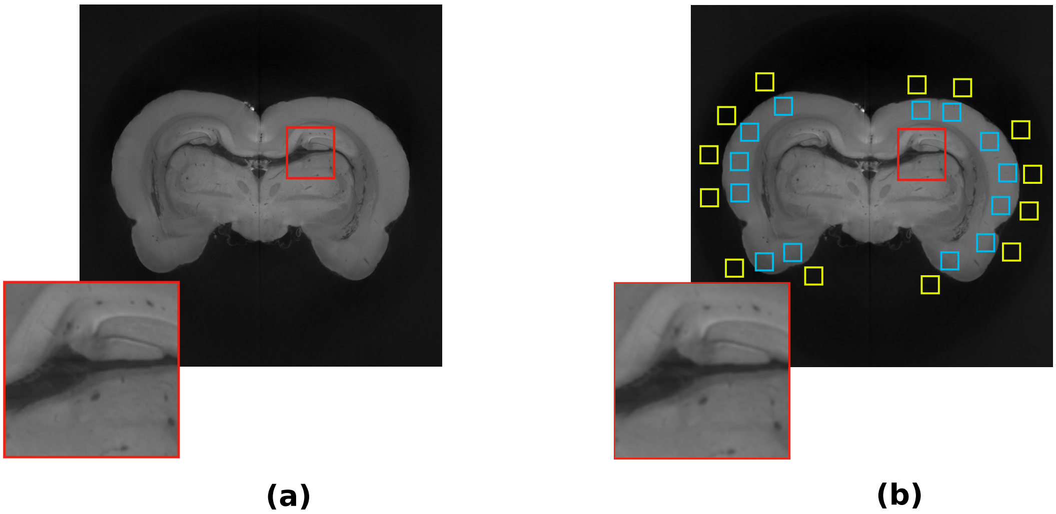


Figure 4.16: Analytical reconstructions of the DPC sinogram of a rat brain with a zoomed area. (a) Reconstruction with the Hilbert-filtered DBRP. (b) Reconstruction with the Hilbert-filtered DBP. The light blue and yellow squares in (b) indicate the ROIs used to computed the average SNR and CNR values reported in Tab. 14.

The first experiment focuses on the convergence of the algorithm: an analytical sinogram, created from $f^{(2)}$ and having 100 views × 256 pixels with additional Gaussian noise of $\sigma = 8\%\mu$ (this σ has been chosen to be slightly bigger than that estimated on the real data presented in the following), is reconstructed by means of ADMM and the relative norm of the difference between consecutive iterations, i.e. $||\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}||_2^2 / ||\mathbf{x}^{(k)}||_2^2$, is measured. The ADMM runs for 100 iterations. Figure 4.17

Table 14: SNR and CNR scores for the analytical reconstructions of the rat brain with 721 (a), 400 (b) and 300 (c) views. These values represent averages computed over the ROIs displayed in Fig. 4.16(b).

	HF-DBRP	HF-DBP
SNR	27.37	24.69
CNR	9.99	9.18

(a)

	HF-DBRP	HF-DBP
SNR	24.83	23.57
CNR	8.22	8.19

(b)

	HF-DBRP	HF-DBP
SNR	22.65	22.24
CNR	7.23	7.15

(c)

shows that ADMM-DFRP and ADMM-DRT have almost exactly the same convergence speed and that the solution remains practically unaltered after 20 iterations. This fact proves that the gridding operators do not alterate the algorithm convergence with respect to standard projectors. For the next reconstructions, the ADMM is stopped when $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2^2 / \|\mathbf{x}^{(k)}\|_2^2 < \epsilon = 5 \cdot 10^{-6}$.

In a second step, the iterative reconstructions of the noisy and undersampled simulated sinogram, employed for the convergence test, are analyzed. Images in Figs. 4.18(b) and 4.18(c) show no visible difference; the PSNR and SSIM analysis (Tab. 16(a)) reveals a slightly better accuracy for the ADMM-DFRP.

Finally the sinogram of the rat brain, also used in the previous section, is reconstructed with the ADMM approach. In this case, the number of projections has been reduced by a factor of 2.5 leading to an undersampled dataset with an array of 300 views x 1493 pixels. Also for real data, the reconstructed slices (Figs. 4.19(a) and 4.19(b)) look almost identical at visual inspection; the analysis, reported in Tabs. 16(a) and 16(b), highlights a slight better performance for the ADMM-DFRP.

From a computational point of view, the proposed operators are ideal for iterative algorithms, since these methods generally call the forward projector and its Hermitian adjoint few times per iteration, often representing the computational bottleneck of the entire reconstruction process. For this last dataset, on an Intel(R) Core(TM) i7-3520M CPU 2.90GHz, a single conjugate-gradient sub-iteration within the ADMM (see Appendix) takes around 40s with the proposed operators (ADMM-DFRP), and around 830s for ADMM-DRT. Since the stopping criterion terminates the procedure after approximately 20 iteration, each involving 15 sub-iterations, the time elapsed for the reconstructions, if run on a single core, is around 3.5h for the ADMM-DFRP and 69h for the ADMM-DRT.

4.3.7 Summary

Novel projectors have been introduced to address analytical and iterative reconstruction of tomographic differential phase contrast datasets, when the forward model is represented by the derivative of the Radon transform, $\mathcal{R}^{(1)}$. These operators are based on the gridding method and feature a complexity of $O(N^2 \log_2 N)$. The gridding parameters, offering the best trade-off between accuracy and computational performance, have been experimentally optimized on the basis of the minimal oversampling strategy, that requires the utilization of a Kaiser-Bessel kernel for the interpolation in Fourier space.

The performance of the proposed differentiated gridding forward projector (DFRP) and backprojector (DBRP) have been assessed with respect to a standard space-based implementation of $\mathcal{R}^{(1)}$ and $\mathcal{R}^{(1)*}$.

Table 15: Times elapsed to perform the analytical reconstructions of DPC sinograms of different sizes. The used hardware is an Intel(R) Core(TM) i7-3520M CPU 2.90GHz.

	HF-DBRP	HF-DBP
805 views × 512 pix.	0.23 s	28.93 s
1608 views × 1024 pix.	0.97 s	$2.40 \cdot 10^2$ s
3217 views × 2048 pix.	4.23 s	$1.20 \cdot 10^3$ s

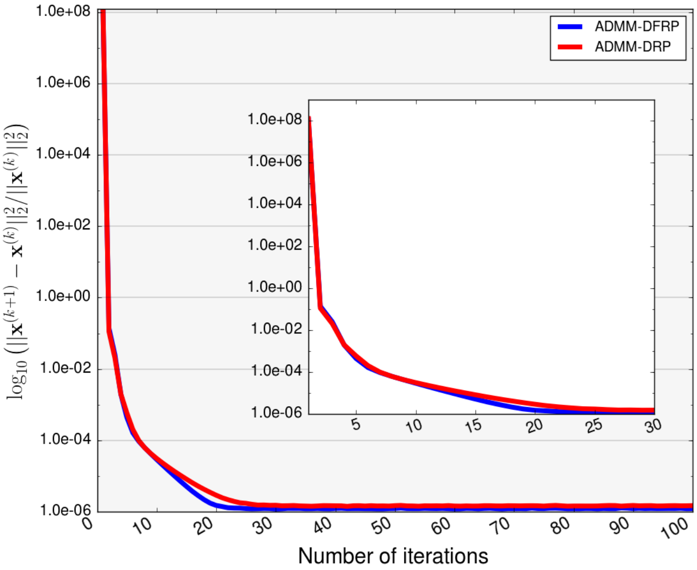


Figure 4.17: Convergence test showing the decrease of $\log_{10}\left(\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2^2/\|\mathbf{x}^{(k)}\|_2^2\right)$ as a function of the number of iterations.

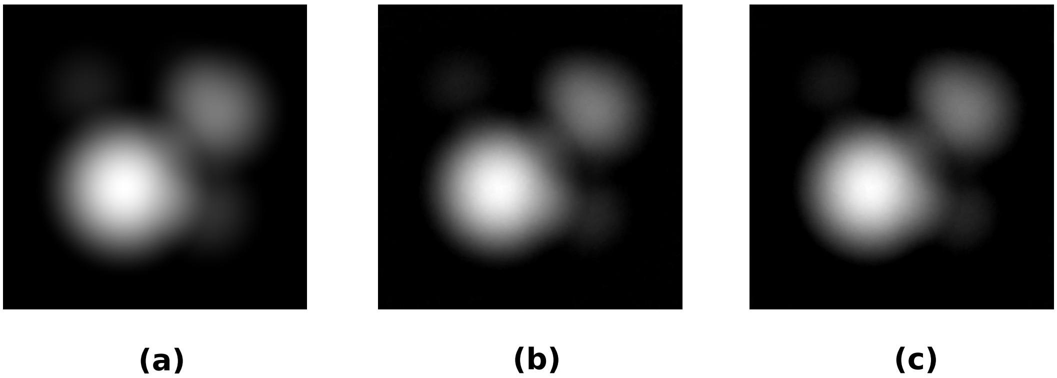


Figure 4.18: Iterative reconstructions of a noisy undersampled DPC sinogram of $f^{(2)}$. (a) Original phantom. (b) Reconstruction with ADMM-DFRP. (c) Reconstruction with ADMM-DRT.

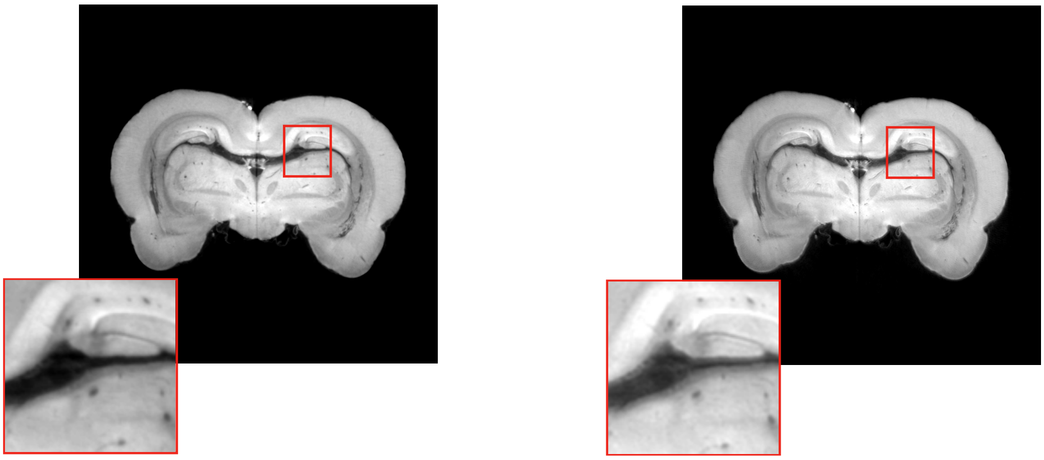
Table 16: (a) PSNR and MSSIM scores for the iterative reconstructions shown in Figs. 4.18(b) and 4.18(c). (b) Average SNR and CNR computed for multiple ROIs at different distances from the image center for the iterative reconstructions shown in Figs. 4.19(a) and 4.19(b). The SNR and CNR scores were averaged over the ROIs displayed in Fig. 4.16(b).

	ADMM-DFRP	ADMM-DRT
PSNR	31.42	30.76
MSSIM	0.74	0.67

(a)

	ADMM-DFRP	ADMM-DRT
SNR	29.34	29.01
CNR	13.18	12.63

(b)


(a)

(b)

Figure 4.19: Iterative reconstructions of a DPC undersampled sinogram of a rat brain with a zoomed area. (a) Reconstruction with ADMM-DFRP. (b) Reconstruction with ADMM-DRT.

The comparison has shown that the differentiated gridding operators can yield analytical and iterative reconstructions of the same quality as space-based operators. The great advantage of the proposed operators lies in the computational efficiency, since calculation times can be reduced by 3 orders of magnitude.

Appendix: LASSO solved by the alternate direction method of multipliers

The LASSO problem corresponds to a L_1 penalized linear regression of the form:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \lambda \, ||\mathbf{x}||_1 \right] \quad . \tag{4.47}$$

Considering the reconstruction problem presented in Sec.4.3.6, $\mathbf{A}$, i.e., $\mathbf{A} = \mathcal{R}^{(1)}$, is the matrix representation of the forward projector (therefore, $\mathbf{A}^\dagger$ is the Hermitian adjoint operator or backprojector),

i.e., $\mathbf{b}$ is the DPC sinogram, $\mathbf{x}$ is the unknown slice of the object and $\lambda \in \mathbb{R}$ is the regularization weight.
The LASSO problem can be numerically tackled by the alternate direction method of multipliers (ADMM) [27], that maps (4.47) into the minimization of the following augmented Lagrangian:

$$\mathcal{L}_\mu(\mathbf{x}, \mathbf{u}, \mathbf{m}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \sum_{\mathbf{k}} \|\mathbf{u}_\mathbf{k}\|_1 + \mathbf{m}^T(\mathbf{x} - \mathbf{u}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{u}\|_2^2 \tag{4.48}$$

where $\mathbf{m}$ are the Lagrangian multipliers and $\mathbf{u}$ is an auxiliary variable or the dual image.
The ADMM iteratively minimizes $\mathcal{L}_\mu$ by sequentially solving smaller problems; each iteration involves two sub-optimizations with respect to $\mathbf{x}$ and to $\mathbf{u}$, followed by the update of $\mathbf{m}$:

1. $\mathbf{x}^{(k+1)} \longleftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}_\mu\left(\mathbf{x}, \mathbf{u}^{(k)}, \mathbf{m}^{(k)}\right)$
2. $\mathbf{u}^{(k+1)} \longleftarrow \underset{\mathbf{u}}{\operatorname{argmin}}\, \mathcal{L}_\mu\left(\mathbf{x}^{(k+1)}, \mathbf{u}, \mathbf{m}^{(k)}\right)$
3. $\mathbf{m}^{(k+1)} \longleftarrow \mathbf{m}^{(k)} + \mu\left(\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)}\right)$

In step (1), the conjugate gradient method (CG) is applied to the following linear system:

$$(\mathbf{A}^\dagger\mathbf{A} + \mu\mathbf{I})\mathbf{x} \ = \ \mathbf{A}^\dagger\mathbf{b} + \mu\left(\mathbf{u}^{(k)} - \frac{\mathbf{m}^{(k)}}{\mu}\right) \quad . \tag{4.49}$$

The subproblem of step (2) is solved through a shrinkage operation:

$$\mathbf{u}^{(k+1)} = \max\left\{\left|\mathbf{x}^{(k+1)} + \frac{\mathbf{m}^{(k)}}{\mu}\right| - \frac{\lambda}{\mu}, 0\right\} \operatorname{sgn}\left(\mathbf{x}^{(k+1)} + \frac{\mathbf{m}^{(k)}}{\mu}\right) \quad . \tag{4.50}$$

Several parameters are required by the ADMM: the stopping criterion, the number of sub-iterations of the CG, λ and μ. The more the conditioning number of $\mathbf{A}^\dagger\mathbf{A}$ differs from 1, the more sub-iterations are required by the CG. This number was set to 15. The parameter λ rules the trade-off between fidelity and penalty term, for this reason, it should be selected depending on the amount of noise affecting the data. Suited values for λ and μ have been found through several experiments with simulated data. The reconstruction of simulated and real data are, in particular, performed with $\lambda = \mu = 1.0$. It is important to point out that the trends presented in Sec.4.3.6 resulted experimentally indipendent from the choice of these two parameters.

4.4 Perspectives for Fourier-based tomographic reconstruction

The geometry and the computational facilities usually provided by synchrotron imaging beamlines motivate the usage of the gridding method or iterative algorithms based on gridding projectors. Various studies have shown that analytical and iterative Fourier-based tomographic reconstruction can perform very similarly in terms of accuracy to FBP or iterative methods making use of standard projectors. Synchrotron imaging beamlines like the Swiss Light Source (SLS), the European Synchrotron Radiation Facility (ESRF) and the Advanced Photon Source (APS) offer CPU clusters to process the raw projection data, which represents the most suited enviroment for the utilization of Fourier-based reconstruction methods. Only recently, ESRF and APS have been equipped with GPU clusters as well.

Putting for a moment the specific context of this project aside and considering the broader picture, the question that naturally arises is: *does it make sense to push further the study of Fourier-based reconstruction methods?* To address this question, we will focus on two different aspects: complex geometries and computational efficiency.

Fourier-based methods have been explored for more complex geometries like fan-beam [164, 171] and cone-beam [165] since the late 90s, but this line of research did not have much of a follow-up (actually for all geometries). For the cone-beam configuration, the reason may be connected to the memory bottleneck. The GM requires to keep the entire volume in memory and, considering that oversampling is needed, the reconstruction of a real dataset can easily result computationally not practical for standard machines. If the minimal oversampling strategy [28, 168] applied to the study conducted in [165] and considering future upgrades of computing machines in terms of memory capacity, GM-based cone-beam reconstruction will probably be feasible and competitive with existing algorithms.

Fourier-based methods greatly outperform their real-space counterparts in speed, if the implementations are all considered on a single core. However, the last two decades have witnessed the introduction and fervid development of real-space tomographic projectors implemented on GPUs [107, 108, 210–214] (just to cite few of the numerous works published on the topic). How do the gridding projectors perform with respect to GPU-implemented operators? It is not trivial to answer to this question. Beside the algorithmical part, there are so many code-related and machine-architecture details that can substantially change the outcome of such comparison. Although the tests were performed on different machines, we can look at a couple of figures in [28] related to the FGP and in [215] related to GPU-based projectors: to reconstruct a sinogram with 3200 views × 2048 pixels, (a non-optimized) FGP with minimal oversampling requires 1.47s; to reconstruct a sinogram with 3072 views × 2048 pixels, ASTRA [108] takes 1.2s, NiftyRec [107] 35s. The presented figures indicate the possibility that gridding projectors with minimal oversampling can compete with operators implemented on GPUs.

4.5 Coupling projector-backprojector for ADMM

The following section represents a reprinted manuscript published as: F. Arcadu, F. Marone and M. Stampanoni, *"An experimental study on the coupling projector-backprojector in iterative tomographic reconstruction"*, submitted, 2016 [30].

4.5.1 Abstract

The performance of an iterative reconstruction algorithm for X-ray tomography is strongly determined by the features of the used forward and backprojector. For this reason, a large number of studies has focused on the to design of projectors with increasingly higher accuracy and speed. To what extent the accuracy of an iterative algorithm is affected by the mathematical affinity and the similarity between the actual implementation of the forward and backprojection, referred here as *coupling projector-backprojector*, has been an overlooked aspect so far.

The experimental study presented here shows that the reconstruction quality and the convergence of an iterative algorithm greatly rely on a good matching between the implementation of the tomographic operators. In comparison, other aspects like the accuracy of the standalone operators, the usage of physical constraints or the choice of stopping criteria may even play a less relevant role.

4.5.2 Introduction

Iterative reconstruction for X-ray tomography has been studied since the introduction of the first CT scans in the mid 70s [36]. Differently from the filtered backprojection (FBP) algorithm [95], iterative methods are non-linear and less computationally efficient, as the forward projector and its adjoint operator, the backprojector, are generally called few times per iteration. In contrast to FBP, iterative methods can, however, provide high quality reconstructions of tomographic underconstrained datasets, characterized by poor signal-to-noise ratio (SNR), little number of views and/or missing data.

In general, iterative algorithms consist of the following elements: a solver for the cost function, physical constraints, a regularization scheme linked to the a-priori-knowledge regarding the specimen under study and tomographic projectors.

Four main families of solvers can be identified for iterative reconstruction. Algebraic reconstruction techniques like ART [13], SIRT [14] and SART [15] handle the tomographic problem as a system of equations, which is solved by means of the Kaczmarz method [16]. Statistical methods as the maximum likelihood expectation maximization (MLEM) [110], the separable paraboloidal surrogate [17] and the penalized weighted least square method (PWLS) [18, 111] incorporate the statistical model ruling the signal formation at the detector. Recently, modern techniques for convex optimization like the split Bregman method [112] and the alternate direction method of multipliers (ADMM) [113] have also been applied to tomographic reconstruction [114–116]. Finally, the projection-onto-convex-sets method [118] has been mainly used to address the interior tomography problem.

Physical constraints enforce at each iteration strict conditions in the image domain. Setting to zero all negative pixel values and those falling outside the reconstruction circle is a typical example of broadly exploited physical constraints.

Regularization schemes often utilized by iterative algorithms are Tikhonov [125], Huber [126] and total variation (TV) [128]. In particular, a Huber or TV term can steer the cost function towards a piecewise-constant solution, while preserving the spatial resolution.

Several implementations of the tomographic projectors have been proposed since the 70s. The pixel-driven [95, 216, 217], ray-driven [95, 96, 217], distance-driven [94] and slant-stacking [93, 218] approaches are different methods to approximate the Radon transform in real domain. Since the listed approaches feature a complexity of $\mathcal{O}(N^3)$ [93], their implementation on GPUs is a must to make iterative reconstructions computationally feasible [108, 215, 219, 220]. Tomographic projectors with complexity $\mathcal{O}(N^2 \log_2 N)$, based on hierarchical-decomposition [109], the non-uniform Fourier transform [170] or the gridding method [28], are, instead, fast enough to not necessarily require a GPU architecture.

So far, research in iterative reconstruction algorithms has mainly addressed the design of regularization schemes leading to a better SNR-spatial resolution trade-off and the development of tomographic projectors with increasingly higher accuracy and speed. An aspect that has been generally neglected is the role of the *coupling projector-backprojector*, i.e., the level of mathematical affinity and matching between the actual implementation of the forward projector and its adjoint operator.

This work is an empirical investigation of the role played by this aspect on the performance of iterative reconstruction algorithms for X-ray tomography. Ad-hoc experiments with state-of-the-art implementations of different tomographic operators (pixel-driven, ray-driven, distance-driven, slant-stacking, gridding method) have been designed for this purpose. Reconstructions have been performed with both analytical (FBP) and iterative (ADMM, PWLS, MLEM, SIRT) schemes. Results show that the coupling projector-backprojector substantially affects accuracy and convergence of an iterative algorithm. In some cases, the degree of matching between the tomographic projectors can even play a more decisive role for the performance of the iterative method than other factors, like physical constraints, stopping criteria or the accuracy of the standalone projectors.

A mathematical justification of the presented experimental results is not straightforward and is outside the scope of this work. The aim of this study is, instead, to provide convincing experimental evidence that a well-tuned coupling projector-backprojector is an absolute "must" for iterative tomographic reconstruction schemes to avoid systematically sub-accurate results. A practical strategy for measuring the coupling degree is also proposed: this tool could be very useful for users and developers of software packages for iterative tomographic reconstruction to assess and validate the quality of the proposed projector pairs.

4.5.3 Experimental framework

Tomographic projectors

The Radon transform, $\mathcal{R}$, integrates a function $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ over an hyperplane $HY(\mathbf{n},t) = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x}\cdot\mathbf{n} = t\}$, where $\mathbf{n}$ is a unit vector and $t \in \mathbb{R}$ is the signed distance from the origin [92]:

$$\mathcal{R}\{f\}(\mathbf{n},t) := \int\limits_{HY} d\mathbf{x}\, f(\mathbf{x}) = \int\limits_{\mathbb{R}^n} d\mathbf{x}\, \delta(t - \mathbf{x}\cdot\mathbf{n}) f(\mathbf{x}) = \int\limits_{\mathbf{n}^\perp} d\mathbf{x}\, f(t\mathbf{n} + \mathbf{x}) \quad . \tag{4.51}$$

δ is the Dirac function and $\mathbf{n}^\perp = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x}\cdot\mathbf{n} = 0\}$ is the subspace orthogonal to $\mathbf{n}$. For $n = 2$, $f(\mathbf{x}) = f(x_1, x_2)$, $\mathbf{n} = (\cos\theta, \sin\theta)$, HY is a line of equation $\mathbf{x}\cdot\mathbf{n} = x_1\cos\theta + x_2\sin\theta = t$, thus, $\mathcal{R}$ integrates f along lines. The second definition in (4.51) simplifies to:

$$\mathcal{R}\{f\}(\theta,t) := \int\limits_{-\infty}^{+\infty} dx_1 \int\limits_{-\infty}^{+\infty} dx_2\, f(x_1,x_2)\, \delta(x_1\cos\theta + x_2\sin\theta - t) \quad . \tag{4.52}$$

$\mathcal{R}$ is also called forward projector and θ is, here, the angle formed by the detector line and the positive x-semiaxis. The dual transform, i.e., the adjoint of the Radon transform, $\mathcal{R}^*$, is called backprojection. For $n = 2$ and given a generic function $g(\mathbf{x}) = g(x_1, x_2)$, $\mathcal{R}^*$ is defined as [92]:

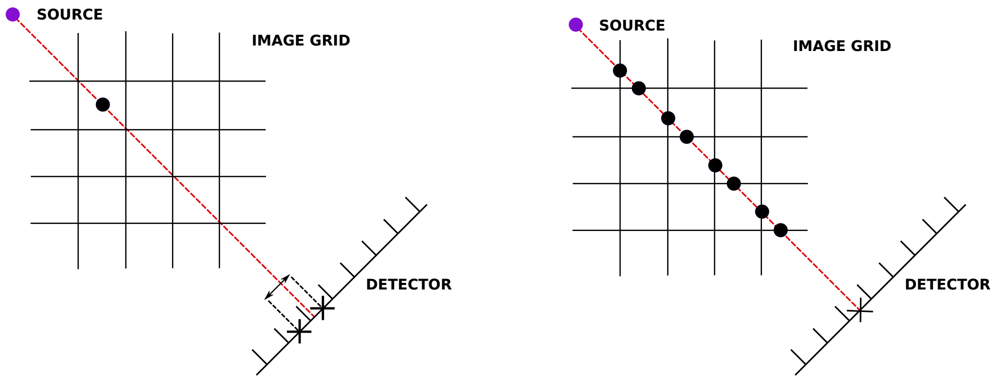
$$\mathcal{R}^*\{g\}(\mathbf{x}) = \frac{1}{2\pi} \int\limits_0^{2\pi} d\theta\, g\left(\theta, x_1\cos\theta + x_2\sin\theta\right) \quad . \tag{4.53}$$

The six implementations of $\mathcal{R}$ and $\mathcal{R}^*$ used in this work are for parallel beam geometry and a brief description is given in the following.

The pixel-driven (PD) approach [95, 216, 217] works by connecting the source point to the selected pixel center until intersection with the detector line, as displayed in Fig.4.20a. The pixel value is assigned on the basis of a linear interpolation scheme to the two detector cells that enclose the ray end point (they are indicated with a cross in Fig.4.20a).

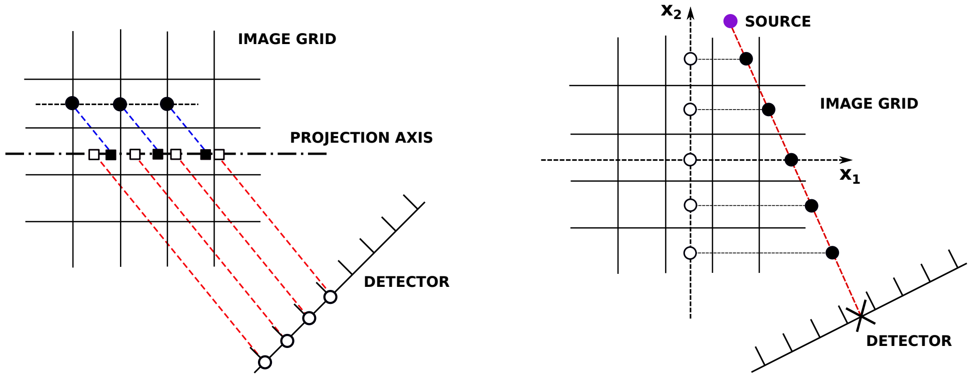
The ray-driven (RD) approach [95, 96, 217] connects the source to the center of a selected detector cell as shown in Fig.4.20b. The Siddon algorithm [181] is used to compute the intersection points of the ray with the image grid (black dots in Fig.4.20b). Each pixel contributes to the selected detector cell according to the ray path length.

The distance-driven (DD) approach [94, 98] in Fig.4.20c projects the pixel boundaries (black dots) of each image row/column and the detector cell boundaries (white dots) onto a common axis (in Fig.4.20c, the black squares are projected pixel boundaries, the white squares projected detector cells). The overlap between the interval defined by the projected boundaries of an image pixel and the one defined by the projected boundaries of a detector cell weights the contribution of the selected



(a) Pixel-driven approach

(b) Ray-driven approach



(c) Distance-driven approach

(d) Slant-stacking approach

Figure 4.20: Schematic representation of the different mechanisms characterizing the pixel-driven, ray-driven, distance-driven and slant-stacking approach for forward projection (and backprojection).

image pixel to the selected detector cell (and viceversa).
The slant stacking (SS) [93] method connects the source to each detector cell and divides the interval $[0, \pi]$ in two regions: one for nearly-vertical lines $0 \leqslant \theta \leqslant \pi/4$ and $3\pi/4 \leqslant \theta \leqslant \pi$; one for nearly-horizontal lines $\pi/4 \leqslant \theta \leqslant 3\pi/4$ (θ is the angle formed by the detector line and the positive x-semiaxis). Figure 4.20d shows how SS works with a nearly-vertical line: the abscissas of the black dots are obtained by using the ray equation and the x_2 coordinates (white dots) of all image pixels. The computed points (x_1, x_2) contribute to the selected detector cell according to a linear interpolation scheme. The same approach is used for nearly-horizontal lines.
The gridding projectors [28, 170] are implementations of $\mathcal{R}$ and $\mathcal{R}^*$ in the Fourier domain and are based on the Fourier slice theorem (FSM) [92]. For the forward operation, the input image grid is, first, multiplied with the deapodization matrix and, then, Fourier transformed (FFT-2D). The Fourier Cartesian grid is convolved with a compact kernel to obtain Fourier samples on a polar grid. According to the FSM, the inverse Fourier transform (IFFT-1D) of a polar slice at angle θ corresponds to the object projection acquired at angle θ. The accuracy and efficiency of gridding projectors rely entirely on the choice of the convolving kernel (that also determines the deapodizer) and the oversampling ratio, α, used for the Fourier grid. In this work, two slightly different implementations are considered [28]: one using a prolate spheroidal wavefunctions kernel and $\alpha = 2$ (abbreviated with WF); the other using a Kaiser-Bessel kernel and $\alpha = 1.5$ (abbreviated with KB).

Degree of coupling projector-backprojector

Given a generic linear operator $\mathcal{A} : \mathbb{C}^{n_1} \longrightarrow \mathbb{C}^{n_2}$, the adjoint, $\mathcal{A}^*$, is defined as follows:

$$\mathcal{A}^* : \mathbb{C}^{n_2} \longrightarrow \mathbb{C}^{n_1} \quad \text{such that} \quad \langle \mathbf{y}, \mathcal{A}(\mathbf{x}) \rangle = \langle \mathcal{A}^*(\mathbf{y}), \mathbf{x} \rangle \quad \forall \, \mathbf{x} \in \mathbb{C}^{n_1}, \forall \, \mathbf{y} \in \mathbb{C}^{n_2} \quad , \tag{4.54}$$

where $< \dots >$ is the notation for the inner product. Definition (4.54) can be used to measure how well a computer implementation of $\mathcal{A}$ matches the computer implementation of $\mathcal{A}^*$. The two inner products in (4.54) are numerically evaluated with $\mathbf{x}$ and $\mathbf{y}$ being vectors of randomly generated numbers. If the ratio $r = \langle \mathcal{A}^*(\mathbf{y}), \mathbf{x} \rangle \, / \, \langle \mathbf{y}, \mathcal{A}(\mathbf{x}) \rangle$ matches 1 up to a reasonably sufficient numerical precision, the implementations of $\mathcal{A}$ and $\mathcal{A}^*$ can be considered well coupled.
For the tomographic case, a good coupling is achieved when the backprojector foresees the same exact operations of the forward projector, but in reverse order and switching the roles of input/output arrays for object and sinogram. The coupled implementations of $\mathcal{R}$ and $\mathcal{R}^*$ listed in (4.5.3) feature $r = 1$ up to the 7th digit. When not coupled, $r = 1$ at most up to the 4th digit.

Reconstruction algorithms

Analytical reconstructions are here performed with filtered backprojection (FBP) [95], that inverts the Radon transform by applying the linear operator $\mathcal{R}^* \circ \Delta$, where Δ is the ramp or Ram-Lak filter. The trade-off between SNR and spatial resolution of FBP reconstructions depends on the type of window superimposed to the Ram-Lak filter [12]. For this reason, FBP is used here with four different filters [106]: a pure Ram-Lak filter that provides the highest spatial resolution and poorest SNR (abbr. RAMP); a Ram-Lak filter combined with a Shepp-Logan window (abbr. SHLO); a Ram-Lak filter combined with a Hanning window (abbr. HANN); a Ram-Lak filter combined with a Parzen window that provides the poorest spatial resolution and highest SNR (abbr. PARZ).

Four different iterative reconstruction algorithms have been selected for this study: the alternate direction method of multipliers (ADMM) with TV regularization [115], the penalized weighted least square (PWLS) with Huber penalty [18, 111], the maximum-likelihood expectation maximization (MLEM) [110] and the simultaneous iterative reconstruction technique (SIRT) [13]. The number of iterations is set to around 100, when studying the algorithm convergence. For the other experiments, the stopping criterion and regularization strength are optimized according to the characteristics of the considered dataset. Iterative reconstructions are run with a range of different stopping criteria and weights of the penalty term. We define the optimal number of iterations and regularization strength as those providing the best reconstruction accuracy, after appropriate exploration of the parameter space. Nevertheless, it is important to point out that the presented trends in the performance of the iterative algorithms as a function of the coupling projector-backprojector are independent from the choice of the regularization parameters and confirmed also in case of a suboptimal selection.

Dataset and image quality assessment

The Shepp-Logan (SL) phantom [129] is used to create the simulated datasets for this study. Since this phantom consists exclusively of roto-translated ellipses, its forward projection can be computed analytically [12]. An analytical forward projection can be used in two ways: (i) as reference when measuring the accuracy of a projector; (ii) as tomographic dataset not coupled to a specific operator used within the selected reconstruction algorithm. A selection of experiments presented in Section 4.5.4 and 4.5.5 were also performed with different simulated objects and real datasets: the observed trends are comparable to those obtained with the SL phantom and are, therefore, independent from the chosen object.

The discretized forward projection of an object is also called sinogram, which corresponds to a matrix $\in \mathbb{R}^{M \times N}$; M is the number of views and N the number of detector cells. In this study, projections are always homogeneously distributed in $[0, \pi)$. A sinogram in parallel beam geometry is considered undersampled, when $M < N\pi/2$ [12]. FBP reconstructions of undersampled datasets are affected by radially arranged line artifacts [12]. To simulate projections with a low photon statistics, Gaussian noise with variance σ is added to the computed forward projection.

Four different analytical forward projections of the SL phantom are used in the experimental sections: a well-sampled, noiseless SL sinogram with 402 views × 256 pixels, abbreviated as SL-FULL; an undersampled, noiseless SL sinogram with 50 views × 256 pixels, abbreviated as SL-UNDER; a well-sampled, noisy SL sinogram with 402 views × 256 pixels and additional Gaussian noise with $\sigma = 3\%$ of the SL-FULL mean value, abbreviated as SL-NOISE; an undersampled noisy sinogram with 75 views × 256 pixels and additional Gaussian noise with the same σ of the SL-NOISE, abbreviated as SL-UCONSTR.

The image quality is measured by the peak-signal-to-noise ratio (PSNR) [144], defined as:

$$\text{PSNR} = 10 \log_{10}\left(\frac{\max\{\mathbf{r}\}^2}{\text{MSE}}\right) = 20 \log_{10}\left(\frac{|\max\{\mathbf{r}\}|}{\sqrt{\text{MSE}}}\right) \quad , \tag{4.55}$$

where the mean squared error (MSE) is:

$$\mathrm{MSE} = \frac{1}{PQ} \sum_{i=0}^{P-1} \sum_{j=0}^{Q-1} (f[i,j] - r[i,j])^2 \quad . \tag{4.56}$$

$\mathbf{r}, \mathbf{f} \in \mathbb{R}^{P \times Q}$ are the reference and the image to be evaluated, respectively. The PSNR is preferable over the MSE because more sensitive: as $(f[i,j] - r[i,j])^2$ appears at the denominator, even small differences can elicit non negligible variations of the PSNR value. In this study, the reference is either SL or its analytical forward projection. When comparing an analytical or iterative reconstruction to SL, the PSNR is computed within the reconstruction circle.

4.5.4 Operator coupling in analytical reconstruction

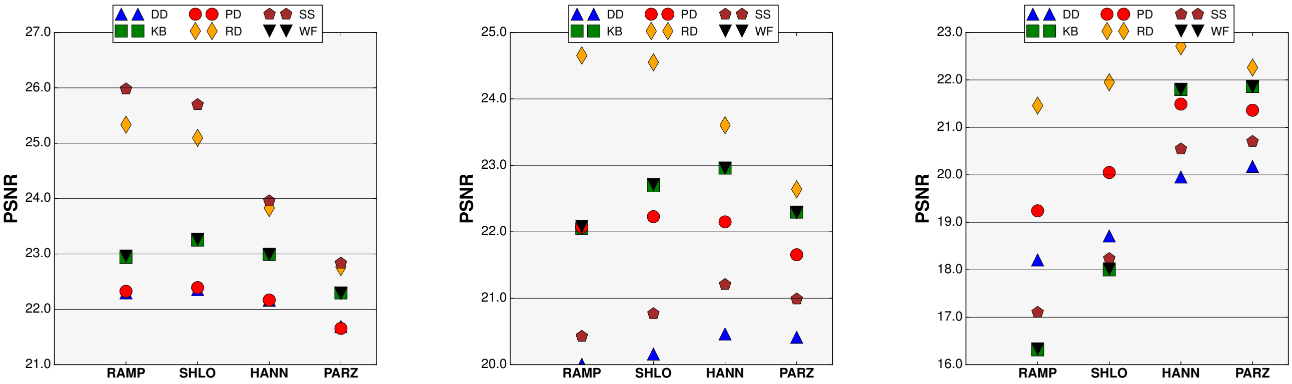
The following FBP tests provide a first indication of the role played by the coupling projector-backprojector in iterative tomographic reconstruction. Reconstructed slices are not displayed here, because differences are usually not detectable at visual inspection.
The accuracy of the standalone forward projectors DD, KB, PD, RD, SS and WF with respect to SL-FULL is reported in Tab.17. The standalone backprojectors are used to perform FBP reconstructions with different filters of SL-FULL, SL-UNDER and SL-UCONSTR[6] and the corresponding results are illustrated in Fig.4.21. The analysis in Tab.17 and Fig.4.21 suggest two facts. (i) The accuracy of the standalone projector is not a good predictor of the accuracy of the standalone backprojector in analytical reconstruction: e.g., KB has the lowest PSNR value in Tab.17, but it provides higher quality reconstruction of SL-FULL than PD and DD (Fig.4.21a). (ii) The performance of a backprojector is highly dependent on the characteristics of the dataset: e.g., SS has the highest PSNR score in Tab.17 and the best reconstruction quality for SL-FULL (Fig.4.21a), but it performs poorly when reconstructing underconstrained datasets (SL-UNDER and SL-UCONSTR in Fig.4.21b and 4.21c).
The experiment in Fig.4.22 evaluates the reconstruction accuracy of well-sampled noiseless sinograms created by DD (Fig.4.22a), KB (Fig.4.22b) and PD (Fig.4.22c).

Table 17: Accuracy of the standalone forward projectors with respect to SL-FULL.

	DD	KB	PD	RD	SS	WF
PSNR	39.49	37.64	39.35	39.35	45.53	37.57

---

[6] Results with SL-NOISE show the same trends characterizing the reconstruction of SL-FULL, SL-UNDER and SL-UCONSTR and therefore are not shown.

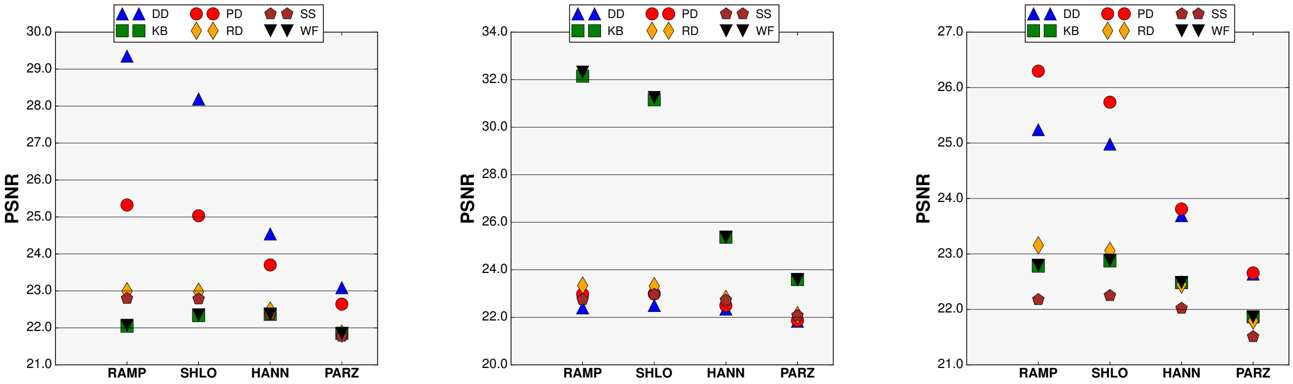


(a) SL-FULL

(b) SL-UNDER

(c) SL-UCONSTR

Figure 4.21: Accuracy of the standalone backprojectors in performing FBP reconstruction with different filters (RAMP, SHLO, HANN, PARZ) of SL analytical sinograms. Reconstruction of (a) SL-FULL, (b) SL-UNDER and (c) SL-UCONSTR.

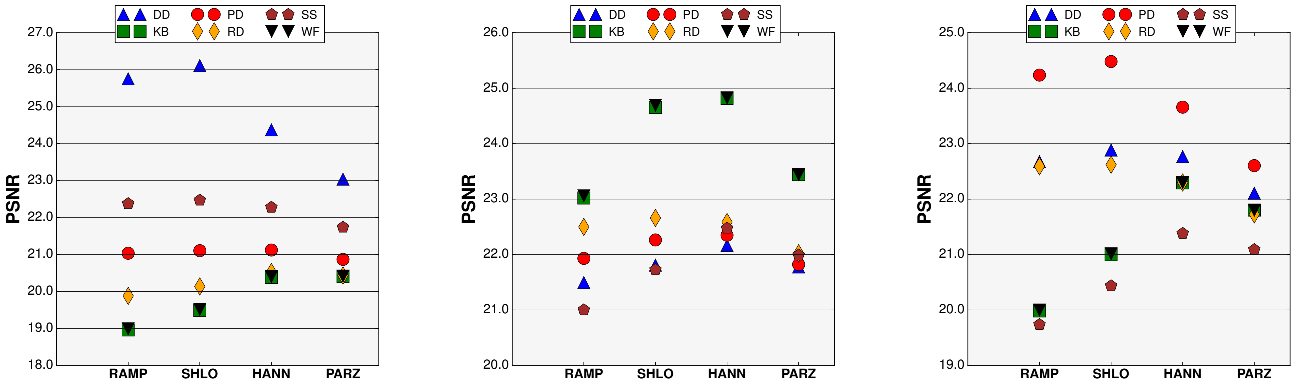


(a) FBP of DD sinogram

(b) FBP of KB sinogram

(c) FBP of PD sinogram

Figure 4.22: FBP reconstructions of sinograms with 402 views × 256 pixels created by the DD, KB and PD forward projectors. The reconstructions are performed with different filters by the DD, KB, PD, RD, SS and WF backprojectors.



(a) FBP of a DD undersampled sinogram (b) FBP of a KB noisy sinogram (c) FBP of a PD underconstr. sinogram

Figure 4.23: FBP reconstructions of (a) an undersampled sinogram with 100 views × 256 pixels created by the DD, (b) a noisy sinogram with 402 views × 256 pixel and additional Gaussian noise ($\sigma = 2\%$ of SL-FULL mean) created by KB and (c) an underconstrained sinogram with 100 views × 256 and additional Gaussian noise ($\sigma = 2\%$ of SL-FULL mean) created by PD. The reconstructions are performed with different filters by the DD, KB, PD, RD, SS and WF backprojectors.

The effect of the coupling projector-backprojector is clear: regardless of the filter choice, the best reconstruction quality is achieved when the backprojector matches the operator used to compute the input sinogram. The weaker the action of the filter, the more pronounced the impact of the coupling on the reconstruction accuracy. The results of the FBP reconstructions in Fig.4.23 show that the role of the coupling remains important even when dealing with undersampled (Fig.4.23a), noisy (Fig.4.23b) or underconstrained (Fig.4.23c) datasets. Considering that the performance of the standalone backprojectors can strongly vary as a function of the dataset (Fig.4.21), it is remarkable that undersampling and noise fail at breaking the effect of the coupling projector-backprojector. The important role of the coupling projector-backprojector is also clear when the sinograms are computed by RD, SS and WF (not shown).

4.5.5 Operator coupling in iterative reconstruction

To study the coupling effect on the convergence of iterative algorithms, SL-FULL is reconstructed with ADMM, PWLS, MLEM and SIRT. In each test, a different pair of forward and backward operators is used (Fig. 5-7). Only results for selected combinations of tomographic operators are shown in this section for illustration. The observed trends are however confirmed by all combinations.

ADMM converges and reaches the lowest value of the cost function when the backprojector matches the forward operator (Fig.4.24). When the backprojector does not match the forward operator, three different scenarios are observed. (i) ADMM converges but the cost function does not reach the minimum value (SS and PD curves in Fig.4.24a). (ii) ADMM simply does not converge (KB and WF curves in Fig.4.24b). (iii) ADMM reaches the lowest value of the cost function before diverging (DD curve in Fig.4.24a).

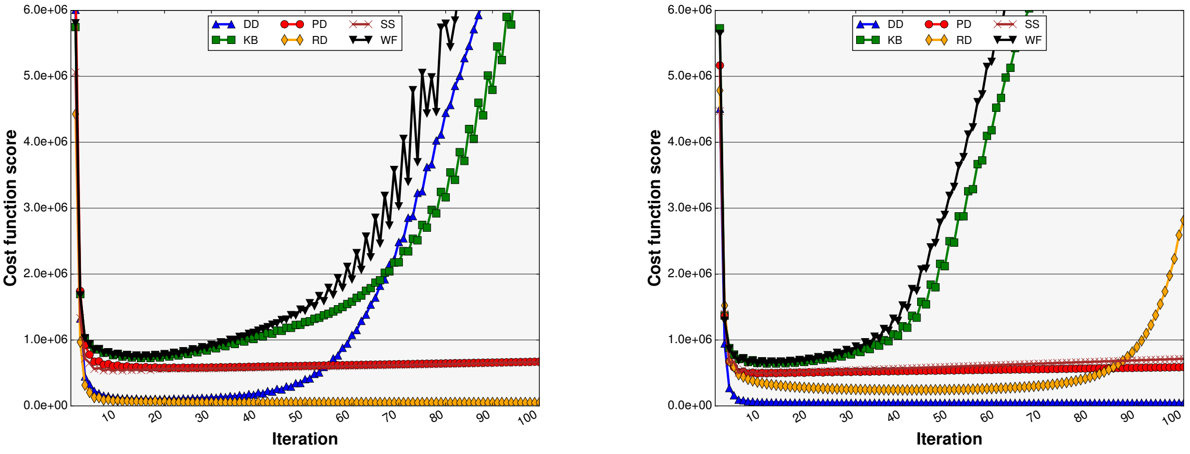
Differently from ADMM, the convergence of PWLS is not endangered by a mismatch between tomographic operators. Nevertheless, the cost function curve of PWLS with coupled operators is the lowest at each point after few initial iterations. This is visible in the insets of Fig.4.25a and 4.25b.

MLEM and SIRT behave similarly to PWLS: the matching between forward projector and backprojector is not essential to guarantee convergence, but is required to obtain the lowest cost function curve at each point, as shown in the insets of Fig.4.26a and 4.26b. Despite this similarity to PWLS, MLEM and SIRT can, instead, easily "explode" with an undersampled or noisy dataset if the operators are not coupled. For this reason, no reconstruction of underconstrained datasets done by SIRT and only few cases with MLEM are shown in the following. SIRT and MLEM share a common aspect: the computation of the diagonal matrix $\mathbf{C} = \{c_{jj} = 1/\sum_i a_{ij}\}$ is necessary, where $\{a_{ij}\}$ are the elements of the matrix representation of $\mathcal{R}$. The c_{jj}'s can be efficiently calculated as $\mathcal{R}^*(\mathbb{1})$, $\mathbb{R}^{M \times N} \ni \mathbb{1} = \{(\mathbb{1})_{ij} = 1 \ \forall i, j\}$. This computation can be rather sensitive and produce very high values at the image boundaries, compromising the stability of the iterative procedure especially when using uncoupled projectors. On the other hand, since ADMM and PWLS do not involve potentially sensitive computations, tests of these algorithms were not restricted to specific datasets or projector pairs.

The results in Fig.4.24, 4.25 and 4.26 clearly illustrate the influence of the coupling projector-backprojector on the convergence of all considered iterative procedures: the best performance is achieved only when the operators match. The level of the cost function of an iterative algorithm after a certain amount of iterations is not completely related to the reconstruction accuracy, or, in other words, reaching the minimum of the cost function does not necessarily mean reaching the closest possible approximation to the original phantom. Additional experiments focusing on the reconstruction accuracy have been performed. Reconstructions are displayed when differences can be perceived at visual inspection.

Table 18 presents the results of ADMM reconstructions of SL-UNDER with the PD forward projector. The best quality is achieved when the PD backprojector is used. Nevertheless, differences are relatively small and the reconstructions look very similar. The coupling has a much stronger effect when reconstructing SL-NOISE, as shown in Fig.4.27: the best ADMM reconstruction is obtained when the operators match (SS, in this case) and differences in PSNR are up to 3.6 dB. At visual inspection, reconstructions in Fig.4.27b, 4.27c and 4.27d are slightly more degraded than in Fig.4.27a, as suggested by the PSNR score. Results in Fig.4.28 show once again the great impact of the coupling effect on the reconstruction accuracy in presence of noise. Since KB and WF are both based on the gridding method and are highly coupled (as also resulting from the previous analysis), the reconstruction in Fig.4.28d is nearly identical to the one performed with matching operators in Fig.4.28a. The combination of a noisy underconstrained dataset and poorly coupled operators leads, instead, to strongly degraded ADMM reconstructions (Fig.4.28b and Fig.4.28c).

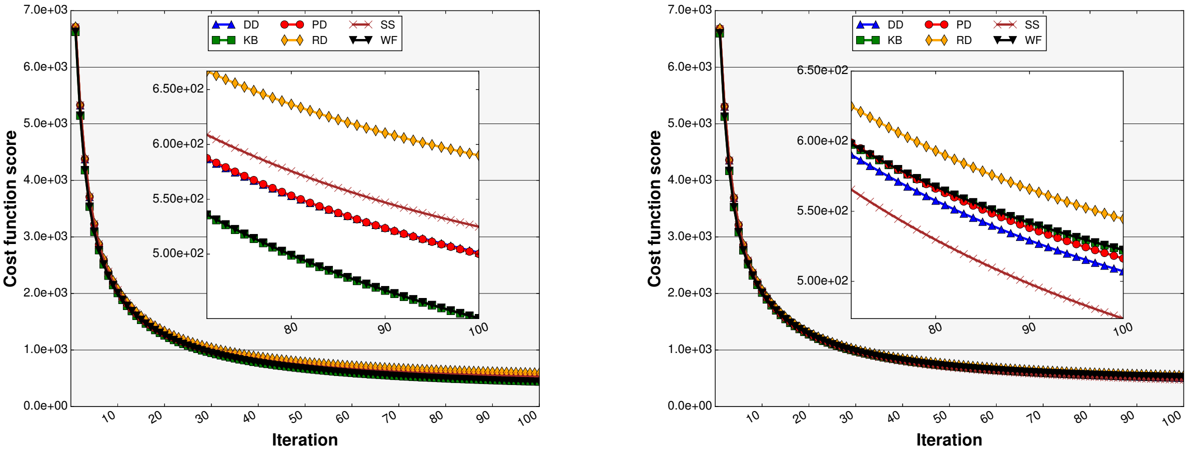
The PSNR values in Tab.19(a) and 19(b) correspond, respectively, to PWLS reconstruction of SL-UNDER using the KB forward projector with KB, DD, SS and WF backprojectors and of SL-NOISE using the SS forward projector with SS, DD, WF and PD backprojectors.For PWLS, the coupling projector-backprojector has slightly more impact in presence of undersampled data than of purely noisy data: the spread of PSNR values in Tab.19(a) is, indeed, a bit larger than for the values in Tab.19(b). Similarly to the results of Fig.4.28, the PWLS reconstruction of underconstrained datasets with coupled projectors has the highest accuracy (Fig.4.29a), whereas severe artifacts can occur when



(a) ADMM $-\mathcal{R}=$RD

(b) ADMM $-\mathcal{R}=$DD

Figure 4.24: Study of convergence of the ADMM, using RD (Fig.4.24a) or DD (Fig.4.24b) as forward projectors combined to all six backprojectors considered in this study.



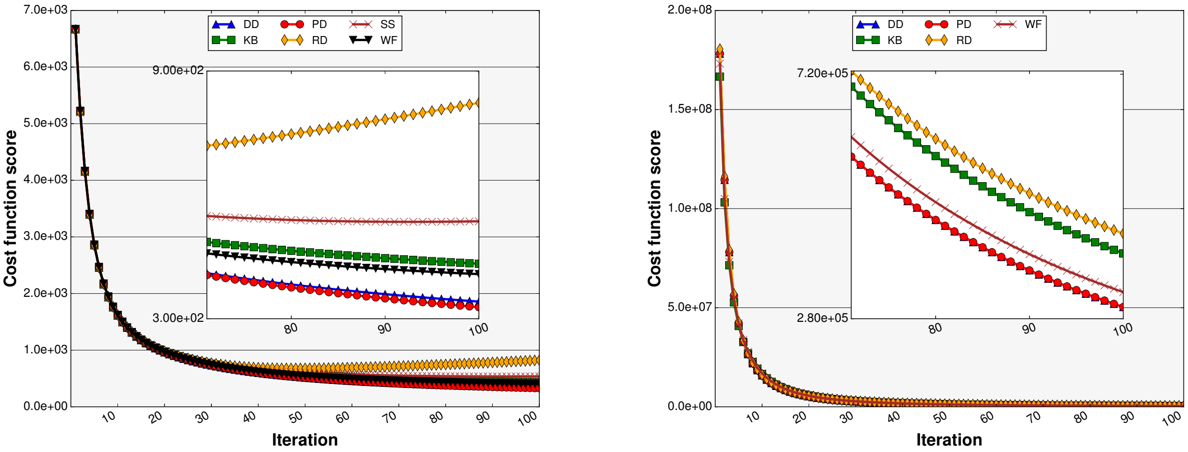
(a) PWLS $-\mathcal{R}=$WF

(b) PWLS $-\mathcal{R}=$SS

Figure 4.25: Study of convergence of the PWLS, using WF (Fig.4.25a) or SS (Fig.4.25b) as forward projectors combined to all six backprojectors considered in this study.

reconstructing an underconstrained dataset with uncoupled operators (Fig.4.29b).

Reconstructions with MLEM and SIRT are very sensitive to the coupling effect with both undersampled



(a) MLEM $-\mathcal{R}$=PD

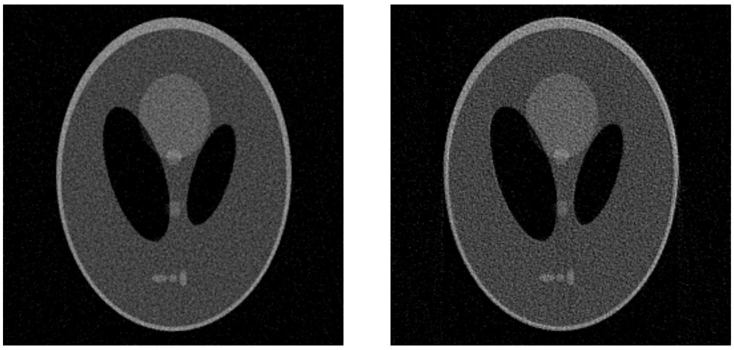
(b) SIRT $-\mathcal{R}$=PD

Figure 4.26: Study of convergence of the MLEM and SIRT, both using PD as forward projector combined to all six backprojectors considered in this study.

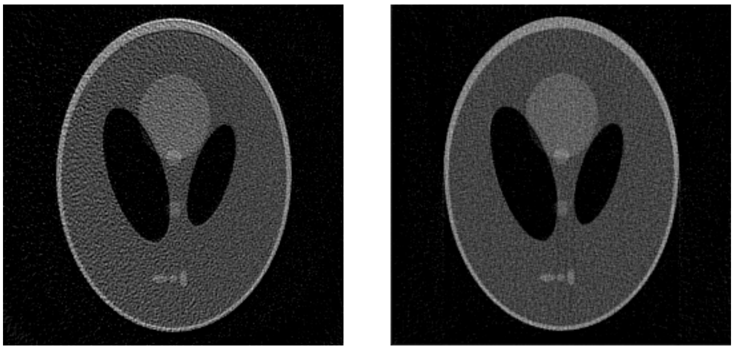
and noisy datasets. Several reconstruction attempts for SL-UNDER, SL-NOISE and SL-UCONSTR using these algorithms with non-matching operators failed, as the procedure quickly diverges after few iterations. Figure 4.30 shows an experiment with MLEM, PD forward projector and PD, RD and KB backprojectors: the reconstruction with coupled operators (Fig.4.30a) is once again characterized by the highest accuracyThe last experiment is designed to roughly estimate the impact of the coupling projector-backprojector on the reconstruction quality with respect to other two fundamental components: physical constraints (i.e., setting to zero all negative pixels at each iteration) and optimal number of iterations. As example we show here the results for SL-UCONSTR and the ADMM. The highest PSNR in Tab.20 corresponds to case (1), where all three components (coupling, constraints, optimal number of iterations) are present. The interesting result is that case (2), that relies only on coupled operators, achieves a better reconstruction quality than case (3), where constraints and optimal number of iterations are kept, but the operators are not matching. This experiment gives a hint of the fact that, in some cases, the coupling projector-backprojector could even play a more decisive role than other crucial factors on the accuracy of an iterative algorithm. To validate the generality of these last results, further in-depth analysis is required, subject of future work.

Table 18: PSNR scores of ADMM reconstructions of SL-UNDER using PD as forward projector and PD, KB, RD, WF as backprojectors.

	R*=PD	R*=KB	R*=RD	R*=WF
PSNR	22.06	21.67	21.29	21.46



(a) R^*=SS – PSNR=20.13 (b) R^*=DD – PSNR=17.15



(c) R^*=KB – PSNR=16.50 (d) R^*=RD – PSNR=18.51

Figure 4.27: ADMM reconstructions of SL-NOISE using SS as forward projector and SS, DD, KB, RD as backprojectors.

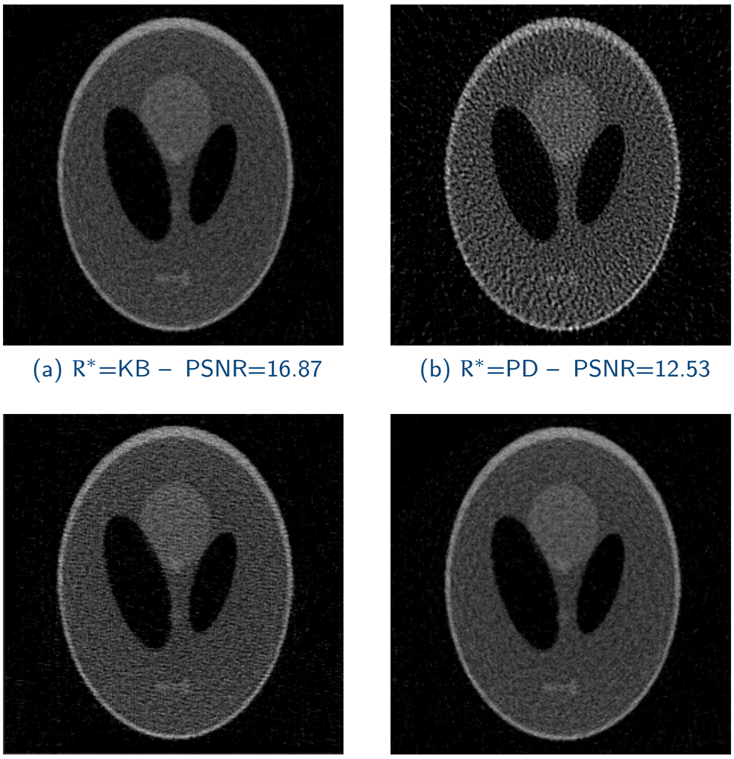
Table 20: Three different ADMM reconstructions of SL-UCONSTR. Case (1): coupled operators + constraints + optimal number of iterations. Case (2): coupled operators. Case (3): constraints + optimal number of iterations.

	Case 1	Case 2	Case 3
PSNR	19.69	18.97	18.10

Table 19: PSNR scores of PSWS reconstructions of SL-UNDER using PD as forward projector (left) and SL-NOISE using SS as forward projector (right).

	R^*=KB	R^*=DD	R^*=SS	R^*=WF
PSNR	22.09	19.15	19.24	21.50

	R^*=SS	R^*=DD	R^*=WF	R^*=PD
PSNR	23.04	22.51	22.39	22.47



(a) R*=KB – PSNR=16.87 (b) R*=PD – PSNR=12.53

(c) R*=RD – PSNR=14.17 (d) R*=WF – PSNR=16.86

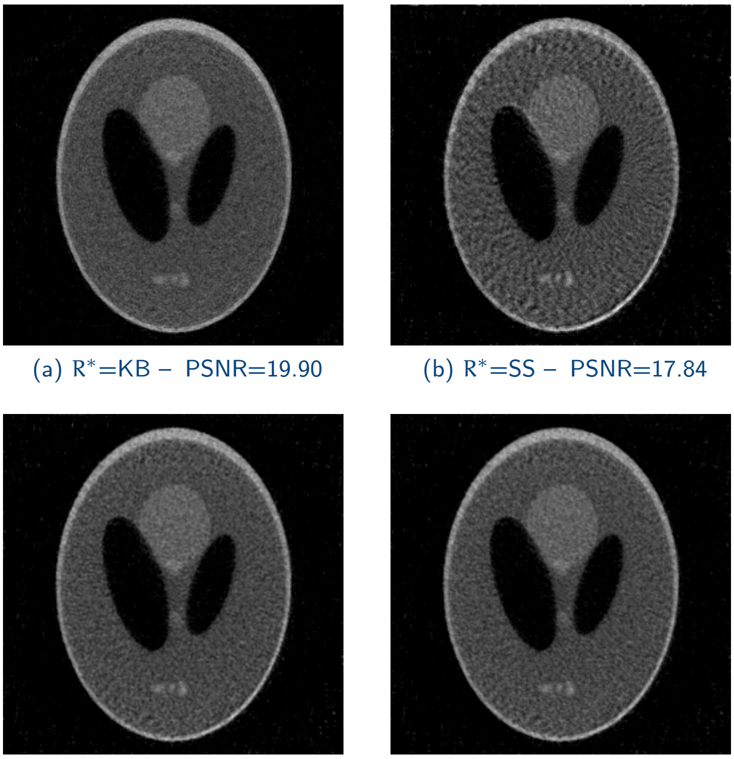
Figure 4.28: ADMM reconstructions of SL-UCONSTR using KB as forward projector and KB, PD, RD, WF as backprojectors.

4.5.6 Conclusion

This work is an experimental study on the impact of the coupling projector-backprojector in iterative reconstruction schemes. Since iterative algorithms call the tomographic operators few times per iteration, it can be expected that the level of matching between the actual implementation of the forward projector and backprojector can deeply affect the performance of the entire iterative procedure.

A framework consisting of four iterative methods (the alternate direction method of multipliers, the penalized weighted least squares, the maximum-likelihood expectation maximization and the simultaneous iterative algebraic technique) working with six different projectors (distance-driven, pixel-driven, ray-driven, slant-stacking and two gridding methods) has been conceived to test the aforementioned hypothesis.

All iterative experiments on simulated data clearly show that the performance of every selected method is deeply affected by the coupling projector-backprojector in terms of convergence and accuracy. The best convergence behaviour and the highest reconstruction quality are systematically obtained when the tomographic operators match. This conclusion holds regardless of the nature of the input tomographic dataset in terms of angular sampling or SNR. Moreover, there is indication that the coupling projector-backprojector may represent one of the major players determining the performance of an



(a) R*=KB – PSNR=19.90 (b) R*=SS – PSNR=17.84

(c) R*=DD – PSNR=19.61 (d) R*=PD – PSNR=19.62

Figure 4.29: PWLS reconstructions of SL-UCONSTR using KB as forward projector and KB, SS, DD, PD as backprojectors.

(a) R*=PD – PSNR=20.74 (b) R*=RD – PSNR=10.63 (c) R*=KB – PSNR=19.42

Figure 4.30: MLEM reconstructions of SL-UNDER using PD as forward projector and PD, RD, KB as backprojectors.

iterative algorithm, even with respect to physical constraints or optimal number of iterations.

The results of this study indicate that it would be strongly advisable for users and developers of software packages for iterative tomographic reconstructions to always select projector pairs with a high mathematical affinity and to carefully assess and validate the degree of coupling of the used implementations. This strategy is important to avoid results systematically characterized by suboptimal accuracy.

Chapter 5

ADMM for CT reconstruction

This chapter focuses on a numerical optimization technique able to solve a large set of problems and that only recently has been applied to tomographic reconstruction: the *alternate direction method of multipliers*, abbreviated as *ADMM*.

The ADMM has the upside of being extremely versatile when it comes to incorporating different forward models, regularization schemes and constraints and yields reconstructions of high quality after a relatively small number of iterations.

We discuss the usage of the ADMM to tackle the reconstruction of underconstrained tomographic datasets. Its accuracy, convergence and computational efficiency are here experimentally characterized.

5.1 Alternate direction method of multipliers

5.1.1 Mathematical formulation

The ADMM solves numerical optimization problems of the form [27]:

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) + g(\mathbf{z}) \\ &\text{subject to} \quad \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \quad , \end{aligned} \tag{5.1}$$

where $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g(\mathbf{z}) : \mathbb{R}^m \longrightarrow \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{p \times n}$ and $\mathbf{B} \in \mathbb{R}^{p \times m}$. The second line of Eq. (5.1) is called "equality constraint" or simply "constraint". The functions f and g are assumed to be convex, i.e. [221]:

$$f\left(\beta\mathbf{x} + (1-\beta)\mathbf{y}\right) \leqslant \beta f(\mathbf{x}) + (1-\beta) f(\mathbf{y}) \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad \forall\, \beta \in \mathbb{R} \quad , \tag{5.2}$$

and equivalently for g. First, the ADMM creates the augmented Lagrangian corresponding to problem (5.1):

$$\mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) = f(\mathbf{x}) + g(\mathbf{z}) + \boldsymbol{\alpha}^T \left(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\right) + \frac{\mu}{2} \left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\right\|_2^2 \quad , \tag{5.3}$$

with $\mu > 0$. $\boldsymbol{\alpha}$ is the vector of Lagrangian multipliers also called *dual variable*. The initial constrained optimization problem (5.1) is transformed into an unconstrained problem:

$$\underset{\mathbf{x},\mathbf{z},\boldsymbol{\alpha}}{\operatorname{argmin}}\ \mathcal{L}_\mu(\mathbf{x},\mathbf{z},\boldsymbol{\alpha}) \quad . \tag{5.4}$$

The solution to (5.4) is, then, iteratively approximated with the following sequence of steps [27]:

$$\begin{aligned}
\mathbf{x}^{(k+1)} &= \underset{\mathbf{x}}{\operatorname{argmin}}\,\mathcal{L}_\mu\left(\mathbf{x},\mathbf{z}^{(k)},\boldsymbol{\alpha}^{(k)}\right)\\
\mathbf{z}^{(k+1)} &= \underset{\mathbf{z}}{\operatorname{argmin}}\,\mathcal{L}_\mu\left(\mathbf{x}^{(k+1)},\mathbf{z},\boldsymbol{\alpha}^{(k)}\right)\\
\boldsymbol{\alpha}^{(k+1)} &= \boldsymbol{\alpha}^{(k)} + \mu\left(\mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{z}^{(k+1)} - \mathbf{c}\right) \quad .
\end{aligned} \tag{5.5}$$

The following pattern characterizes the ADMM: incorporation of the equality constraint into $\mathcal{L}_\mu$; split of the minimization of $\mathcal{L}_\mu$ into partial minimizations with respect to $\mathbf{x}$ and $\mathbf{z}$; computation of the solution through a sequential update of the three steps in (5.5). The ADMM works also when the solutions of the $\mathbf{x}$- and $\mathbf{z}$-subproblem are roughly approximated. The roles played by the variables $\mathbf{x}$ and $\mathbf{z}$ are not entirely symmetric, meaning that in general the order of appearance of the subproblems in (5.5) matters. The Lagrangian can be also expressed in the *scaled form* by introducing the primal residual variable $\mathbf{r} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}$. The second and third term in Eq. (5.3) become [27]:

$$\begin{aligned}
\boldsymbol{\alpha}^T\left(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\right) + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\|_2^2 = \boldsymbol{\alpha}^T\mathbf{r} + \frac{\mu}{2}\|\mathbf{r}\|_2^2 &= \frac{\mu}{2}\left\|\mathbf{r} + \frac{1}{\mu}\boldsymbol{\alpha}\right\|_2^2 - \frac{1}{2\mu}\|\boldsymbol{\alpha}\|_2^2\\
&= \frac{\mu}{2}\|\mathbf{r} + \mathbf{u}\|_2^2 - \frac{\mu}{2}\|\mathbf{u}\|_2^2 \quad ,
\end{aligned} \tag{5.6}$$

where $\mathbf{u} = \boldsymbol{\alpha}/\mu$ is the scaled dual variable. By introducing $\mathbf{u}$ in Eq. (5.5), the ADMM steps become [27]:

$$\begin{aligned}
\mathbf{x}^{(k+1)} &= \underset{\mathbf{x}}{\operatorname{argmin}}\left(f(\mathbf{x}) + \frac{\mu}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^{(k)} - \mathbf{c} + \mathbf{u}^{(k)}\right\|_2^2\right)\\
\mathbf{z}^{(k+1)} &= \underset{\mathbf{z}}{\operatorname{argmin}}\left(g(\mathbf{z}) + \frac{\mu}{2}\left\|\mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{z} - \mathbf{c} + \mathbf{u}^{(k)}\right\|_2^2\right)\\
\mathbf{u}^{(k+1)} &= \mathbf{u}^{(k)} + \mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{z}^{(k+1)} - \mathbf{c} \quad .
\end{aligned} \tag{5.7}$$

This second form of the ADMM allows to express the scaled dual variable at an iteration k as the sum of all residuals until that iteration, i.e., $\mathbf{u}^{(k)} = \mathbf{u}^{(0)} + \sum_{j=0}^{k}\mathbf{r}^{(j)}$.

The ADMM converges if both the $\mathbf{x}$- and $\mathbf{z}$-subproblem admit a solution and if the non-augmented Lagrangian, $\mathcal{L}_{\mu=0}$, has a saddle point, i.e. $\exists\,(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\alpha}^*)$ such that $\mathcal{L}_0(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\alpha}) \leqslant \mathcal{L}_0(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\alpha}^*) \leqslant \mathcal{L}_0(\mathbf{x},\mathbf{z},\boldsymbol{\alpha}^*)\ \forall\,\mathbf{x},\mathbf{z},\boldsymbol{\alpha}$ [27]. If these assumptions hold, it follows that [27]:

- the residual converges: $\lim\limits_{k\to\infty}\|\mathbf{r}^{(k)}\|_2^2 = 0$;

- the objective converges: $\lim_{k\to\infty}\left(f(\mathbf{x}^{(k)})+g(\mathbf{z}^{(k)})\right)=p^*:=\inf\left\{f(\mathbf{x})+g(\mathbf{z})\,\middle|\,\mathbf{Ax}+\mathbf{Bz}=\mathbf{c}\right\}$;
- the dual variable converges: $\lim_{k\to\infty}\|\boldsymbol{\alpha}^{(k)}\|_2^2=\boldsymbol{\alpha}^*$.

Experimental studies [27, 222] have shown that the ADMM is suited to optimization problems, that do not require solutions of high accuracy. Generally speaking, if a rough solution is satisfactory for the problem under study, the ADMM can converge in few tens of iterations, otherwise a large unknown number of iterations may be needed to improve accuracy. For CT reconstruction, the "rough" approximated solution provided by the ADMM results comparable and sometimes even improved at visual inspection with respect to the reconstructions obtained with other iterative algorithms, as the experiments presented in 5.2 show.
A practical stopping criterion for the ADMM requires two distinct ϵ-values [27, 223]: a ϵ_{pri} for the norm of the primal residual, $\|\mathbf{r}^{(k)}\|_2$; a ϵ_{dual} for the norm of the dual residual, $\|\mathbf{s}^{(k)}\|_2$, defined as $s^{(k+1)}=\mu\mathbf{A}^T\mathbf{B}(\mathbf{z}^{(k+1)}-\mathbf{z}^{(k)})$.

5.1.2 General patterns for CT reconstruction

Problem (5.1) for CT image reconstruction simplifies to:

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{Ax}-\mathbf{c}\|_2^2+\lambda R(\mathbf{x}) \tag{5.8}$$

where $\mathbf{A}=\mathcal{R}$, $\mathbf{x}$ is the unknown tomographic slice and $\mathbf{c}$ is the input sinogram. The first member of problem (5.8) is also called *fidelity term* and is differentiable with respect to $\mathbf{x}$ for any $\mathbf{x}$ (differently from the non-squared term $\|\mathbf{Ax}-\mathbf{c}\|_2$). $R(\mathbf{x})$ is a function incorporating the a priori knowledge available for the object $\mathbf{x}$ and it is also called *regularization term*. $R(\mathbf{x})$ enforces, e.g., a specific condition on the p-norm of $\mathbf{x}$ or promotes sparsity in a certain domain.
Consider now $R(\mathbf{x})$ of the form $R(\mathbf{x})=\|\mathbf{Gx}\|_p$, where $\mathbf{G}$ represents the gradient operator.
A strategy to decouple the fidelity and regularization terms is to introduce an auxiliary variable $\mathbf{z}$ [103]:

$$\begin{aligned}&\text{minimize} && \frac{1}{2}\|\mathbf{Ax}-\mathbf{c}\|_2^2+\lambda\|\mathbf{z}\|_p\\ &\text{subject to} && \mathbf{Gx}-\mathbf{z}=0 \quad .\end{aligned} \tag{5.9}$$

The augmented Lagrangian corresponding to problem (5.9) is:

$$\mathcal{L}_\mu(\mathbf{x},\mathbf{z},\boldsymbol{\alpha})=\frac{1}{2}\|\mathbf{Ax}-\mathbf{c}\|_2^2+\lambda\|\mathbf{z}\|_p+\boldsymbol{\alpha}^T(\mathbf{Gx}-\mathbf{z})+\frac{\mu}{2}\|\mathbf{Gx}-\mathbf{z}\|_2^2 \quad . \tag{5.10}$$

To obtain closed-form expressions for the subproblems of procedure (5.5), the partial derivatives of $\mathcal{L}_\mu$ with respect to $\mathbf{x}$ and $\mathbf{z}$ must be computed. For the $\mathbf{x}$-subproblem (the $\mathbf{z}$-subproblem is considered in detail in the next subsection):

$$\begin{aligned} 0 &= \frac{\partial}{\partial \mathbf{x}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) = \mathbf{A}^\mathsf{T}(\mathbf{A}\mathbf{x} - \mathbf{c}) + \mathbf{G}^\mathsf{T}\boldsymbol{\alpha} + \mu \mathbf{G}^\mathsf{T}(\mathbf{G}\mathbf{x} - \mathbf{z}) \\ &= \underbrace{(\mathbf{A}^\mathsf{T}\mathbf{A} + \mu \mathbf{G}^\mathsf{T}\mathbf{G})}_{\tilde{\mathbf{A}}}\mathbf{x} - \underbrace{(\mathbf{A}^\mathsf{T}\mathbf{c} - \mathbf{G}^\mathsf{T}\boldsymbol{\alpha} + \mu\mathbf{G}^\mathsf{T}\mathbf{z})}_{\tilde{\mathbf{c}}} = \tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{c}} \quad . \end{aligned} \tag{5.11}$$

Since $\tilde{\mathbf{A}}$ is symmetric and positive definite, the conjugate gradient (CG) technique [224] can be used to compute an approximated solution of the $\mathbf{x}$-subproblem:

$$\mathbf{x}^{(k+1)} = \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}_\mu\left(\mathbf{x}, \mathbf{z}^{(k)}, \boldsymbol{\alpha}^{(k)}\right) = \mathsf{CG}\left\{\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{c}}\right\} \quad . \tag{5.12}$$

The number of iterations required by the CG to obtain a reasonable approximation of $\mathbf{x}^{(k+1)}$ depends on the conditioning number of $\tilde{\mathbf{A}}$, defined as the ratio between the maximum and minimum singular value of $\tilde{\mathbf{A}}$, i.e., $\kappa = \lambda_{\mathsf{max}}(\tilde{\mathbf{A}})/\lambda_{\mathsf{min}}(\tilde{\mathbf{A}})$. The higher κ, the more iterations are needed for CG to achieve a certain reconstruction accuracy. The preconditioned conjugate gradient (PCG) can be used to substantially improve the CG convergence. The preconditioner is a symmetric positive definite matrix used to decrease the conditioning number of the problem to be solved by CG. At the cost of one matrix-vector multiplication before the CG main loop and one inside it, preconditioning allows to reach the minimum of the cost function in less iterations. Ad-hoc strategies for iterative tomographic reconstruction have been proposed, for example, in [225–229].

A slightly different formulation of the ADMM for CT is the so called *plug-and-play (ADMP)* [230]. The plug-and-play formulation simply replaces the equality constraint in Eq.(5.9) with $\mathbf{x} - \mathbf{z} = 0$. The ADMP Lagrangian, therefore, looks like:

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{c}\|_2^2 + \lambda R(\mathbf{z}) + \boldsymbol{\alpha}^\mathsf{T}(\mathbf{x} - \mathbf{z}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \\ &= \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{c}\|_2^2 + \lambda R(\mathbf{z}) + \frac{1}{2\mu}\|\mu(\mathbf{x} - \mathbf{z}) + \boldsymbol{\alpha}\|_2^2 - \frac{1}{2\mu}\|\boldsymbol{\alpha}\|_2^2 \quad , \end{aligned} \tag{5.13}$$

where the last quadratic term can be neglected since:

$$\min_{\mathbf{x},\mathbf{z}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) = \min_{\mathbf{x},\mathbf{z}} \left\{\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{c}\|_2^2 + \lambda R(\mathbf{z}) + \frac{1}{2}\left\|\mathbf{x} - \mathbf{z} + \frac{\boldsymbol{\alpha}}{\mu}\right\|_2^2\right\} \tag{5.14}$$

The first subproblem becomes:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \mathbf{x}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) = \mathbf{A}^\mathsf{T}(\mathbf{A}\mathbf{x} - \mathbf{c}) + \left(\mathbf{x} - \mathbf{z} + \frac{\boldsymbol{\alpha}}{\mu}\right) \\ &= \underbrace{(\mathbf{A}^\mathsf{T}\mathbf{A} + \mu\mathbf{I})}_{\tilde{\mathbf{A}}^{(p)}}\mathbf{x} - \underbrace{\left(\mathbf{A}^\mathsf{T}\mathbf{c} + \mathbf{z} - \frac{\boldsymbol{\alpha}}{\mu}\right)}_{\tilde{\mathbf{c}}^{(p)}} = \tilde{\mathbf{A}}^{(p)}\mathbf{x} - \tilde{\mathbf{c}}^{(p)} \quad . \end{aligned} \tag{5.15}$$

Without considering the partial derivative of $\mathcal{L}_\mu$ with respect to $\mathbf{z}$, the second subproblem becomes:

$$\min_{\mathbf{z}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) = \min_{\mathbf{z}} \left\{ \frac{1}{2} \left\| \mathbf{x} - \mathbf{z} + \frac{\boldsymbol{\alpha}}{\mu} \right\|_2^2 + \lambda R(\mathbf{z}) \right\} = \min_{\mathbf{z}} \left\{ \frac{1}{2} \|\mathbf{z} - \tilde{\mathbf{x}}\|_2^2 + \lambda R(\mathbf{z}) \right\} \quad . \tag{5.16}$$

The last form of Eq.(5.16) explicits the fact that, regardless of the choice of $R(\mathbf{z})$, the $\mathbf{z}$-subproblem corresponds exactly to a denoising problem. The appealing feature of the ADMP is that the fidelity and the regularization terms are effectively decoupled from each other: the $\mathbf{x}$-subproblem does not contain the regularization term differently from Eq. (5.11), where $\mathbf{G}$ is present; the $\mathbf{z}$-subproblem does not contain the matrix $\mathbf{A}$. The "plug-and-play" denomination of the ADMP comes from the fact that any denoising algorithm can be directly used to obtain an approximated solution of the $\mathbf{z}$-subproblem.

5.1.3 Regularization schemes

The reconstruction of underconstrained tomographic datasets represents in general an ill-posed underdetermined problem, due to presence of noise in the data and the fact that the number of unknowns is larger than the number of equations (namely, $\mathbf{A} \in \mathbb{R}^{mn \times n^2}$ and $m < n$). To guarantee a stable inversion of $\mathbf{Ax} = \mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}$ ($\hat{\mathbf{b}}$ indicates the noise-free data, $\mathbf{e}$ the noise component), a regularization scheme must be employed.

The regularization term, $R(\mathbf{x})$, added to $1/2\|\mathbf{Ax} - \mathbf{b}\|_2^2$ can be interpreted in different ways: (i) it prevents the solution from overfitting the data, thus, assimilating the noise affecting $\mathbf{b}$; (ii) from a Bayesian point of view, $R(\mathbf{x})$ imposes a prior distribution on the model parameters; (iii) it represents some a priori knowledge available on $\mathbf{x}$ used to steer the optimization towards an expected result.

Standard and state-of-the-art regularization schemes for iterative tomographic reconstruction can be grouped in three categories: $\mathbb{L}_p$-norm, dictionary-based and non-local priors.

$\mathbb{L}_p$-norm priors

Considered $\mathbf{x} \in \mathbb{R}^n$ and $p \in \mathbb{R} \mid p > 0$, the $\mathbf{L}_p$-norm of $\mathbf{x}$ is defined as:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \quad , \tag{5.17}$$

and for $p = \infty$

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, ..., |x_n|\} \quad . \tag{5.18}$$

$\mathbb{L}_p$-norms with $p \geqslant 1$ (included $p = \infty$) satisfy the following three properties: (i) only the zero vector has zero norm; (ii) the norm of a vector is positive homogeneous with respect to multiplication by a scalar; (iii) the triangle inequality holds. Instead $\mathbb{L}_p$-norms with $0 < p < 1$ are not complete, because the triangle inequality is not satisfied. The case $p = 0$ does not correspond to a real norm and the notation $\|\mathbf{x}\|_0$ represents the number of non-zero components of $\mathbf{x}$ [127]:

$$\|\mathbf{x}\|_0 = \sum_{i=1}^{n} (1 - \delta(x_i)) \quad . \tag{5.19}$$

Each $\mathbb{L}_p$-norm is characterized by a n-th dimensional "ball" with a surface of specific shape. The "balls" of $\mathbb{L}_0, \mathbb{L}_{1/2}, \mathbb{L}_1, \mathbb{L}_2$ for $n = 2$ are depicted in Fig. 5.1 (colored shapes) and correspond to two segments on the axis (green), a concave figure (orange), a square rotated by 45°(purple), a circle (light blue) and a square (blue), respectively. To graphically understand the role played by a $\mathbf{L}_p$-norm term, consider the following problem:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_p \quad \text{such that} \quad \mathbf{x} \in r \, : \, x_2 = x_1 \tan_\theta + q \quad . \tag{5.20}$$

Problem (5.20) is solved by increasing the p-norm "ball" until it intersects the straight line r in one point (Fig. 5.1). For $p = \infty$, the solution to (5.20) is found when the blue square is increased in size until its right top corner touches r; the solution for $p = 2$ corresponds to the tangent point on r of the circle with radius $\rho = q \cos\theta$; when $p \leqslant 1$, the solution coincides with the point $(0, q)$. This simple example anticipates the connection between $\mathbb{L}_p$-norm regularization terms with $p \leqslant 1$ and *compressive sensing*: the solution to (5.20) with $p \leqslant 1$ is sparse on the basis $\{(1,0)^T, (0,1)^T\}$, as $(0, q)$ has only one non-zero component. This can occur for $p > 1$ only in particular cases (e.g., r being vertical or horizontal).

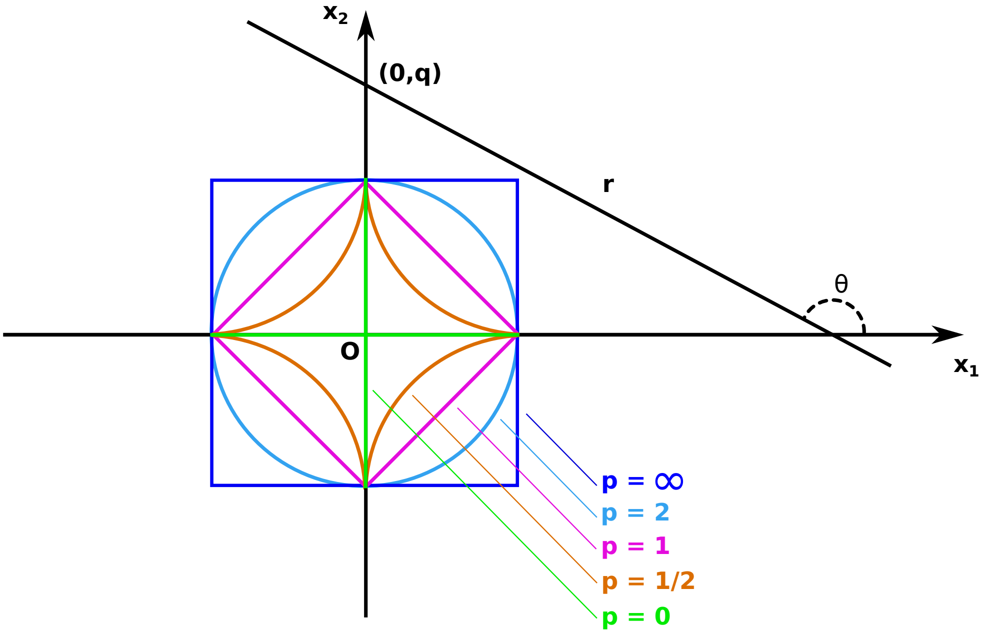


Figure 5.1: $\mathbb{L}_p$ balls for $p = 0, 0.5, 1, 2, \infty$ in the two dimensional case. Fixed a value for p, problem (5.20) is solved by expanding the corresponding ball until it intersects the line r in one point.

Compressive sensing (CS) [127, 231, 232] deals with the recovery of signals that are sparse on a certain frame, even when the number of available samples is not sufficient according to the Nyquist-Shannon theorem. A signal $\mathbf{x} \in \mathbb{R}^n$ is sparse on a frame $\mathbf{D} \in \mathbb{C}^{n \times p}$ with columns $\{\mathbf{d}_1, \mathbf{d}_2, ...\}$ (called *atoms*) and $n \leqslant p$ if:

$$\mathbf{x} = \mathbf{D}\mathbf{c} \quad \mathbf{c} \in \mathbb{R}^p \,\Big|\, \|\mathbf{c}\|_0 = k \text{ and } k \ll p \quad . \tag{5.21}$$

If $p = n$, the frame corresponds to an orthonormal basis and $< \mathbf{d}_i, \mathbf{d}_j >= \delta_{ij}$; for $p > n$, the frame corresponds to an oversampled dictionary, where the first n columns are linearly dependent. The general formulation for CS tomographic reconstruction is:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{c}\|_0 \right\} \quad \text{s.t.} \quad \mathbf{c} = \mathbf{D}^{-1}\mathbf{x} \quad , \tag{5.22}$$

which is equivalent to seek for a reconstruction $\mathbf{x}$ whose forward projection matches $\mathbf{b}$ and has a certain level of sparsity on the frame defined by the atoms $\{\mathbf{d}_1, \mathbf{d}_2, ...\}$. The 0-norm term is non-convex as any other p-norm term with $p < 1$ (Fig. 5.1). Since convex cost functions are easier to tackle in terms of numerical optimization, $\|...\|_0$ can be replaced with $\|...\|_1$, that is convex and yields a solution with reasonable sparsity on the selected frame, as shown by the example of problem (5.20) and Fig. 5.1. When $\|...\|_1$ is used instead of $\|...\|_0$ inside (5.22), the problem is called *least absolute shrinkage and selection operator (LASSO)* [233].

A type of regularization that has found wide-spread successful applications in signal processing and image reconstruction is *total variation (TV)* [128], defined in discretized form as:

$$\|\mathbf{x}\|_{TV} = \|\nabla \mathbf{x}\|_1 = \sum_{i=1}^{n} \Big(|\nabla_{x_1}\mathbf{x}| + |\nabla_{x_2}\mathbf{x}| \Big) \quad , \tag{5.23}$$

which is a LASSO-form of problem (5.22) with $\mathbf{D}^{-1} = \mathbf{G} =$ matrix representation of the gradient operator ∇. A TV term in (5.22) encourages the solution to be piecewise-constant, while preserving discontinuities corresponding to edges. As image denoising method, TV offers a better SNR/spatial resolution trade-off compared to simple filtering schemes (Gaussian blurring, mean filter, ...) [128]. To better understand the idea of TV, let assume that $\tilde{\mathbf{x}}$ is the computed solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$, with $\mathbf{b}$ affected by white noise. Figure 5.2 shows that $\tilde{\mathbf{x}}$ can be thought as $\tilde{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{e}$, where $\bar{\mathbf{x}}$ is the noiseless target object and $\mathbf{e}$ the noise component backprojected from the input data. A large number of natural images and specimens studied with CT can be roughly considered piecewise constant, as they feature high level of sparsity in the gradient domain. The gradient of a white noise image, instead, is not sparse. Therefore, TV steers the solution of the reconstruction problem towards a signal sparse in the gradient domain, which should correspond to the piecewise constant target object, and suppresses all components not characterized by such property. In the ADMM scheme of (5.10), $\mathbf{G}$ is the matrix representation of ∇ and the $\mathbf{z}$-subproblem becomes:

$$0 = \frac{\partial}{\partial \mathbf{z}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) = \lambda \frac{d}{d\mathbf{z}} \|\mathbf{z}\|_1 - \boldsymbol{\alpha} + \mu\,(\mathbf{z} - \mathbf{G}\mathbf{x}) \quad , \tag{5.24}$$

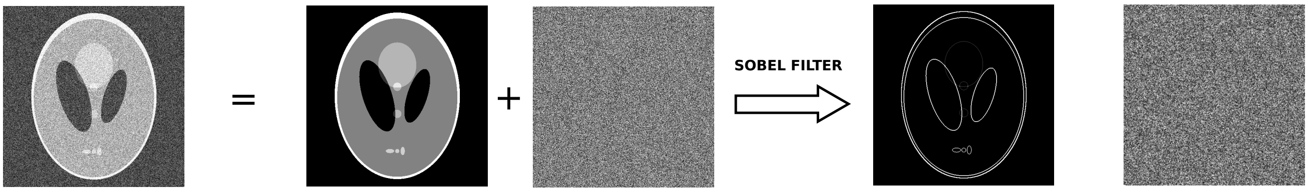
which leads to the shrinkage function [234, 235]:

$$\mathbf{z}^{(k+1)} = \max\left\{\left|\mathbf{G}\mathbf{x}^{(k+1)} + \frac{\alpha^{(k)}}{\mu}\right| - \frac{\lambda}{\mu}\,,\, 0\right\} \operatorname{sgn}\left(\mathbf{G}\mathbf{x}^{(k+1)} + \frac{\alpha^{(k)}}{\mu}\right) \quad . \tag{5.25}$$

The $\mathbf{z}$-subproblem of the ADMM with TV regularization can be, therefore, computed very efficiently. The $\mathbf{z}$-subproblem of ADMP with TV regularization can be solved, instead, by the split-Bregman [112] or Chambolle [236] TV denoising method, that can be more effective than (5.25) but at a higher computational cost.
TV suits the reconstruction of piecewise constant images, but not images with smooth regions (e.g. ramps): staircase artifacts are generated by TV in the attempt of transforming smooth regions into piecewise constant ones. Regularization schemes called *generalized* or *higher order total variation (GTV or HOTV)* [237–240] have been proposed to address the regularization/denoising of smooth objects. These methods make use of higher order derivatives of the TV argument. The GTV implementation discussed in [239] was implemented inside the ADMP. The FFT-based approach for GTV introduced by [241] is computationally efficient when limited to 2D and is tested inside the ADMP in 5.2.7. GTV solves in general the following problem [240, 241]:

$$\begin{gathered}\underset{\mathbf{x}}{\operatorname{argmin}}\left\{\frac{1}{2}\|\mathbf{x}-\tilde{\mathbf{x}}\|_2^2 + \|\mathbf{x}\|_{\mathrm{gtv}}^{\lambda 1,\lambda 2}\right\} = \underset{\mathbf{x}}{\operatorname{argmin}}\left\{\frac{1}{2}\|\mathbf{x}-\tilde{\mathbf{x}}\|_2^2 + \lambda_1\|\mathbf{G}\mathbf{x}\|_{1,2}^{(2,n)} + \lambda_2\|\mathbf{H}\mathbf{x}\|_{1,2,3}^{(3,n)}\right\} \\ \|\mathbf{s}\|_{1,2}^{n_1,n_2} := \sum_{i=0}^{n_1-1}\left(\sum_{j=0}^{n_2-1} s_{i+jn_1}^2\right)^{1/2} \quad ,\end{gathered} \tag{5.26}$$

where $\tilde{\mathbf{x}}$ is the noisy input image, $\mathbf{G}$ and $\mathbf{H}$ are the gradient and Hessian operator, respectively. The FFT-based algorithm for GTV introduced by [241] features as input parameters λ_1, λ_2 and a stopping criterion (chosen to be, once again, a fixed number of iteration). For the regularization parameters of the ADMP, we set $\lambda = \lambda_1 = c \cdot \lambda_2$, $c \in \mathbb{R}_0$.



Noisy reconstruction Signal component Noise component Gradient of the signal component Gradient of the noise component

Figure 5.2: A noisy reconstruction can be considered as the sum of the target object plus a noise component. The target object is piecewise constant and, therefore, has very sparse gradient. The noise component is, instead, dense in the gradient domain. TV can steer the optimization towards a solution featuring sparse gradient, while suppressing the noise component.

Dictionary-based regularization

The *orthogonal matching pursuit (OMP)* [242] is a greedy algorithm that can directly solve problem (5.22), without replacing $\|...\|_0$ with $\|...\|_1$. In particular, the OMP can solve problems of the form:

$$\begin{aligned} &\underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{c}\|_0 \quad \text{subject to} \quad \|\mathbf{Dc}-\mathbf{x}\|_2^2 \leqslant \epsilon \\ &\underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{Dc}-\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\mathbf{c}\|_0 \leqslant k \quad . \end{aligned} \tag{5.27}$$

The OMP can be used with an oversampled dictionary $\mathbf{D}$, that is either precomputed, i.e. constructed with a mathematical formula, or learnt through appropriate training on a specific dataset.

Creating an oversampled dictionary for a real-size image would require long computations and non-negligible memory capacity to finally store a matrix, where each column has the same size of the input image. To keep $\mathbf{D}$ small in size, the common strategy is to divide the real-size image $\mathbf{X} \in \mathbb{R}^{n^2}$ into overlapping patches $\mathbf{x}_i$ with $\sqrt{n} \times \sqrt{n}$ pixels [243–245] and to use $\mathbf{x}_i$ as inputs to either compute the dictionary or to solve one of (5.27) through the OMP. Moreover, working with overlapping patches prevents the appearance of block artifacts, that would occur with non-overlapping patches, and binds the dictionary-based regularization to be local [246].

The OMP solves (5.27) by computing a series of inner products between the overlapping patches of the image and the columns of the dictionary. For the second problem of (5.27) and for a given patch $\mathbf{x}_i$, e.g., the OMP retains the columns of $\mathbf{D}$, $\{\mathbf{d}_{i_1}, \mathbf{d}_{i_2}, ..., \mathbf{d}_{i_k}\}$, yielding the k highest inner products. The selected patch is, then, expressed as a linear combination of such columns:

$$\mathbf{x}_i = \mathbf{Dc} = c_1\mathbf{d}_{i_1} + c_2\mathbf{d}_{i_2} + ... + c_k\mathbf{d}_{i_k} \qquad \text{s.t.} \qquad c_j = \langle \mathbf{x}_i, \mathbf{d}_{i_j} \rangle \quad . \tag{5.28}$$

A precomputed dictionary performing well for image denoising is the one based on the *discrete cosine transform (DCT)* [246, 247]. Two established methods to create trained dictionaries for image denoising/reconstruction are the *principal component analysis (PCA)* [248] and the *K-SVD algorithm* [246].

Figure 5.3 shows how dictionary-based denoising works. First, a precomputed or learnt dictionary is created, then, the overlapping patches of the noisy image are used as input for the OMP algorithm to solve one of the problems in (5.27). The denoised image is constructed by replacing each original patch with a linear combination of the "most representative" atoms of $\mathbf{D}$ as explained in (5.28). In theory, a learnt dictionary offers a more sparse representation than a precomputed one for images similar to those that were used for the training. As example, let us consider two SRXTM datasets of distinct samples of the same kind. If the dictionary is learnt on the tomographic reconstructions of the first dataset, the reconstructions of the second dataset will result more sparse on this dictionary than on a precomputed one. When a training dataset is not available or is not reliable enough, a precomputed dictionary is a viable option, as it will yield a reasonably good sparse representation. How to create a learnt dictionary than performs better than a precomputed one for a specific target dataset is still an open issue.

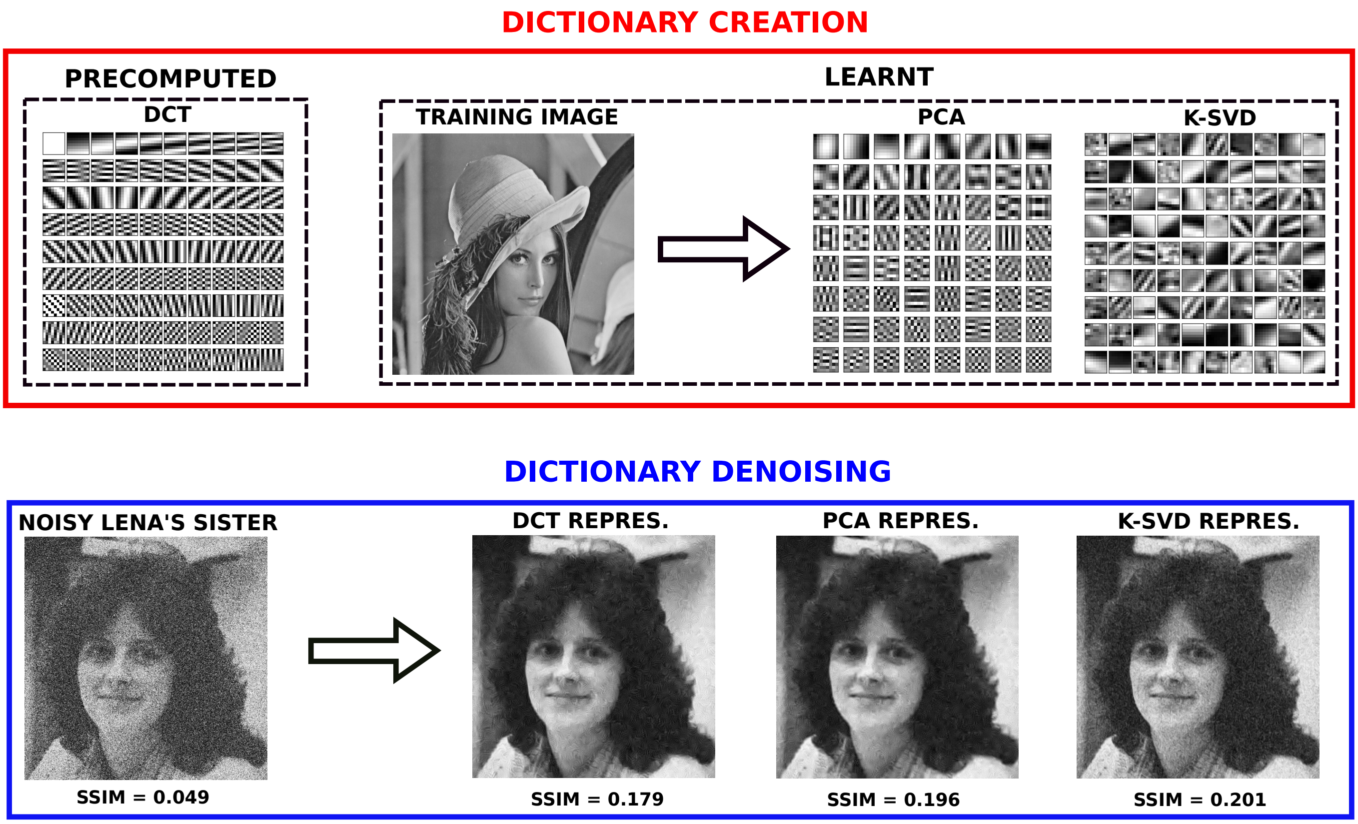


Figure 5.3: Procedure to perform dictionary-based denoising. First, a precomputed or learnt/trained dictionary is created (red box). Second, the overlapping patches of the noisy image are used as input for the OMP algorithm to solve one of the problems in (5.27). Each individual noisy patch is replaced with a linear combination of the "most representative" dictionary atoms, as shown in (5.28). In case a learnt dictionary is used, the training and target images should be similar to a certain extent.

Non-local priors

The denoising methods described so far, scan the neighborhood of each pixel and perform some type of local averaging. The *non-local means (NLM)* algorithm [249, 250] removes the constraint of locality, such that a pixel can be averaged with any similar pixel in a preselected window of search. Given a noisy image $\mathbf{x} \in \mathbb{R}^n$, the NLM denoised version $\mathbf{x}^{\text{nlm}}$ is computed as [250]:

$$\begin{gathered} x_i^{\text{nlm}} = \sum_{j \in \Omega_s} \omega_{ij} x_j \\ 0 \leqslant \omega_{ij} \leqslant 1 \quad , \quad \sum_j \omega_{ij} = 1 \quad , \quad \omega_{ij} \propto \exp\left(-(\mathbf{x}_{\Omega_i} - \mathbf{x}_{\Omega_j})^T \mathbf{W} (\mathbf{x}_{\Omega_i} - \mathbf{x}_{\Omega_j})\right) \quad , \end{gathered} \tag{5.29}$$

where Ω_s is the window of search, defining the level of non-locality, ω_{ij} is the weight encoding the similarity between pixel i and pixel j that, by construction, has an exponential dependence on the $\mathbf{W}$-weighted square difference between the neighborhood Ω_i of pixel i and Ω_j of pixel j. The mechanism behind the NLM is depicted in Fig. 5.4. The drawback of the NLM is that it cannot be utilized in the context of energy minimization, thus, it cannot be generalized to other cases than denoising.

The *non-local total variation (NLTV)* [251] is based on the same idea as NLM, but offers, in addition, a tool that can be used for variational regularization theory. The convex NLTV functional of an image $\mathbf{x} \in \mathbb{R}^n$ is [251]:

$$R(\mathbf{x}) = \sum_{i=0}^{n} |\nabla_i^{\omega}\mathbf{x}| \quad , \quad (\nabla_i^{\omega}\mathbf{x})_j := \sqrt{\omega_{ij}}\,(x_j - x_i) \quad , \quad \omega_{ij} = \omega_{ji} \quad , \tag{5.30}$$

where ∇_i^{ω} is the non-local gradient operator characterized by positive, symmetric weights $\omega \in [0,1]$ that correspond to those in Eq. (5.29). Since the non-local gradient is computed between a given pixel i with all the n image pixels, $\nabla_i^{\omega}\mathbf{x} \in \mathbb{R}^{n^2}$. The non-local gradient can be also interpreted as the gradient of a graph having the given image pixels for nodes. NLTV has been applied to image denoising [251, 252] and used as regularization scheme for iterative reconstruction algorithms [202, 253–255].

Non-local denoising methods tested inside the ADMP in 5.2.7 are the *non-local means (NLM)* algorithm [249, 250], and the *non-local TV (NLTV)* computed with the split-Bregman method [202]. These two algorithms are computationally efficient when limited to 2D problems. The NLM and NLTV share the following input parameters: the neighborhood size (Ω_i and Ω_j in (5.29)), the search radius (defining Ω_s in (5.29)) and the degree of filtering, h, (embedded inside the weighting matrix

Figure 5.4: Sketch showing the mechanism of the NLM. Given a pixel i with neighborhood Ω_i, all pixels within the window of search Ω_s around pixel i are used to compute the similarity weights of Eq. (5.29). The weights range in [0,1] and their value depends on the weighted square difference between Ω_i and the neighborhood of the investigated pixel. For example, $\mathbf{x}_{\Omega_j}$ is clearly more similar to $\mathbf{x}_{\Omega_i}$ than $\mathbf{x}_{\Omega_k}$, therefore $\omega_{ij} > \omega_{ik}$.

$\mathbf{W}$ in (5.29)), ruling the exponential decay of the weights (ω_{ij} in formula (5.29)). NLTV [202] has a variational formulation and is solved by the ADMM; this regularization scheme requires, therefore, its own ADMM-parameters μ^{nltv}, λ^{nltv} and stopping criterion, that here is chosen to be a fixed number of iterations. If the NLTV is used inside the ADMP, the regularization strength λ (in (5.16)) coincides with λ^{nltv} and all other parameters are chosen according to the object under study. NLM is formulated, instead, as an image filter and, therefore, lacks an input parameter that can be directly connected to the regularization strength of the ADMP. The ADMP with NLM regularization can be used, but, from a mathematical point of view, stops to work as a "formal" ADMM method.

5.2 Characterization of ADMM & ADMP for CT

Individual components of ADMM (ADMP) are here separately studied to understand their impact on convergence and reconstruction accuracy.

An exhaustive experimental study would require a huge dataset comprising different phantoms, levels of noise and undersampling. This study is, instead, aimed at finding general trends characterizing ADMM (ADMP). For this reason, the collection of results, that are here presented, should be considered "indicative" rather than "absolute".

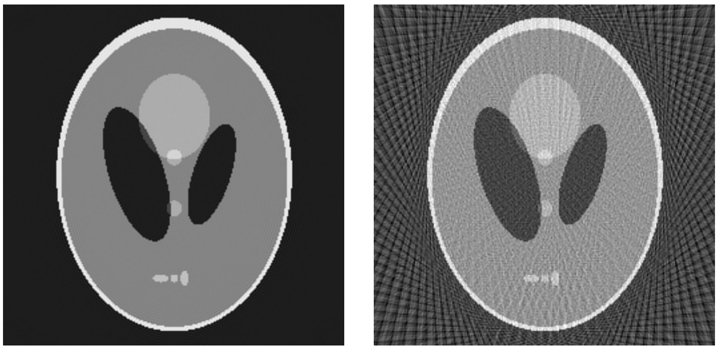
Some parts of this experimental characterization involve comparison of the ADMM to different iterative techniques for CT reconstruction. The iterative methods used for comparison are the separable paraboloidal surrogate used to solve a *penalized weighted least squares (PWLS)* cost function with Huber penalty [17], the *simultaneous iterative reconstruction technique (SIRT)* [13], and the *maximum likelihood expectation-maximization (MLEM)* [110].

When not specified otherwise, the ADMM utilizes gridding projectors with minimal oversampling, TV regularization with $\lambda = \mu = 1.0$, 4 sweeps of CG and physical constraints, i.e., negative pixels are set to zero at each iteration.

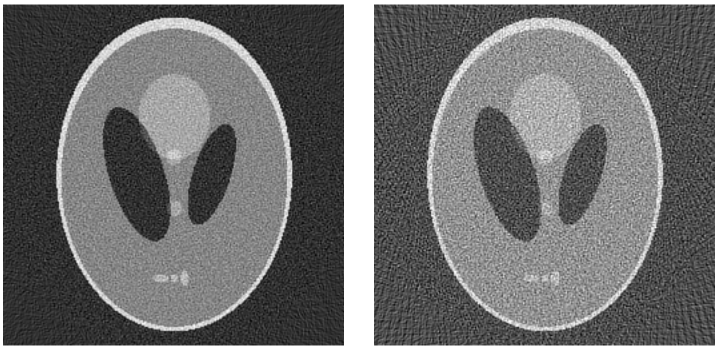
5.2.1 Dataset

To study the general trends of the ADMM (ADMP), the dataset comprises 4 analytical SL sinograms with 256 detector cells. The first sinogram, indicated as "PERF" (which abbreviates "perfect") is noiseless and well-sampled with 402 views. The second sinogram, indicated as "UNDS" (which abbreviates "undersampled") is noiseless with 50 views and simulates a purely undersampled dataset. The third sinogram, indicated as "NOIS" (which abbreviates "noisy") has 402 views and additional Gaussian noise with standard deviation $\sigma = 30\%\mu$, with μ being the mean value of PERF; it simulates a purely noisy dataset. The fourth sinogram, indicated as "UNDC" (which abbreviated "underconstrained"), has 75 views and Gaussian noise with standard deviation $\sigma = 25\%\mu$; it simulates an underconstrained dataset. The FBP reconstructions of the four sinograms with Ram-Lak filter are displayed in Fig. 5.5.

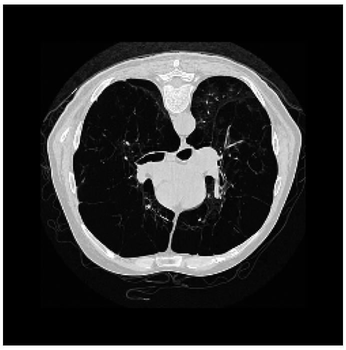
To validate whether our findings are biased by the phantom choice, experiments were repeated with a simulated slice of a human lung (in the 4 modalities of "PERF", "UNDS", "NOIS" and "UNDC") that is displayed in Fig. 5.5(e). In the following text and captions, the SL dataset is always used, unless specified otherwise.



(a) PERF (b) UNDS



(c) NOIS (d) UNDC



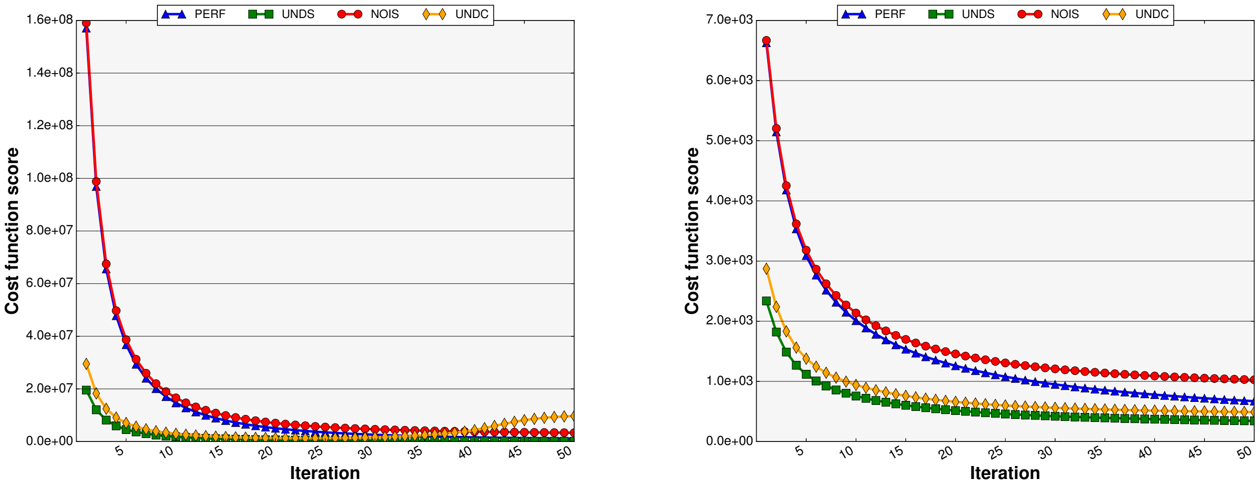
(e) LUNG PHANTOM

Figure 5.5: FBP reconstructions with Ram-Lak filter of the four analytical SL sinograms, PERF, UNDS, NOIS, UNDC and the lung phantom used to validate the findings obtained for the SL dataset, as described in (5.2.1).

5.2.2 Convergence rate, accuracy and spatial resolution

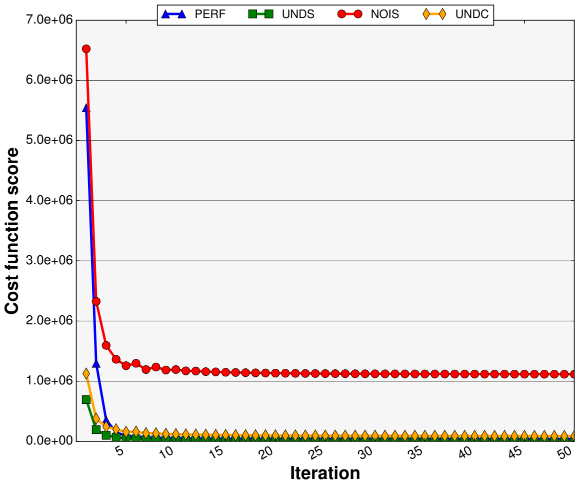
The ADMM is characterized by a very fast convergence rate compared, e.g., to SIRT and PWLS. Figure 5.6 shows that the cost function of the ADMM effectively stops to decrease after around 8-10 iterations for all types of input data. The cost function of SIRT stops to decrease after 20 iterations, when reconstructing UNDS and UNDC, and requires more than 50 iterations to reach a plateau for

PERF and NOIS. PWLS necessitates more than 50 iterations for any kind of input data.
The plots of the PSNR as a function of the number of iterations are displayed in Fig. 5.7. The ADMM reaches the peak of quality for PERF, NOIS and UNDC after around 5 iterations, thus, before the cost function plateau is reached. When reconstructing UNDS, the ADMM saturates the



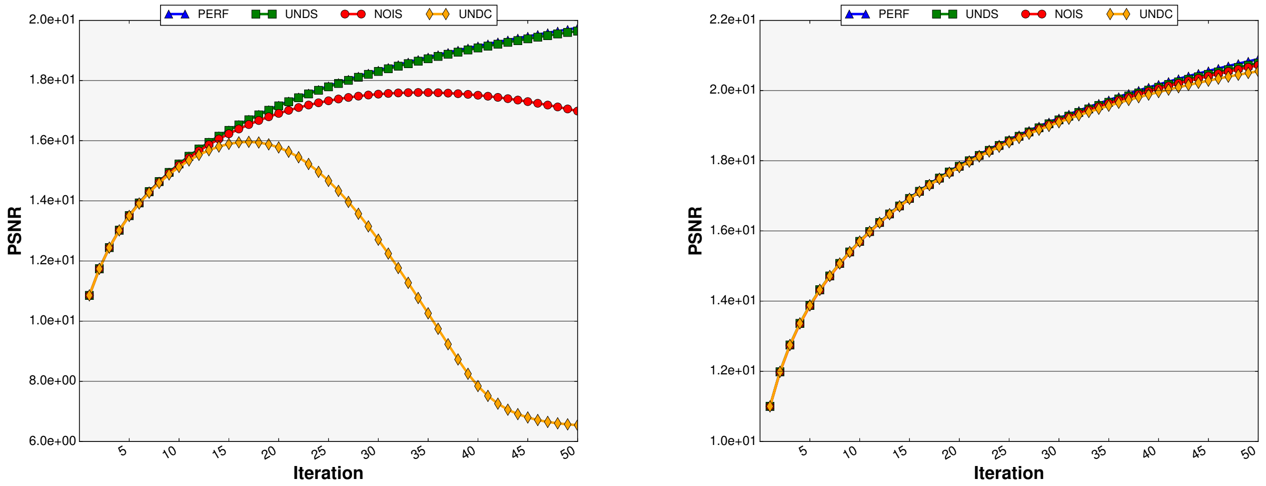
(a) SIRT

(b) PWLS



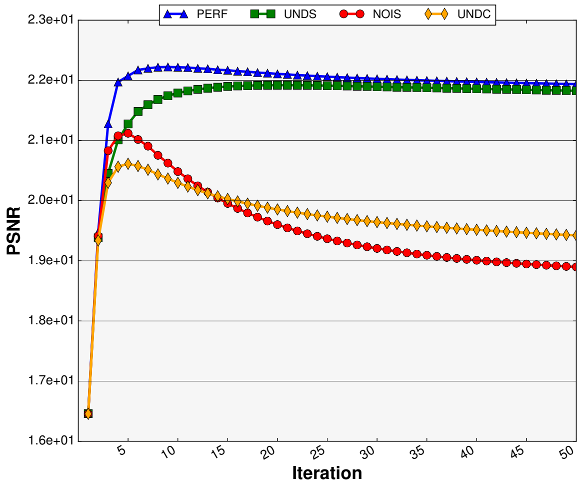
(c) ADMM

Figure 5.6: Convergence plots for SIRT, PWLS and ADMM when tackling the reconstruction of PERF, UNDS, NOIS and UNDC.



(a) SIRT

(b) PWLS



(c) ADMM

Figure 5.7: Reconstruction accuracy as a function of the number of iterations for SIRT, PWLS and ADMM when tackling the reconstruction of PERF, UNDS, NOIS and UNDC.

reconstruction quality few iterations after the cost function plateau. SIRT features a steady increase of the reconstruction quality even after 50 iterations for PERF and UNDS, a peak quality for NOIS after 35-40 iterations and another at 15-20 iterations for UNDC. PWLS is characterized by a monotone increase of the PSNR regardless of the kind of input dataset. Moreover, in the considered range of

50 iterations, the ADMM reaches for each dataset a higher PSNR peak compared to both SIRT and

(a) SIRT (PSNR=19.70) (b) PWLS (PSNR=21.20) (c) ADMM (PSNR=21.37)

Figure 5.8: Reconstructions of UNDC performed by SIRT (15 iterations), PWLS (50 iterations) and ADMM (4 iterations). The red segment drawn on the first reconstruction indicates the position of the line profile used to compare the spatial resolution in Fig. 5.9.

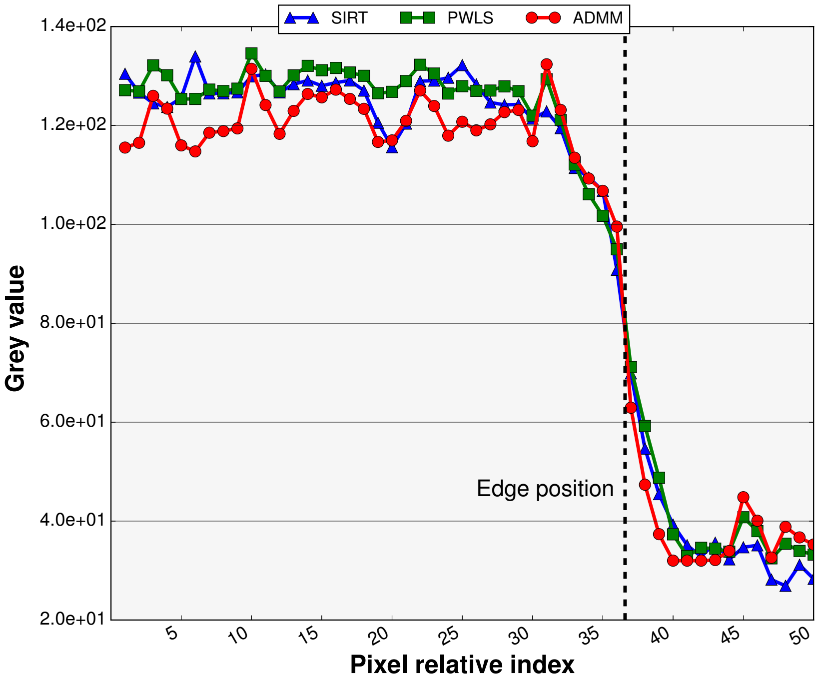


Figure 5.9: Line profiles of the SIRT, PWLS and ADMM reconstructions displayed in Fig. 5.8. The position of the line profile is shown in Fig. 5.8(a).

PWLS.

The ADMM has a great performance in terms of spatial resolution. Consider, for example, the SIRT, PWLS and ADMM reconstructions of UNDC done with 15, 50 and 4 iterations, respectively, in order to pick the best result (PSNR-wise) of each algorithm in the first 50 iterations (Fig. 5.7). The ADMM reconstruction results the sharpest at visual inspection, as also proved by the comparison of edge profiles in Fig. 5.9: the red curve, corresponding to the ADMM edge profile, is slightly above the green and blue curve on the left of the dashed black line, that identifies the edge position, whereas it is clearly underneath the other two curves on the right side.

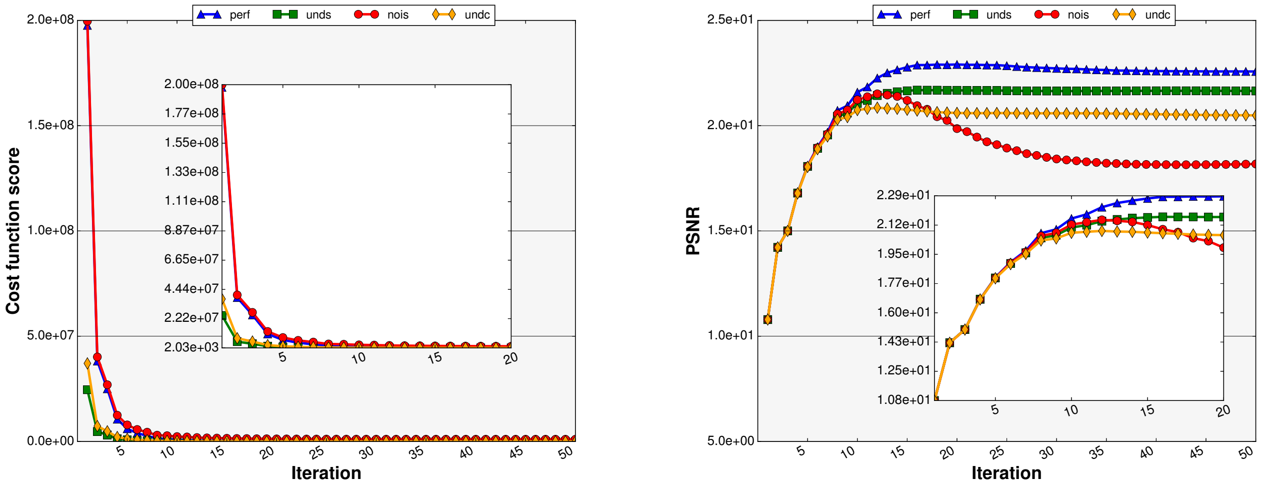
These results are not used to claim that the ADMM is necessarily superior to SIRT and PWLS, but only to show that the algorithm yields satisfactory results at visual inspection with a very little number of iterations. Based on our experience, standard implementations of SIRT generally cannot provide higher reconstruction accuracy than ADMM. PWLS can perform better on simulated data than ADMM if the number of iterations is > 50. Despite this, one has to keep in mind the bias that naturally tends to favor PWLS with respect to ADMM: underconstrained datasets are created with the addition of pure Gaussian noise which matches exactly the formulation of PWLS. Real data acquired at X-ray synchrotron beamline usually do not feature a pure Gaussian noisy component.

5.2.3 Number of CG sub-iterations

The CG plays a fundamental role in the reconstruction process of the ADMM.

When solving a problem of the form $\tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{b}}$ with $\mathbf{A}$ being symmetric and positive definite, the convergence rate of the CG is known to be highly dependent on the conditioning number of $\mathbf{A}$. In the tomographic case, $\mathbf{A} = \mathcal{R}^* \mathcal{R}$ and $\tilde{\mathbf{b}} = \mathcal{R}^* \mathbf{b}$. If the CG alone is used to perform iterative reconstruction, noise (and not undersampling, for instance) can represent a critical source of instability. Figure 5.10 shows that, despite all cost functions decrease with the number of iterations (Fig. 5.10(a)), the PSNR of the NOIS reconstruction reaches a peak around 10-15 iterations and, then, becomes smaller (Fig. 5.10(b)), whereas the PSNR for all other reconstructions saturates approximately after the same number of iterations.

The question, now, is how to choose the optimal number of CG sub-iterations (ncg) for the ADMM. Results in Fig. 5.11 show, unsurprisingly, that the cost function decreases faster when a higher number of CG sub-iterations is utilized, although all curves reach the same plateau after around 12-13 ADMM iterations. A real insight on how to select ncg comes from the plots of Fig. 5.12: figures 5.12(a,c,d) show, in particular, that when the number of iterations is too high, i.e ncg > 6, the PSNR reaches the maximum value after the first ADMM iteration and decreases right after. The problem with this behaviour is that the final reconstruction will be almost exclusively shaped by the action of the first subproblem, whereas the regularization will only marginally contribute. Moreover, stopping the ADMM after 1 iteration means downgrading the algorithm to simple “image processing”, where CG takes care of the reconstruction and, then, denoising follows. This finding is confirmed by the results displayed in Fig. 5.13 of the same experiment conducted on the lung phantom dataset. The optimal number of sub-iterations is considered to be between 4-6, which represents a good compromise between the excessively fast convergence in terms of PSNR when ncg > 6 and the slow convergence provided by ncg=2. In the following, reconstructions are performed with ncg=4. Since ncg=4 is a relatively small number of CG sub-iterations, preconditioning is not really needed to speed up the first subproblem and



(a) (b)

Figure 5.10: Convergence rate and PSNR as a function of the number of iterations for CG used standalone to reconstruct PERF, UNDS, NOIS and UNDC.

performing the additional matrix-vector multiplication required by the PCG could even slow down the first subproblem. Preconditioning is, instead, required when $\tilde{\mathbf{A}} = \mathcal{R}^{(1)*}\,\mathcal{R}^{(1)}$ ($\mathcal{R}^{(1)}$ is the derivative of the Radon transform along the channel direction, i.e. $\mathcal{R}^{(1)}\{f\}(\theta, t) = d\,\mathcal{R}\{f\}(\theta)/dt$), which occurs for the reconstruction of DPC datasets (see 2.2.3). $\tilde{\mathbf{A}} = \mathcal{R}^{(1)*}\,\mathcal{R}^{(1)}$ has a higher conditioning number than $\tilde{\mathbf{A}} = \mathcal{R}^*\,\mathcal{R}$ and the optimal number of CG iterations for the ADMM is around 15 [29]. When $\tilde{\mathbf{A}} = \mathcal{R}^{(1)*}\,\mathcal{R}^{(1)}$, the ADMM can benefit from a preconditioning strategy.

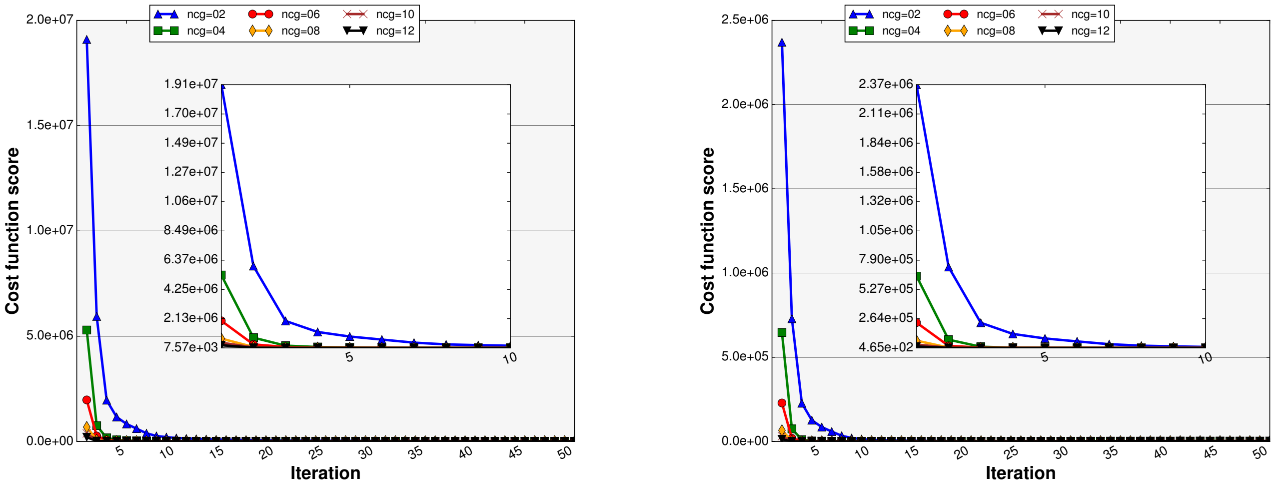
The last experiment compares the classical ADMM with a slightly different version, indicated as ADMM-EXP, where the first subproblem has been slightly modified. The operator $\tilde{\mathbf{A}} = \mathcal{R}^*\,\mathcal{R}$ is applied once outside the main CG loop and twice inside it. Recalling that [92]:

$$\mathcal{R}^*\,\mathcal{R}\{\mathbf{x}\} = 2\pi \cdot (-\Delta)^{-1/2}\{\mathbf{x}\} \quad , \tag{5.31}$$

where $(-\Delta)^{-1/2}$ is the ramp filter operator, one could replace $\mathcal{R}^*\,\mathcal{R}$ in the CG with the right term of Eq. (5.31), that can be efficiently implemented in Fourier space:

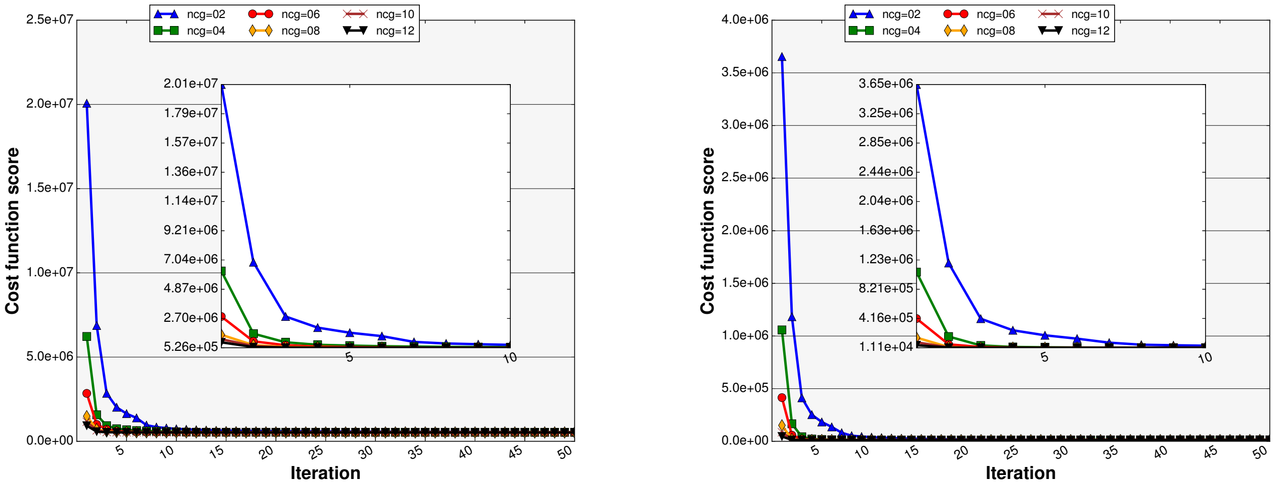
$$(-\Delta)^{-1/2}\{\mathbf{x}\} = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{(-\Delta)^{-1/2}\right\} \cdot \mathcal{F}\{\mathbf{x}\}\right\} = \mathcal{F}^{-1}\left\{\frac{\|\boldsymbol{\omega}\|}{2\pi} \cdot \mathcal{F}\{\mathbf{x}\}\right\} \quad . \tag{5.32}$$

This simple trick allows to substantially speed up the first subproblem, as $\mathcal{R}^*\,\mathcal{R}$ is replaced by 1 FFT-2D and 1 IFFT-2D. Figure 5.14 shows the reconstruction accuracy as a function of the number of iterations for ADMM and ADMM-EXP: for each type of dataset, the ADMM-EXP shows a significantly poorer performance than the ADMM with standard CG. The reason behind this clear decrease of



(a) Reconstruction of PERF

(b) Reconstruction of UNDS

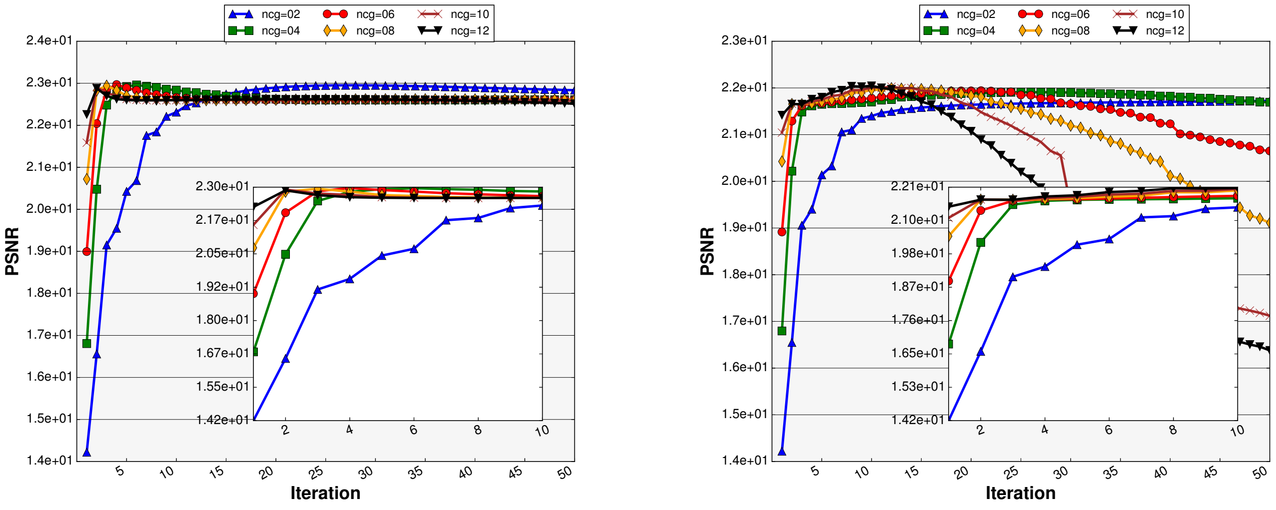


(c) Reconstruction of NOIS

(d) Reconstruction of UNDC

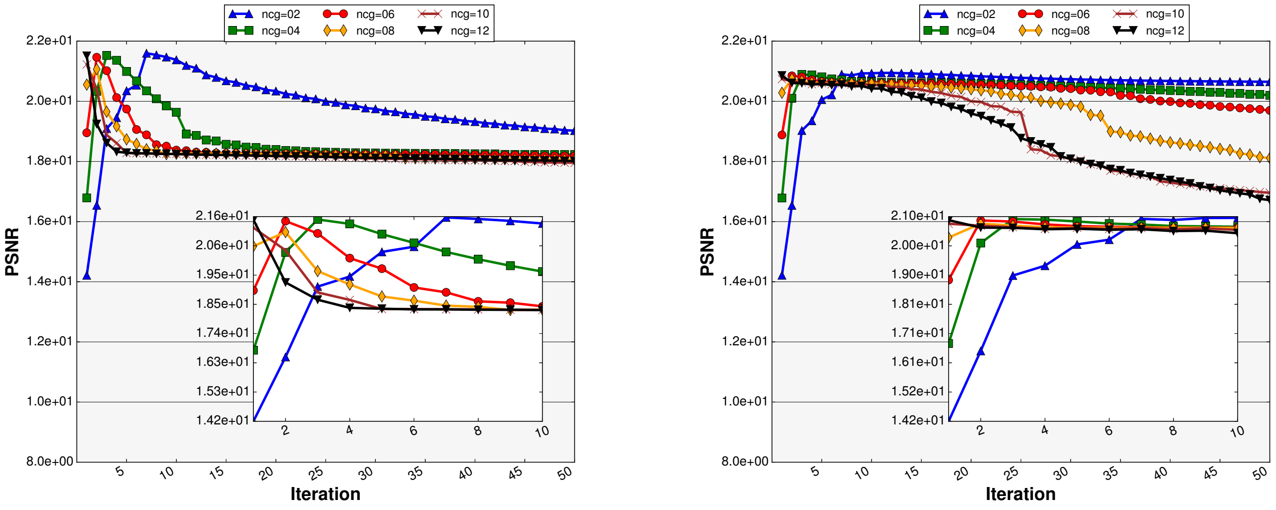
Figure 5.11: Convergence rate of the ADMM for different numbers of CG sub-iterations. Each plot corresponds to the reconstruction of one of the four datasets PERF, UNDS, NOIS and UNDC.

reconstruction quality may arise from a disruption in the coupling of the tomographic operators in the CG: formula (5.32) does not replace every single call of the tomographic projectors, as $\mathcal{R}$ and $\mathcal{R}^*$ need still to be called separately at other points in the CG or ADMM. This creates an evident asymmetry or absence of coupling between the ramp operator of eq. (5.32) and the single calls of the tomographic projectors. In a way, this test is an additional confirmation of the importance of the operator coupling in iterative tomographic reconstruction as already discussed in (4.5).



(a) Case of perfect dataset

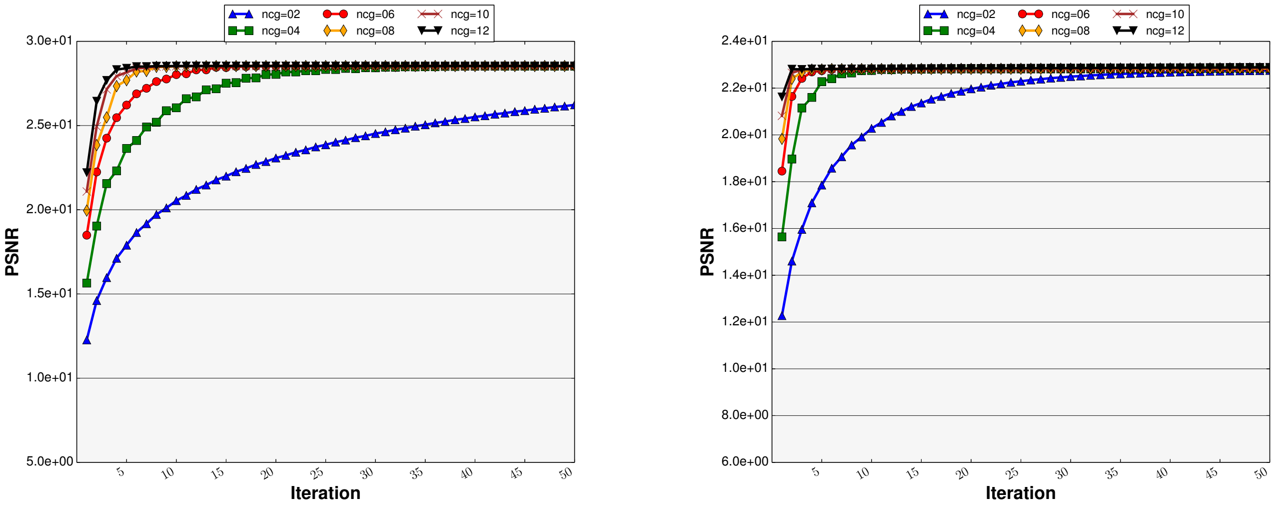
(b) Case of undersampled dataset



(c) Case of noisy dataset

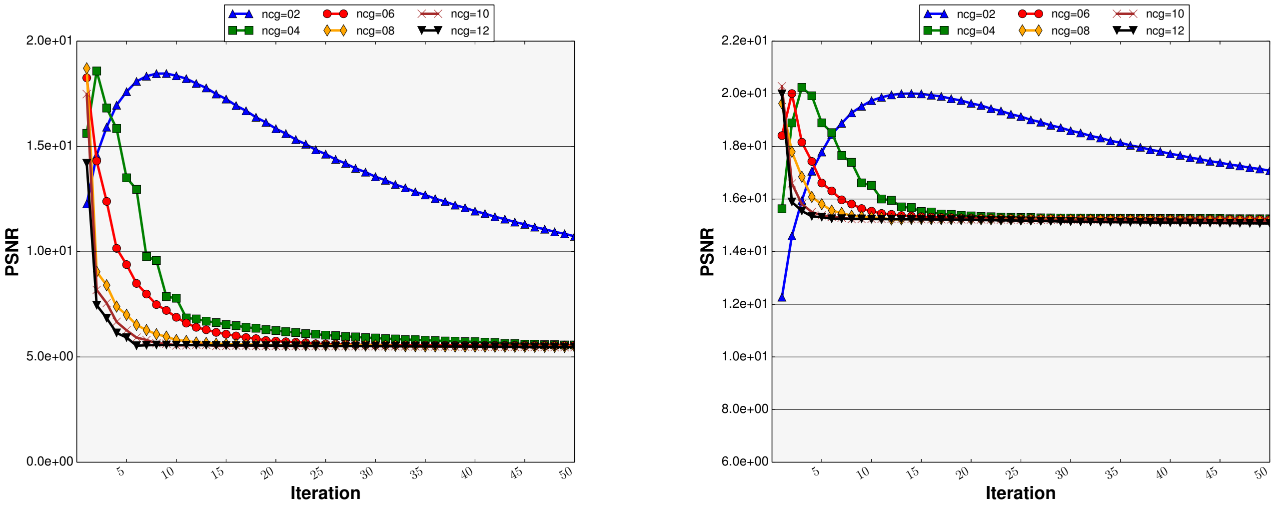
(d) Case of underconstrained dataset

Figure 5.12: Reconstruction accuracy of the ADMM as a function of the number of ADMM iterations for different numbers of CG sub-iterations. Each plot corresponds to the reconstruction of one of the four datasets PERF, UNDS, NOIS and UNDC.



(a) Case of perfect dataset

(b) Case of undersampled dataset



(c) Case of noisy dataset

(d) Case of underconstrained dataset

Figure 5.13: Reconstruction accuracy of the ADMM as a function of the number of ADMM iterations for different numbers of CG sub-iterations using the lung phantom for validation Each plot corresponds to the reconstruction of one of the four datasets PERF, UNDS, NOIS and UNDC.

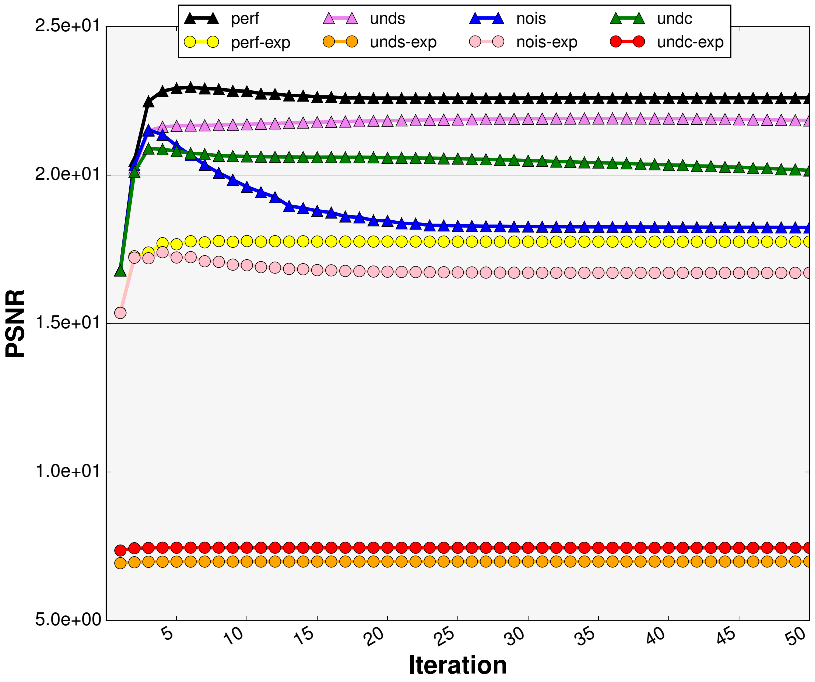


Figure 5.14: Comparison of reconstruction accuracy as a function of the number of iterations between ADMM and ADMM-EXP, the latter briefly described at the end of (5.2.3).

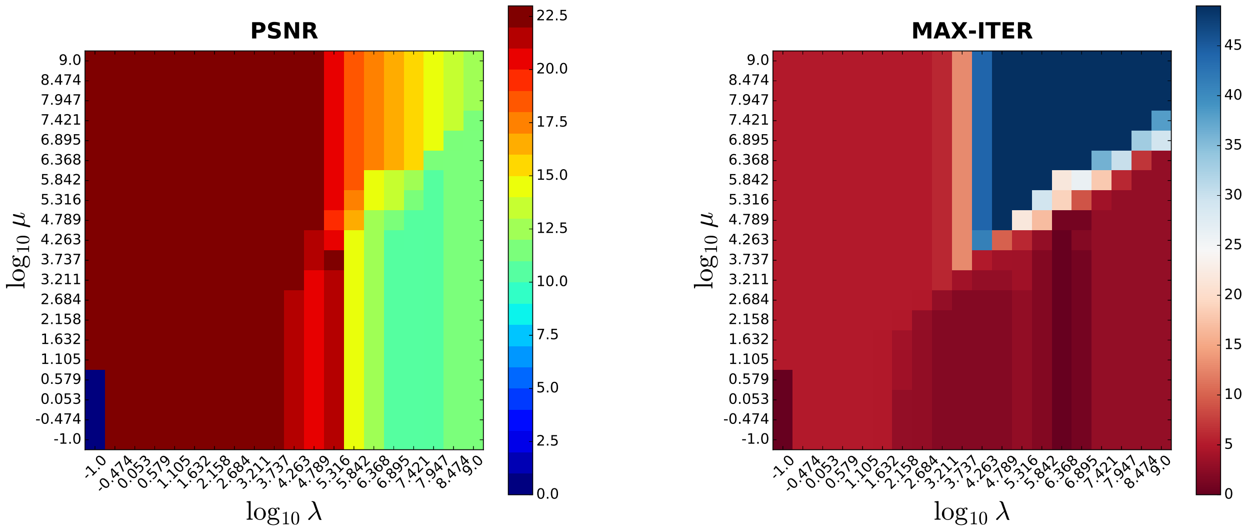
5.2.4 Optimal selection of λ and μ

A "brute-force-search" approach is used to study the influence of λ and μ on the reconstruction quality. Both parameters are spanned in the interval $[10^{-1}, 10^{9}]$. For $\mu < 10^{-1}$ and $\forall\, \lambda$, ADMM does not converge.

Results are presented in the form of coupled maps for each kind of dataset (Fig. 5.15 and 5.16): the map on the left represents the maximum PSNR obtained in 50 iterations; the map on the right indicates the minimum number of iterations required to achieve the highest PSNR.

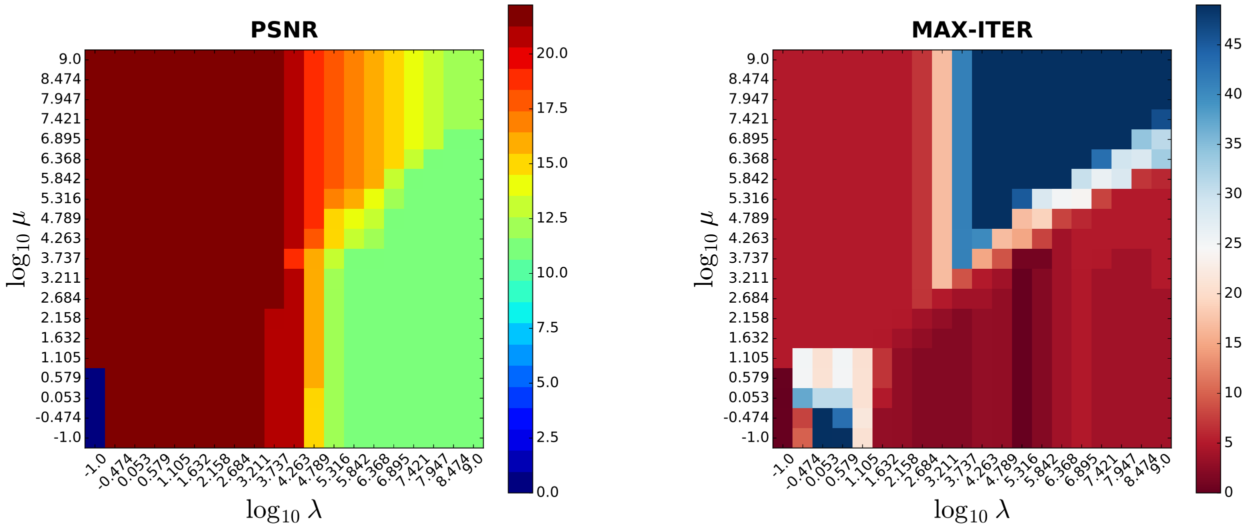
A first important observation is that there is no strong dependence of accuracy and convergence on on the ratio λ/μ, which appears once in the shrinkage operator (Eq. (5.25)) as a threshold level determining which pixels of $\mathbf{z}$ are set to zero or left unaltered. The presence of μ alone inside the CG and the shrinkage function breaks the potential dependence of the ADMM on λ/μ. The occurrence of μ in both subproblems makes this parameter more vital for accuracy and convergence of the ADMM than λ: as stated before, μ is forced to be $> 10^{-1}$, regardless of λ, to prevent the iterative process from diverging.

All PSNR maps of Fig. 5.15 and 5.16 feature a rectangular area on the left where the maximum accuracy (PSNR > 20) is achieved for $\forall\, \mu > 10^{1}$ and $\lambda \in [10^{-0.5}, 10^{2.7}]$, which roughly corresponds to a number of iterations between 5-10. For noisy datasets (Fig. 5.16(b,d)), the best reconstruction



(a) Reconstruction of PERF

(b) Reconstruction of PERF

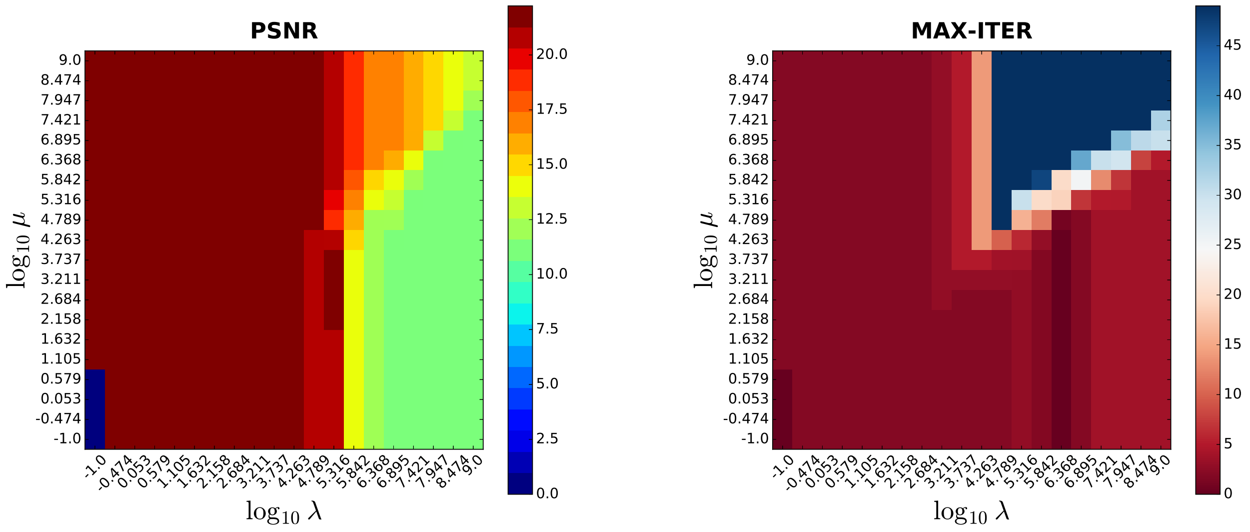


(c) Reconstruction of UNDS

(d) Reconstruction of UNDS

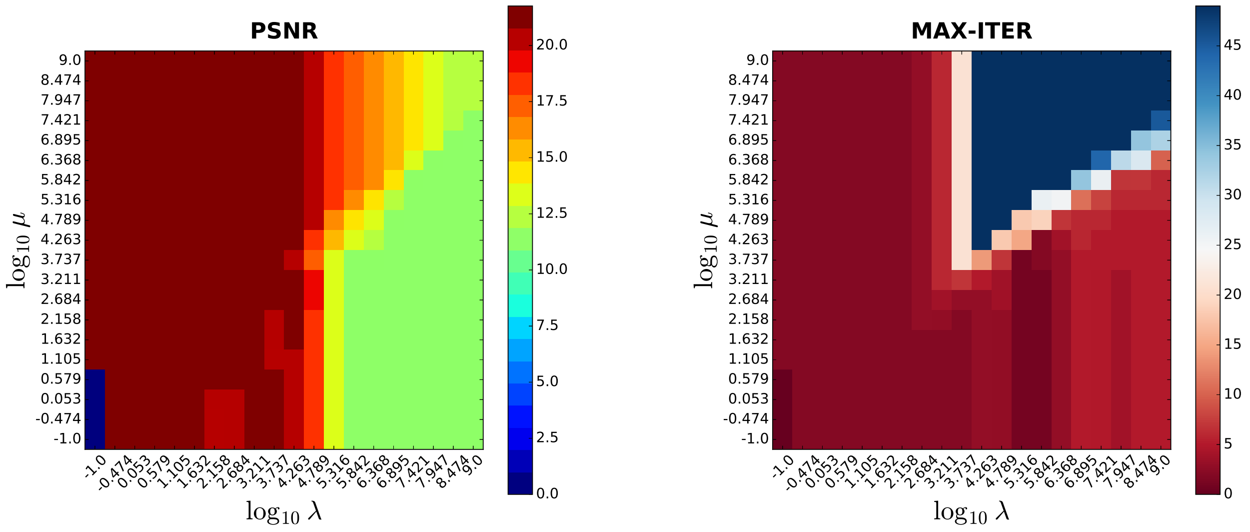
Figure 5.15: On the left column, maps of maximum PSNR, on the right column, maps of number of iterations where the maximum PSNR is achieved of ADMM reconstruction of PERF and UNDS for different values of λ and μ.

quality is obtained with a number of iterations $\leqslant 5$.



(a) Reconstruction of NOIS

(b) Reconstruction of NOIS



(c) Reconstruction of UNDC

(d) Reconstruction of UNDC

Figure 5.16: On the left column, maps of maximum PSNR, on the right column, maps of number of iterations where the maximum PSNR is achieved of ADMM reconstruction of NOIS and UNDC for different values of λ and μ.

The top right area of all maps features very low PSNR, achieved after the chosen maximum number of iterations. The corresponding reconstructions are very blurred, because in this interval of μ and λ the regularization term takes over the fidelity one.

The optimal selection of λ and μ falls in the aforementioned top left rectangular area, corresponding to a number of iterations between 5-10.

Experiments conducted on the lung phantom dataset (Fig. 5.17 and 5.18) confirm the same dependence of accuracy and convergence of the ADMM on the parameters μ and λ.

5.2.5 Physical constraints

$\mathbf{x}$ is the map of a selected physical quantity (e.g. attenuation coefficient), usually characterized only by positive values. This a-priori information can be used during the reconstruction and the negative pixels of x can be totally or partially steered towards 0. The application of these “physical constraints” after the first subproblem of the ADMM corresponds to the following simple operator:

$$x_i^{\text{phys}} = \begin{cases} x_i & \text{if} \;\; x_i > 0 \\ \beta \cdot x_i & 0 \leqslant \beta < 1 \;, \; \text{if} \;\; x_i < 0 \end{cases} \quad . \tag{5.33}$$

(5.33) can be regarded as a shrinkage operation: the difference between (5.33) and (5.25) lies in the fact that the first is applied directly on $\mathbf{x}$ and the second on $\mathbf{z}$. The term “physical” comes from the fact that $\mathbf{x}$ is supposed to be a map of attenuation coefficients that are physical quantities supposed to be $\geqslant 0$.

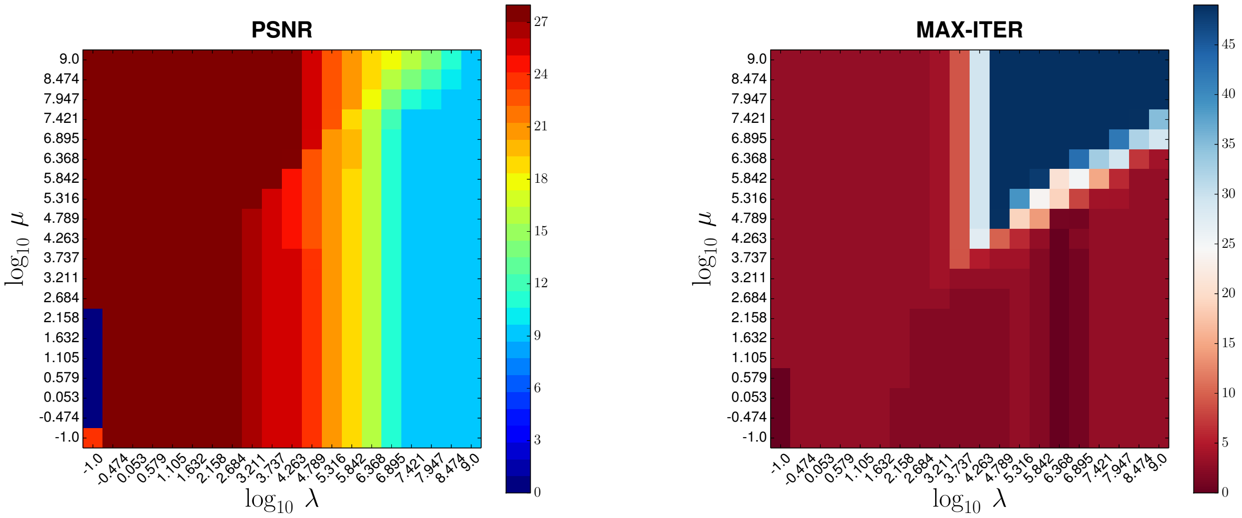
Experiments are here run with the “hard threshold”, i.e. $\beta = 0$. Results in Fig. 5.19 show that physical constraints provide superior results in three cases out of four (PERF, UNDS and NOIS). For the UNDC case, the accuracy of the ADMM with physical constraints drops after around 10 iterations, but the peak of quality is higher with respect to the one achieved without constraints.

These results clearly indicate that a trivial operator like (5.33) substantially impact the reconstruction quality and should be utilized whenever possible (e.g, it cannot be used with interior tomography datasets). This fact is confirmed by the plots in Fig. 5.20 obtained with the lung phantom dataset, despite these curves look quite different from those in Fig. 5.19.

5.2.6 Warm initialization

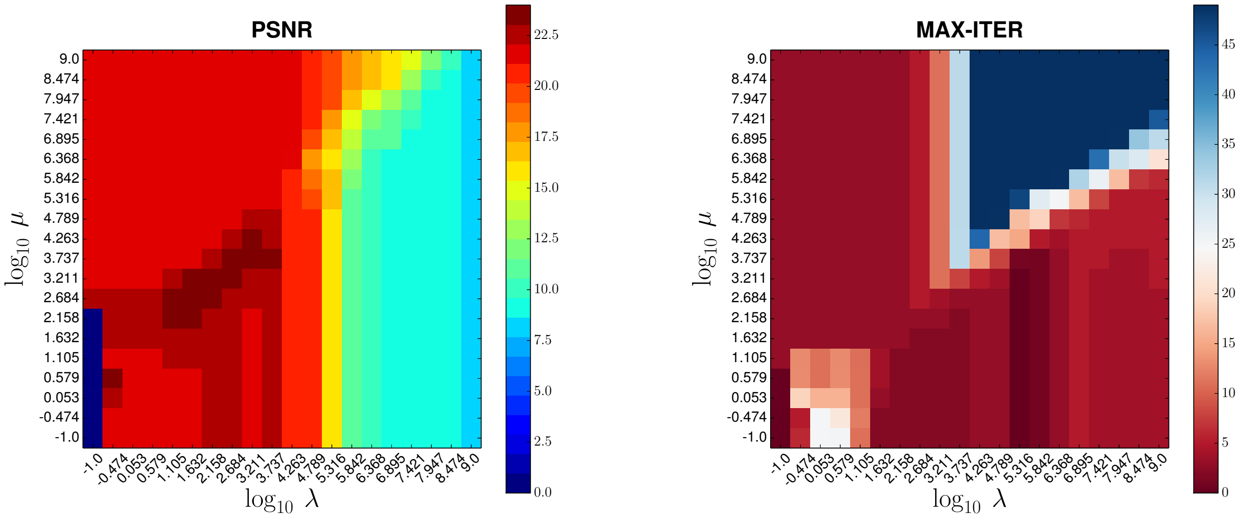
For the first iteration of the ADMM, the $\mathbf{x}$-subproblem needs an initial guess for $\mathbf{x}^{(0)}$. In contrast to a "cold" initialization (trivial guess, $\mathbf{x}^{(0)} = 0$), a “warm initialization” requires an educated guess, To understand whether warm initialization can boost the reconstruction quality or speed up convergence, the four datasets are reconstructed with “cold-initialization”, $\mathbf{x}^{(0)} = 0$, and warm initialization, $\mathbf{x}^{(0)} = \text{FBP}(\mathbf{b})$. The Ram-Lak filter is used to perform the FBP reconstruction to prevent losses in spatial resolution before starting the ADMM procedure.

Results in Fig. 5.21 show that the cold and warm initialization (plots are labeled with an additional “-init”) reach asymptotically the same reconstruction quality, but the accuracy in the first 15 iterations is severely deteriorated in case of warm initialization. For each dataset, the ADMM with warm initialization does not reach the peak quality and generally “stagnates” giving very little contribution



(a) Reconstruction of PERF

(b) Reconstruction of PERF

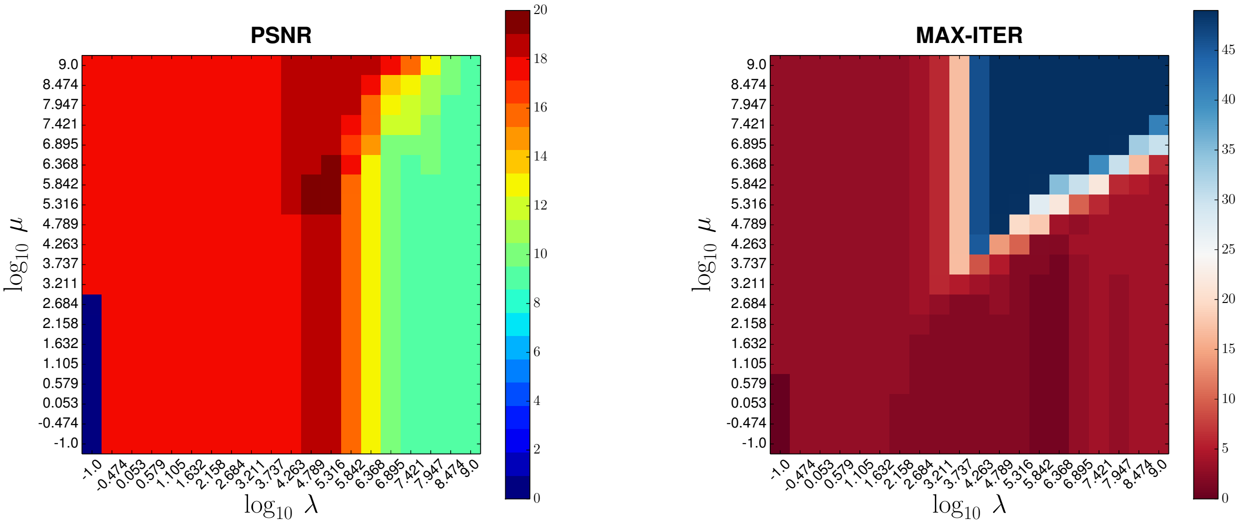


(c) Reconstruction of UNDS

(d) Reconstruction of UNDS

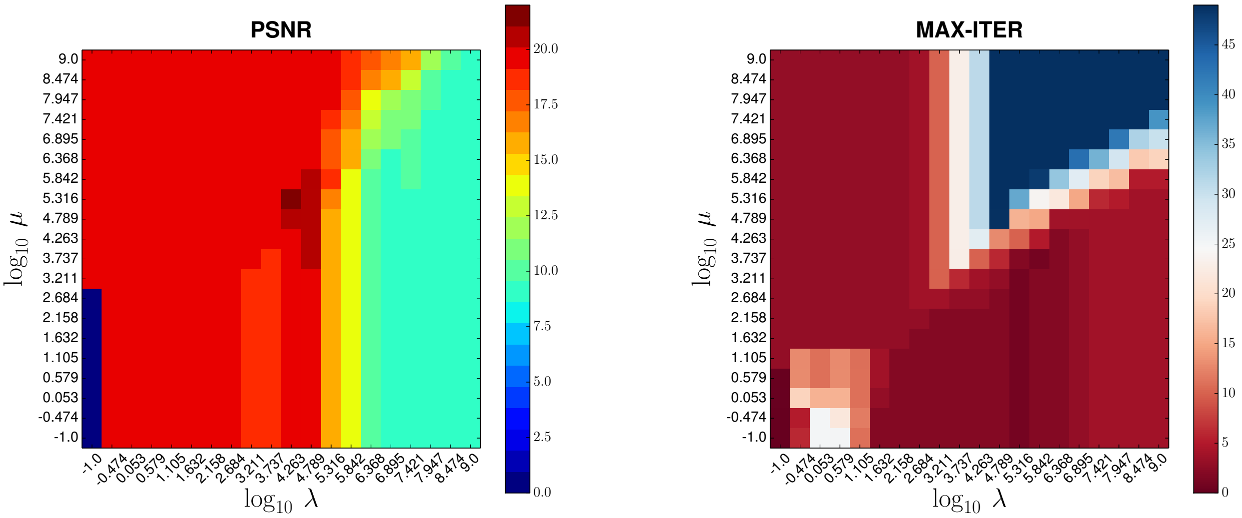
Figure 5.17: On the left column, maps of maximum PSNR, on the right column, maps of number of iterations where the maximum PSNR is achieved of ADMM reconstruction of PERF and UNDS of the lung phantom dataset for different values of λ and μ.

to the accuracy already achieved by FBP. On the basis of these results, the warm initialization strategy is simply discarded.



(a) Reconstruction of NOIS

(b) Reconstruction of NOIS



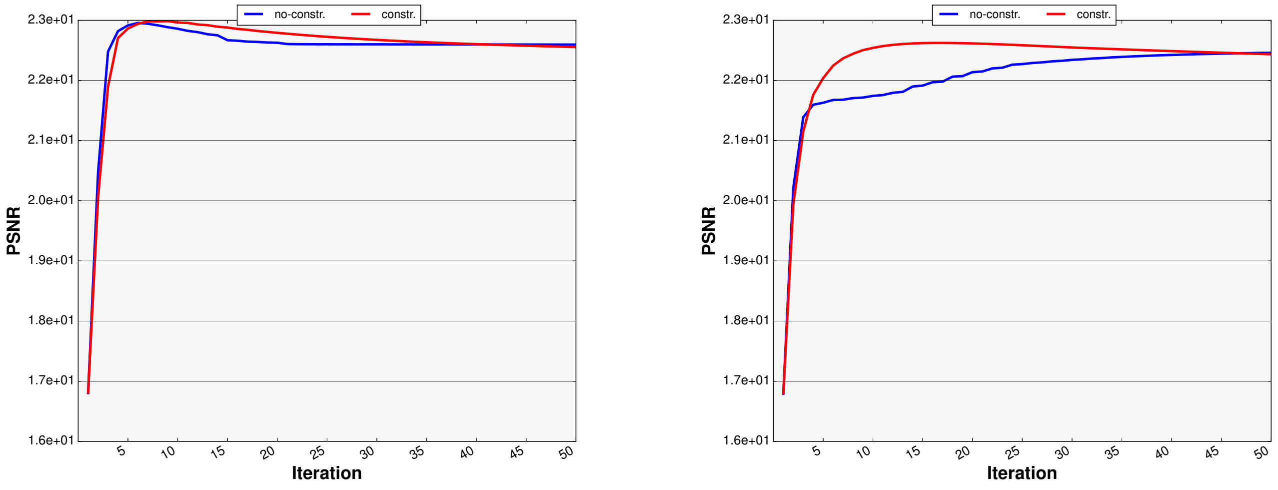
(c) Reconstruction of UNDC

(d) Reconstruction of UNDC

Figure 5.18: On the left column, maps of maximum PSNR, on the right column, maps of number of iterations where the maximum PSNR is achieved of ADMM reconstruction of NOIS and UNDC of the lung phantom dataset for different values of λ and μ.

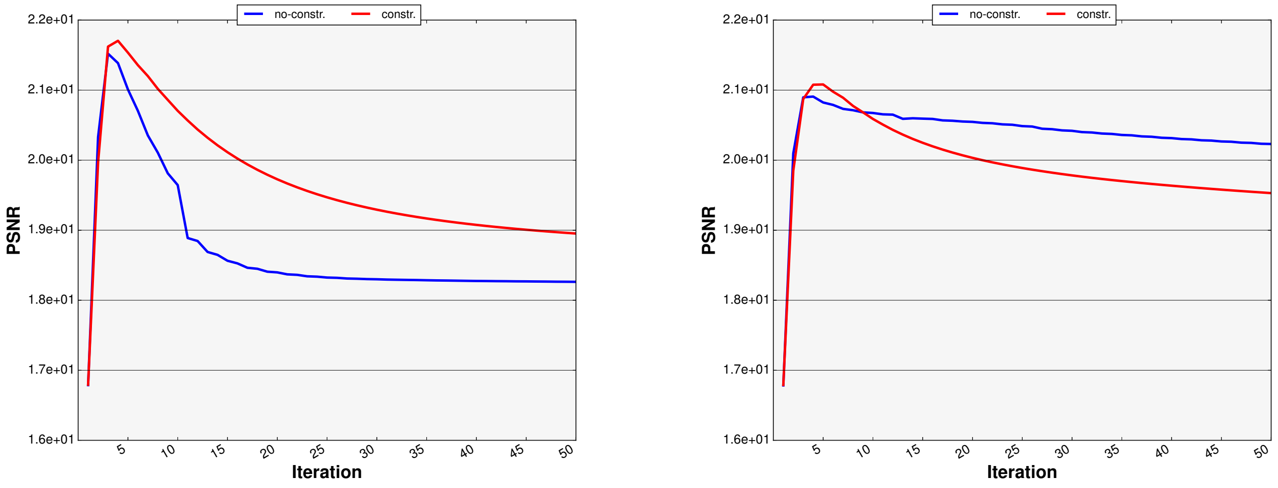
5.2.7 Testing different regularization schemes

The following experiments illustrate the potential advantage of using sophisticated regularization schemes to dump the noise component or the streak artifacts affecting reconstructions of undercon-



(a) Reconstruction of PERF

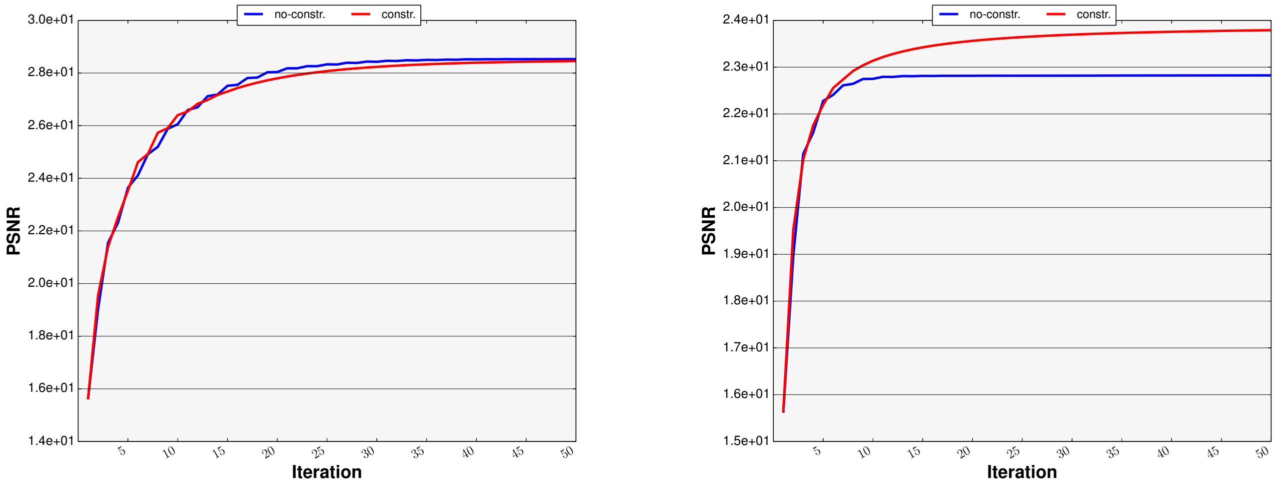
(b) Reconstruction of UNDS



(c) Reconstruction of NOIS

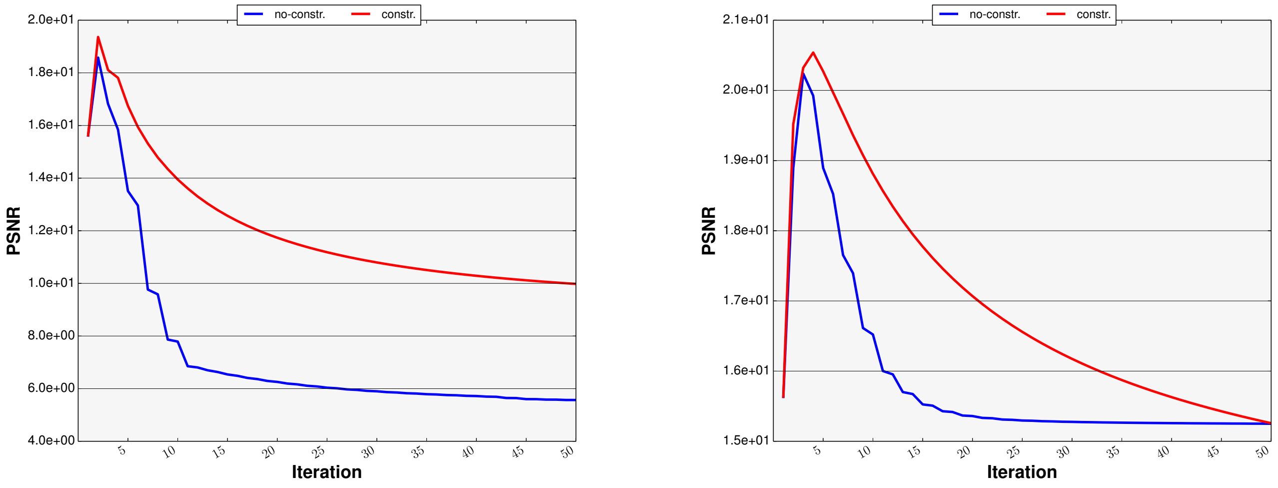
(d) Reconstruction of UNDC

Figure 5.19: Reconstruction accuracy as a function of the number of iterations for the case in which no physical constraints are used (blue curves labeled "no-constr" in the legend) and the case in which they are with $\beta = 0$ (red curves labeled "constr" in the legend).



(a) Reconstruction of PERF

(b) Reconstruction of UNDS



(c) Reconstruction of NOIS

(d) Reconstruction of UNDC

Figure 5.20: Reconstruction accuracy as a function of the number of iterations for the case in which no physical constraints are used (blue curves labeled "no-constr" in the legend) and the case in which they are with $\beta = 0$ (red curves labeled "constr" in the legend) for the lung phantom dataset.

strained datasets. The used algorithms are: ADMM-TV, ADMP using split-Bregman TV (ADMP-TV), ADMP using NLM (ADMP-NLM), ADMP using NLTV (ADMP-NLTV) and ADMP using GTV (ADMP-GTV). Analytical reconstructions with FBP using an additional Parzen window for the ramp filtering are displayed to give an idea of quality of the input datasets. The reconstruction with ADMM-TV is computed with an optimal parameter set. Reconstructions with ADMP are performed with sub-optimal parameters (no brute-force search of the parameter space was performed to find the optimal parameter set) that already yield superior results with respect to the ADMM-TV.

When dealing with purely noisy data, sophisticated regularization schemes provide higher quality reconstructions than ADMM-TV already at visual inspection: the noisy pattern characterizing Fig. 5.22 (b) is partially or completely removed in Fig. 5.22 (c,d,e,f), although with different SNR-spatial resolution trade-offs. The edge line profiles in Fig. 5.23 prove that an effective noise removal can be achieved without substantial losses in terms of spatial resolution.

The purely undersampled case represents a different kind of challenge: regularization schemes are designed to filter out random noise, whereas streak artifacts represent a type of structured noise that can be "interpreted" by the denoising method as part of the true signal to restore. To prevent enhancement of the streak artifacts, the regularization parameters must be chosen carefully. The pattern due to undersampling can be completely removed by sophisticated regularization schemes, but at the cost of severely decreasing the spatial resolution. The ADMP reconstructions in Fig. 5.24 offer a

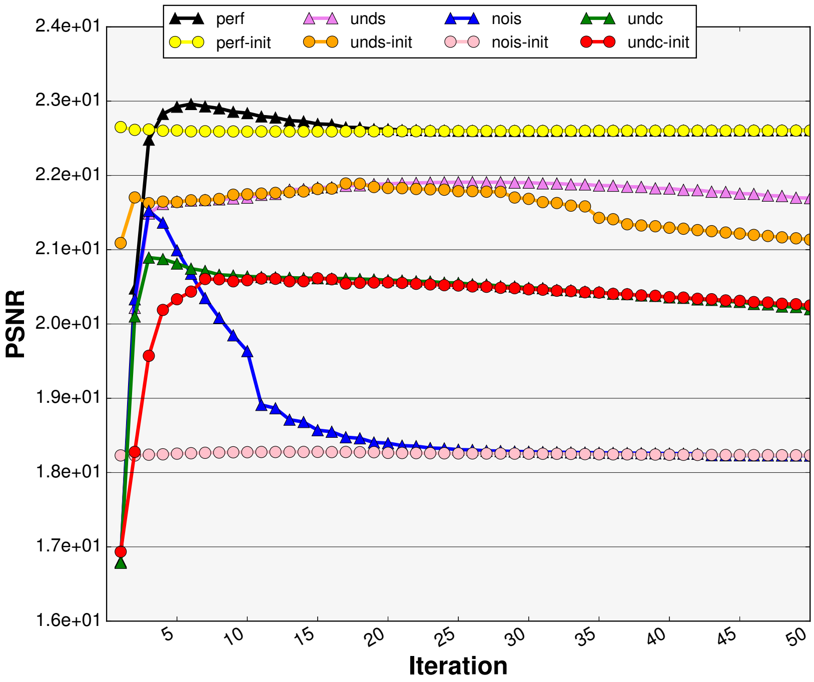
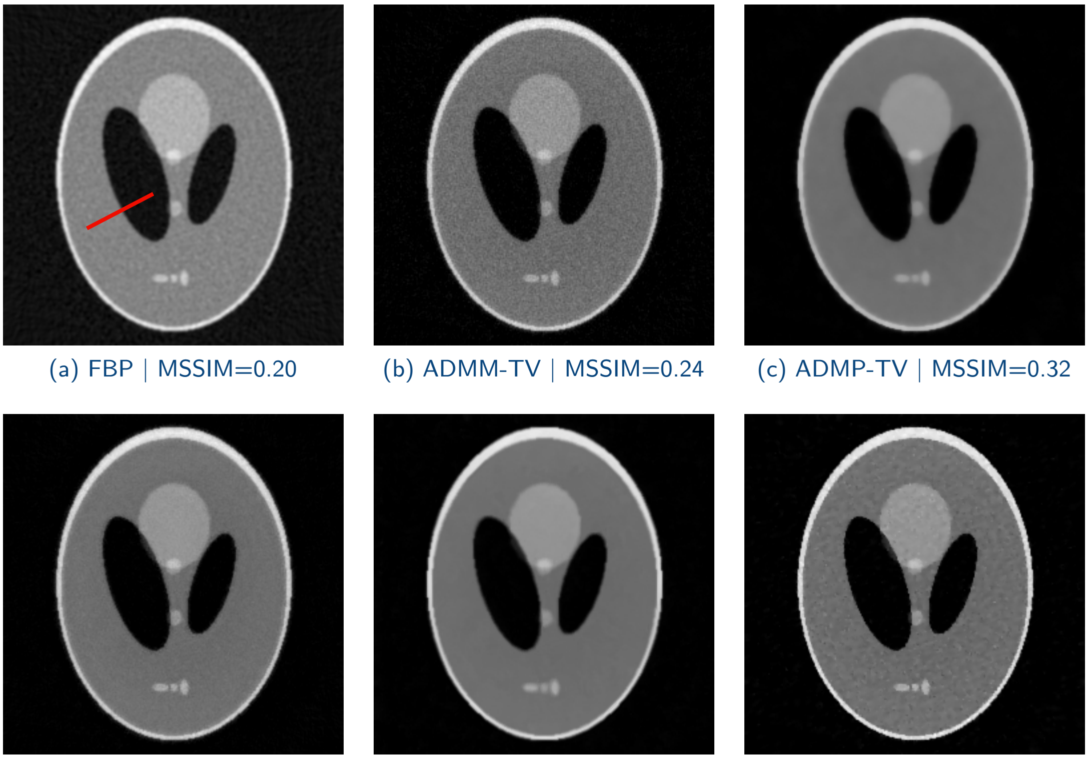


Figure 5.21: Comparison of reconstruction accuracy as a function of the number of iterations between "cold initialization", i.e., $\mathbf{x}^{(0)} = 0$, and "warm initialization", i.e., $\mathbf{x}^{(0)} = \text{FBP}(\mathbf{b})$ (with Ram-Lak filter) used to feed the $\mathbf{x}$-subproblem at the first ADMM iteration. Results with warm initialization feature the additional label "init".


(a) FBP | MSSIM=0.20 (b) ADMM-TV | MSSIM=0.24 (c) ADMP-TV | MSSIM=0.32

(d) ADMP-NLM | MSSIM=0.26 (e) ADMP-NLTV | MSSIM=0.33 (f) ADMP-GTV | MSSIM=0.30

Figure 5.22: Reconstructions of NOIS with FBP (Parzen filter), ADMM-TV, ADMP-TV, ADMP-NLM, ADMP-NLTV, ADMP-GTV. The red segment in (a) identifies the position of the line profiles shown in Fig. 5.23.

reasonable visual compromise between artifacts reduction and spatial resolution (as also demonstrated by the edge profile analysis in Fig. 5.25). The visual improvement with respect to ADMM-TV is, however, inferior to what experienced in the case of a purely noisy dataset.

The case of an underconstrained dataset in Fig. 5.26 shows, once again, that great visual results can be achieved by making use of sophisticated regularization schemes. The analysis of the edge profiles for the reconstructions of Fig. 5.26 is shown in Fig. 5.27.

In the last experiment, we show that "upgrading" a regularization scheme from 2D to 3D generally improves the reconstruction of a piecewise constant volume. A 3D modified SL phantom with 256 × 256 × 256 pixels is used to construct three datasets, 3D-NOIS, 3D-UNDS and 3D-UNDC, which represent the 3D versions of NOIS, UNDS and UNDC. These new datasets are reconstructed by the ADMP-TV, where both subproblems operates slice by slice, and ADMP-TV3, where the $\mathbf{x}$-subproblem operates slice by slice and the $\mathbf{z}$-subproblem on the entire volume. Despite not showing great differences at visual inspection, the volume computed by ADMP-TV3 is more accurate than the one computed by ADMP-TV, as proved by the MSSIM values shown in Tab. 21 (and calculated on the entire volume). The performance of the iterative reconstruction can substantially improve when TV works in 3D, due to the higher sparsity characterizing the object volume-wise rather than slice-wise.

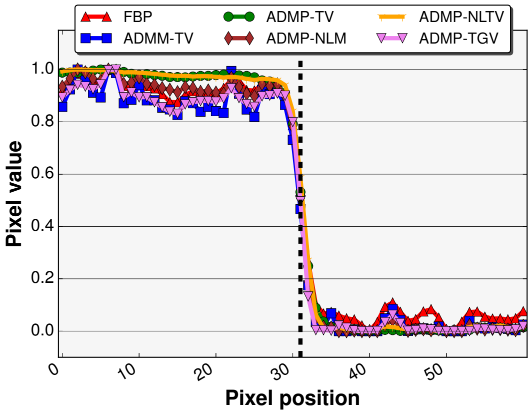


Figure 5.23: Edge line profiles for the reconstructions shown in Fig. 5.22 corresponding to the segment indicated in Fig. 5.22(a). The black vertical dashed line identifies the edge location.

	3D-NOIS	3D-UNDS	3D-UNDC
ADMP-TV	0.18	0.20	0.17
ADMP-TV3	0.30	0.36	0.27

Table 21: MSSIM values for the reconstruction of SL-3D-NOISE, SL-3D-UNDER and SL-3D-UCONSTR computed by ADMP-TV, where TV operates on individual slices, and ADMP-TV3, where TV operates on the entire volume.

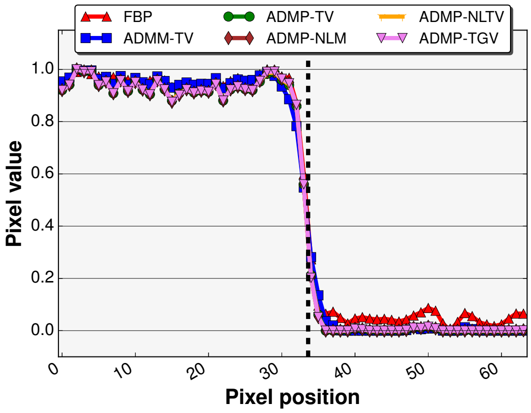
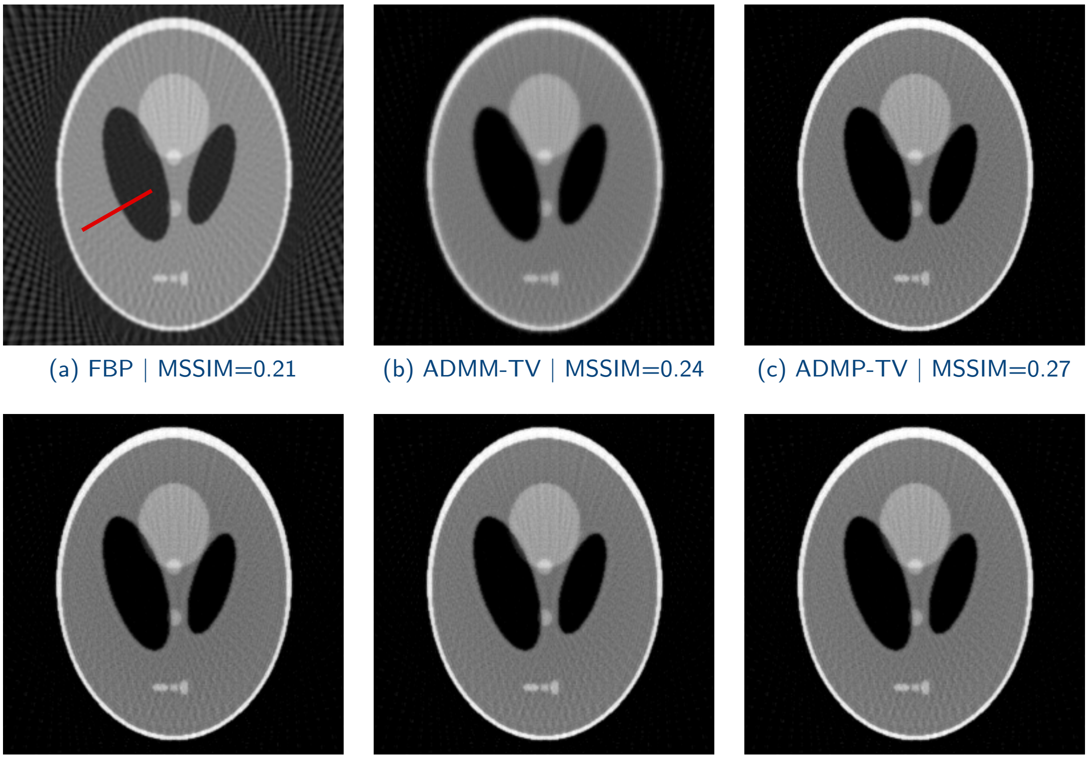


Figure 5.25: Edge line profiles for the reconstructions shown in Fig. 5.24 corresponding to the segment indicated in Fig. 5.24(a). The black vertical dashed line identifies the edge location.



(a) FBP | MSSIM=0.21 (b) ADMM-TV | MSSIM=0.24 (c) ADMP-TV | MSSIM=0.27

(d) ADMP-NLM | MSSIM=0.28 (e) ADMP-NLTV | MSSIM=0.27 (f) ADMP-GTV | MSSIM=0.28

Figure 5.24: Reconstructions of UNDS with FBP (Parzen filter), ADMM-TV, ADMP-TV, ADMP-NLM, ADMP-NLTV, ADMP-GTV. The red segment in (a) identifies the position of the line profiles shown in Fig. 5.25.

The computational cost, in terms of both runtime and memory allocation, prevents the usage of many sophisticated regularization (e.g. NLM, NLTV and GTV) schemes in a 3D fashion for datasets of real sizes.

The results presented in this paragraph should not be considered conclusive, as they are biased by the choice of the used phantom (in this case, SL). A phantom characterized by a different structural complexity or intrinsic CNR than the SL, being eventually smooth instead of piecewise-constant, would certainly turn the presented results upside down.

5.3 Practical tips for ADMM users

5.3.1 Choosing the stopping criterion

The $\mathbb{L}_2$-norm of the difference between consecutive reconstructions, i.e. $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2$, can be used to decide when to stop an iterative algorithm: if $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2 > \epsilon$, the procedure continues, otherwise it stops. This stopping criterion works well, if the iterative algorithm features a strictly monotone increase of the reconstruction quality as a function of the number of iterations like the

(a) FBP | MSSIM=0.17 (b) ADMM-TV | MSSIM=0.20 (c) ADMP-TV | MSSIM=0.24

(d) ADMP-NLM | MSSIM=0.27 (e) ADMP-NLTV | MSSIM=0.26 (f) ADMP-GTV | MSSIM=0.30

Figure 5.26: Reconstructions of UNDC with FBP (Parzen filter), ADMM-TV, ADMP-TV, ADMP-NLM, ADMP-NLTV, ADMP-GTV. The red segment in (a) identifies the position of the line profiles shown in Fig. 5.27..

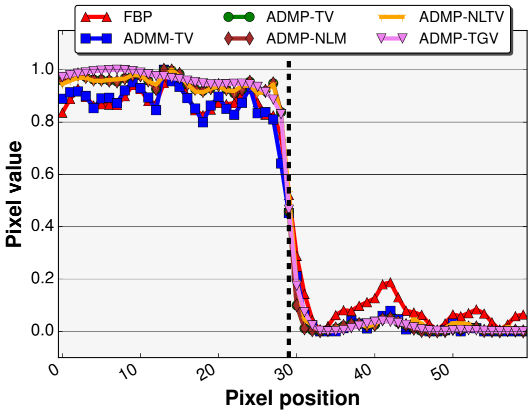


Figure 5.27: Edge line profiles for the reconstructions shown in Fig. 5.26 corresponding to the segment indicated in Fig. 5.26(a). The black vertical dashed line identifies the edge loacation.

PWLS (Fig. 5.7(b)).

The accuracy of ADMM (ADMP) reconstructions generally peaks after very iterations; at that point, $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2$ may not be small enough to comply to a standard ϵ-threshold criterion. When reconstructing a dataset for the first time, a good strategy is to run the ADMM (ADMP) on a single slice for around 50 iterations, saving each intermediate result. The set of reconstructions is, then, visually inspected and analyzed with the PSNR to know at which iteration the accuracy reaches the maximum. The number of iterations obtained with this procedure is likely to represent a reasonable stopping criterion for the reconstruction of the entire dataset.

5.3.2 Selecting the parameters λ and μ

Various studies have presented criteria to automatically select the optimal regularization strength (λ, in this case) for large ill-posed inverse problems. Examples are the generalized cross-validation method [256], the L-curve criterion [257], the discrepancy principle [258, 259], the subspace [260] and the generalized [202] information criterion. The problem is that these criteria are usually based on mathematical principles not directly correlated to the image quality, as the image quality itself cannot be precisely defined and modeled.

If a reliable criterion for the parameter selection is missing, it is convenient to collect some experience with manual tuning. For the ADMM, only the order of magnitude of λ and μ counts. A good strategy is to select first the appropriate order of magnitude for μ, that rules the intermixing between the $\mathbf{x}$- and the $\mathbf{z}$-subproblem. Once μ is set, λ can be chosen on the basis of the target SNR/spatial resolution trade-off. Several trials will be needed, especially at the beginning, to get a "feeling" with manual parameter tuning.

5.3.3 Selecting the regularization scheme

In 5.1.3, different regularization schemes have been introduced and some of them have been used to run the experiments presented in 5.2.7. Here we briefly discuss how to select the appropriate regularization scheme for a given dataset.

For piecewise constant objects, TV usually offers a very good accuracy/efficiency trade-off. ADMP-TV, implemented with either split-Bregman or Chambolle TV, can yield better results than ADMM-TV but at a higher computational cost.

For smooth objects, TV can generate artifacts and loss of structural information, thus NLM, NLTV and GTV are preferable solutions. Preliminary results, as those presented in Fig. 5.22 and 5.26, have shown that NLTV provides a remarkable SNR/spatial resolution trade-off, while being less incline to generate wavy patterns in piecewise constant ROIs compared to NLM (see reconstruction in Fig. 5.26(d)) and GTV (see reconstructions in Fig. 5.22(f) and Fig. 5.26(f)). GTV seems also to perform poorly in preserving edges when the intrinsic CNR of the object is low: the reconstruction in Fig. 5.26(f) is characterized by very sharp black ellipses, whereas the same does not hold for the middle-sized circle on the top center, that has by construction lower intrinsic CNR with respect to grey "filling" of SL.

It is opinion of the author that dictionary-based regularization schemes should generally be avoided due to the scarse accuracy/efficiency trade-off. The OMP scans all the atoms of the dictionary to find the best candidates to represent a given patch and this operation results highly time-consuming without a GPU implementation [261] or a sparsity-based algorithm [262]. In addition, the action of dictionary-based regularization schemes is bounded to be local, since overlapping patches are used. TV or Huber penalty are, then, much more efficient edge-preserving schemes with a local action.

Chapter 6

Reconstruction strategies for full interior tomography datasets

6.1 The interior tomography problem

Interior tomography (INT) [92] is the problem of reconstructing an object function from truncated projections. Truncated datasets are acquired when the object support $S \not\subseteq \text{FOV}$, i.e. the detector does not fit the object along certain directions. INT scans allow high resolution investigations of small ROIs inside the specimen under study [263], while decreasing the integral dose delivered to the entire volume [264].

Data reconstruction is the major challenge posed by interior tomography.

Until the early 2000s, it was believed that INT data could not yield a unique solution [92]. Consider the complete dataset $p(\theta, t)$ with $\theta \in [0, \pi)$ and $t \in [-1, 1]$ and the partial dataset $\tilde{p}(\theta, t)$ $\theta \in [0, \pi)$ and $a < |t| \leqslant 1$. Even if $\tilde{p}(\theta, t) = 0$ for $|t| \leqslant a$, it can still produce a non-zero local reconstruction inside $r \leqslant a$, i.e. $\tilde{x}(\theta, r) = \text{FBP}(\tilde{p}(\theta, t)) \neq 0$ for $r \leqslant a$ [92, 265]. If $\tilde{\tilde{p}}(\theta, t)$ with $\theta \in [0, \pi)$ and $t \in [-a, a]$ is the INT dataset, $\tilde{\tilde{p}}(\theta, t) = p(\theta, t) - \tilde{p}(\theta, t)$ and it follows that:

$$
\begin{aligned}
\tilde{x}(\theta, r) = \text{FBP}\left(\tilde{\tilde{p}}\right)(\theta, r)\Big|_{|r| \leqslant a} &= \text{FBP}\left(p\right)(\theta, r)\Big|_{|r| \leqslant 1} - \text{FBP}\left(\tilde{p}\right)(\theta, r)\Big|_{|r| \leqslant 1} \\
&= \cancel{\text{FBP}\left(p\right)(\theta, r)}\Big|_{a \leqslant |r| \leqslant 1} + \text{FBP}\left(p\right)(\theta, r)\Big|_{|r| \leqslant a} - \cancel{\text{FBP}\left(\tilde{p}\right)(\theta, r)}\Big|_{a \leqslant |r| \leqslant 1} - \tilde{x}(\theta, r)\Big|_{|r| \leqslant a} \\
&= x(\theta, r)\Big|_{|r| \leqslant a} - \tilde{x}(\theta, r)\Big|_{|r| \leqslant a} \quad .
\end{aligned} \tag{6.1}
$$

Equation (6.1) shows that the reconstruction of the INT dataset, $\tilde{\tilde{p}}(\theta, t)$, equals the exact object inside the FOV of radius a, $x(\theta, r)$ with $|r| \leqslant a$, minus an artifact image, $\tilde{x}(\theta, r)$ with $|r| \leqslant a$, arising from the data outside the FOV. Apart from being not exact, $\tilde{\tilde{x}}(\theta, r)$ is also not unique, as $\tilde{x}(\theta, r)$ can be constructed as [92]:

$$\tilde{x}(\theta, r) = \frac{1}{\pi} \int_{|u|}^{+\infty} dr \, \left(r^2 - |u|^2\right)^{-1/2} g'(r) \quad , \tag{6.2}$$

where g is a generic Schwartz function vanishing in $[-a, a]$.
From a practical point of view, FBP reconstructions of INT datasets are inevitably affected by a low-frequency artifact, that arises from the action of the ramp filter on truncated projections [266–268]. This artifact gives a bowl-shaped profile to constant ROIs with very low values at the image center and very high values close to the border of the reconstruction circle. The CNR of reconstructions affected by the bowl artifact is so degraded, that different structures can be barely recognizable at visual inspection. All considered, INT reconstructions are not exact (therefore, quantitative), not unique and cannot be used for morphological studies due to presence of the bowl artifact.

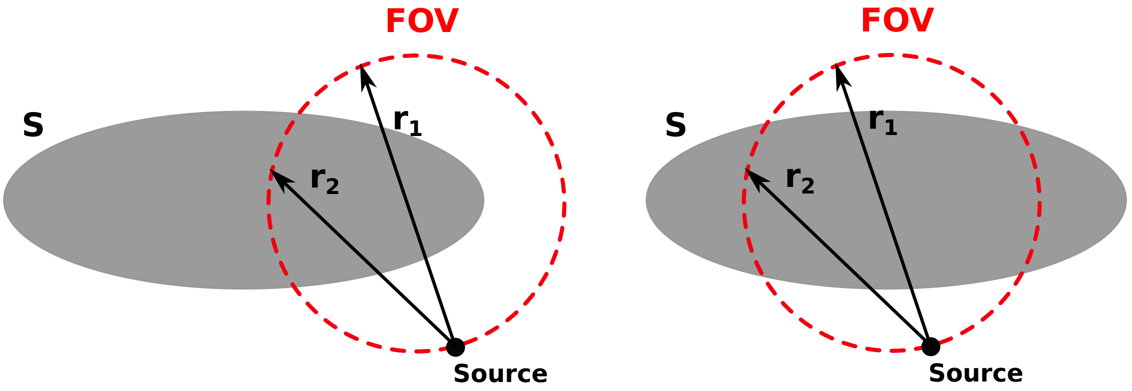
6.2 Overview of reconstruction methods for INT data

Only two methods can provide reconstructions of interior tomography data, that are not affected by the bowl-artifact: *differentiated backprojection* and *sinogram extrapolation*. The idea and theory behind these two methods are briefly explained in the following.

6.2.1 Differentiated backprojection

In 2004, [269] proposed a two-step algorithm for the reconstruction of INT datasets, called *differentiated backprojection (DBP)*. DBP, first, backprojects the derivative of the projections along the channel direction and, then, retrieves the object by inverting a finite Hilbert transform. Differently from a FBP reconstruction, the backprojection of the differential sinogram is only partially affected by the data truncation. The idea of DBP comes from the *local* or *lambda tomography* [270], which retrieves the object from the backprojection of the second derivatives of the projections. The relation between the backprojected differential sinogram and the Hilbert transform is here derived. Considering an object f, $\mathbf{x} = (x_1, x_2)$, $\mathbf{n} = (\cos\theta, \sin\theta)$, $\hat{p}(\theta, r) = \mathcal{F}_t\{p(\theta, t)\}$ and i the imaginary unit, the backprojected differential sinogram results [269]:

$$\begin{aligned} g_0(x_1, x_2) &:= \int_0^{\pi} d\theta \, \frac{\partial p(\theta, \mathbf{x}\cdot\mathbf{n})}{\partial r} = 2\pi \int_0^{\pi} d\theta \int_{-\infty}^{+\infty} dr \, ir \, \hat{p}(\theta, r) \, \exp(2\pi i r \, \mathbf{x}\cdot\mathbf{n}) \\ &= 2\pi \int_0^{\pi} d\theta \int_{-\infty}^{+\infty} dr \, i \, \mathrm{sgn}(r) \, \hat{p}(\theta, r) \, |r| \, \exp(2\pi i r \, \mathbf{x}\cdot\mathbf{n}) \\ &= 2\pi \int_0^{\pi} d\theta \int_{-\infty}^{+\infty} dr \, i \, \mathrm{sgn}(r \sin\theta) \, \hat{f}(r\mathbf{n}) \, |r| \, \exp(2\pi i r \, \mathbf{x}\cdot\mathbf{n}) \end{aligned}$$



(a) (b)

Figure 6.1: According to [269], exact inversion of Eq. (6.3) is possible only along rays crossing opposite sides of the object support, S. In the figures, such rays are indicated with r_1, whereas the inversion is not possible along r_2.

$$= 2\pi \int_{-\infty}^{+\infty} du_1 \int_{-\infty}^{+\infty} du_2 \, i \, \mathrm{sgn}\left(u_2\right) \, \hat{f}\left(u_1, u_2\right) \, \exp\left(2\pi i \, \mathbf{x} \cdot \mathbf{u}\right) = -2\pi \, H_0 \, f\left(x_1, x_2\right) \quad , \tag{6.3}$$

where in the third passage one exploits that $\mathrm{sgn}(r) = \mathrm{sgn}(r \sin\theta)$ for $\theta \in [0, \pi)$ and the FSM, i.e. $\hat{p}(\theta, r) = \hat{f}(r\mathbf{n})$. The symbol H_0 stands for the Hilbert transform along the direction $\theta = 0$. Equation (6.3) can be also rewritten with the Hilbert transform computed along any direction in $[0, \pi)$ [269]:

$$g_\theta\left(x_1, x_2\right) = -2\pi H_\theta f\left(x_1, x_2\right) \quad . \tag{6.4}$$

For computational convenience, the Hilbert transform is usually inverted along the x_1- ($\theta = \pi/2$) or the x_2-axis ($\theta = 0$). In [269], Eq.(6.3) is invertible only along directions crossing opposite boundaries of the object support, thus, characterized by segments outside the object support but still inside the FOV where f is known to be zero (radii r_1 illustrated in Fig. 6.1(a,b)). The inversion formula results [269]:

$$\begin{aligned} f\left(x_1, x_2\right) &= f\left(t\mathbf{n} + s\mathbf{n}^{\perp}\right) \\ &= -\frac{1}{\sqrt{\left(s - L_t\right)\left(U_t - s\right)}} \left(\int_{L_t}^{U_t} ds' \sqrt{\left(s' - L_t\right)\left(U_t - s'\right)} \, \frac{H_\theta f\left(t\mathbf{n} + s'\mathbf{n}^{\perp}\right)}{\pi(s - s')} + C_t \right) \; , \end{aligned} \tag{6.5}$$

where L_t and S_t are the entrance and the exit point of the ray (of the kind of r_1 in Fig.6.1(a,b)) with S and C_t is a constant computed as follows:

$$C_t = -\int_{L_t}^{U_t} ds' \, \sqrt{(s'-L_t)(U_t-s')} \, \frac{H_\theta f\left(t\mathbf{n} + s'\mathbf{n}^\perp\right)}{\pi(s-s')} \quad . \tag{6.6}$$

The DBP method of [269] is extended in [118] to the case in which the FOV exceeds S only from one side (Fig. 6.2(a)). [271–273] introduce DBP also to the case in which FOV $\subset S$ and the attenuations coefficients are known in specific ROIs $\subset$ FOV (Fig. 6.2(b)). In all the presented configurations, DBP yields an exact solution, thus reconstructions are quantitative and free of the bowl artifact.

6.2.2 Sinogram extrapolation

Sinogram extrapolation techniques combined to standard FBP do not provide exact, quantitative reconstructions, but prevent the formation of the bowl artifact. Reconstructions are still usable for morphological analysis, unless the quantitativity is retrieved on the basis of some kind of a priori knowledge regarding the object under study [263]. Sinogram extrapolation was proposed for the first time at the end of the 70s, to remove the bowl artifact appearing on one or two sides of an FBP reconstruction, when the object exceeds the detector length.

In [266], the continuation of the data is performed by adding pieces of a chosen function, that are smoothly attached to the projection extremities. The continuation length is selected on the basis of the known object support and the extrapolated parts gradually fall to zero.

The method introduced by [274] works with partially truncated projections and required the knowledge of the object support. The truncation length of a specific projection is estimated from the data acquired at other angles. An iterative procedure going forward and back from the Radon to the image domain is used to reconstruct the object.

The reconstruction artifacts due to data truncation are mathematically analyzed in [267] for parallel

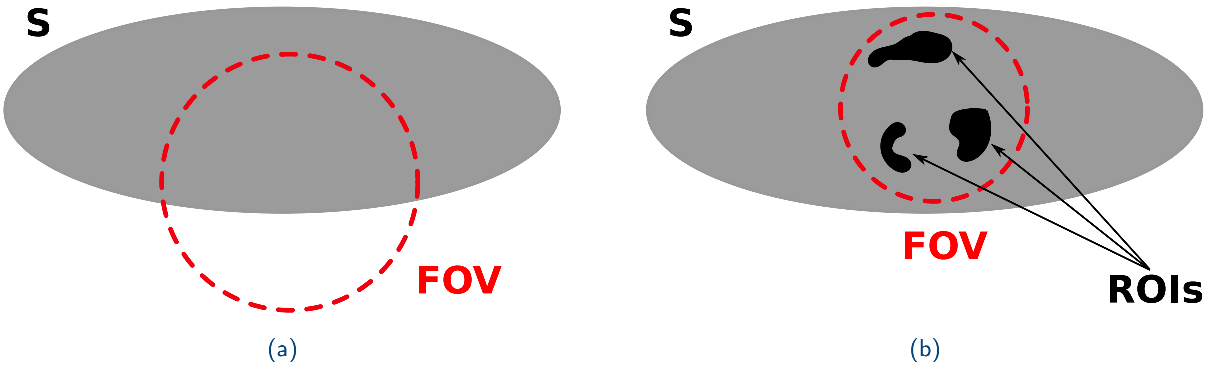


Figure 6.2: The results of [269] are valid for the kind of configurations shown in Fig. 6.1 were extended to (a) in [118] and (b) in [271–273].

and fan-beam geometries.

The connection between bowl artifact and ramp filtering in FBP reconstructions is well explained in [268]. The same study proposes extrapolation methods that do not require the knowledge of the object support. Consider a truncated projection $p(r)$ with $|r| \leqslant a < 1$ (the dependence on θ is dropped because the extrapolation methods have no dependence on the angle), the extrapolated projection, $p^{(e)}(r)$, according to one method results [268]:

$$p^{(e)}(r) \begin{cases} p(r) & -a \leqslant r \leqslant a \\ p(a) \cdot \cos^2\left(\dfrac{\pi(r-a)}{2a}\right) & a < r < 2a \\ p(-a) \cdot \cos^2\left(\dfrac{\pi(r+a)}{2a}\right) & -2a < r < -a \\ 0 & \text{otherwise} \end{cases} , \tag{6.7}$$

and according to the other method [268]:

$$p^{(e)}(r) \begin{cases} p(r) - \dfrac{p(a)+p(-a)}{2} & -a \leqslant r \leqslant a \\ \dfrac{p(a)-p(-a)}{2} \cdot \cos^2\left(\dfrac{\pi(r-a)}{2a}\right) & a < r < 2a \\ \dfrac{p(-a)-p(a)}{2} \cdot \cos^2\left(\dfrac{\pi(r+a)}{2a}\right) & -2a < r < -a \\ 0 & \text{otherwise} \end{cases} . \tag{6.8}$$

A simple but effective extrapolation method, when the object support is not known, is *edge-padding*, discussed in [263, 275, 276]. In this case, $p^{(e)}(r)$ results:

$$p^{(e)}(r) \begin{cases} p(r) & -a \leqslant r \leqslant a \\ p(a) & a < r < 2a \\ p(-a) & -2a < r < -a \end{cases} . \tag{6.9}$$

Reconstructions based on edge-padding can be regarded as quasi-exact, because they are defined up to an "almost" constant term. Consider an object $f = f(x, y)$ with support defined on the unit circle and $|\rho| < 1$ be the radius of the FOV for the INT scan. The FBP-E reconstruction, with edge-padding symmetrically extending in [-1,$-\rho$] and [ρ,1], is of the form:

$$f^{(\text{fbpe})}(x,y) = \int_0^{\pi} d\theta \int_{-\rho}^{+\rho} dt\, P(\theta,t) \cdot h(x\cos\theta + y\sin\theta - t)$$
$$+ \int_0^{\pi} d\theta \int_{-1}^{-\rho} dt\, P(\theta,-\rho) \cdot h(x\cos\theta + y\sin\theta - t)$$

$$
\begin{aligned}
&\qquad + \int_0^{\pi} d\theta \int_{\rho}^{+1} dt\, P(\theta,\rho) \cdot h(x\cos\theta + y\sin\theta - t) \\
&= \int_0^{\pi} d\theta \int_{-\rho}^{+\rho} dt\, P(\theta,t) \cdot h(x\cos\theta + y\sin\theta - t) \\
&\qquad + \int_0^{\pi} d\theta\, \left(P(\theta,-\rho) + P(\theta,\rho)\right) \underbrace{\int_{\rho}^{+1} dt\, h(x\cos\theta + y\sin\theta - t)}_{\text{const}} \\
&= \int_0^{\pi} d\theta \int_{-\rho}^{+\rho} dt'\, P(\theta,t') \cdot h(t-t') + \int_0^{\pi} d\theta\, C(\theta) \quad , \qquad (6.10)
\end{aligned}
$$

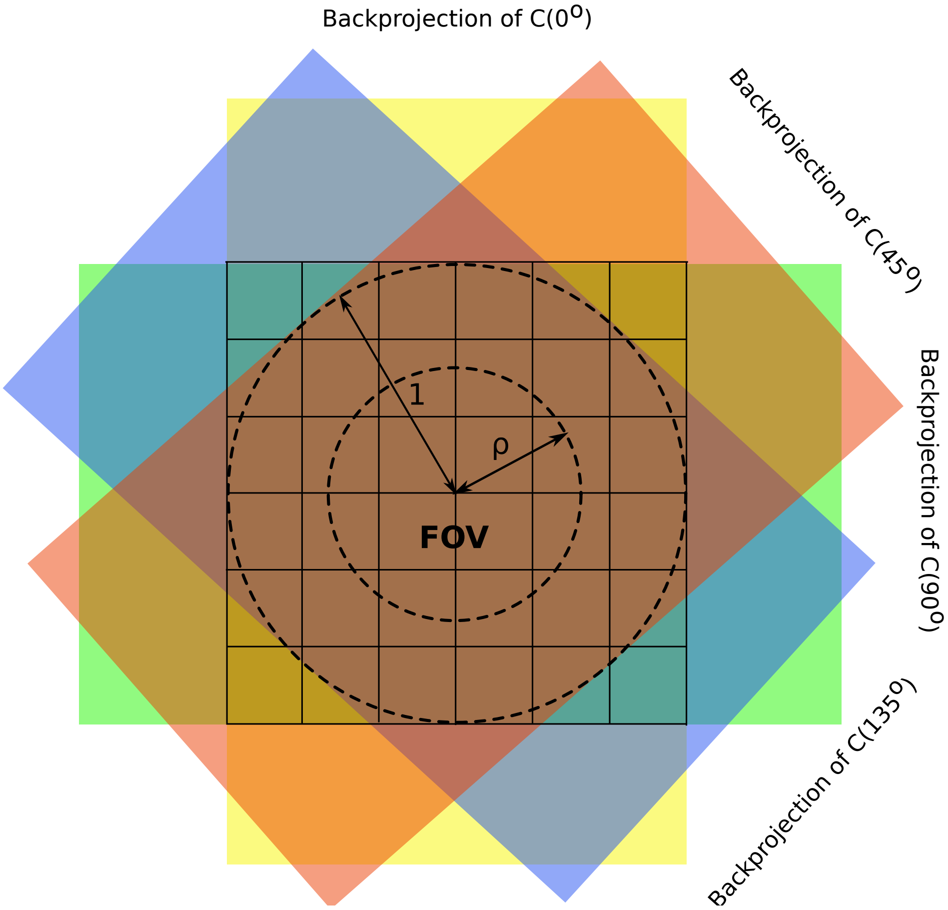


Figure 6.3: Sketch showing the backprojection of the term $C(\theta)$ in equation (6.10) along 4 directions 0, $\pi/4$, $\pi/2$ and $3\pi/4$. The central FOV of radius ρ is homogeneously covered by a constant value.

where $P(\theta, t)$ is the projection point at angle θ and distance t from the origin of the reference frame and $h(t)$ is the inverse Fourier transform of the ramp filter (an even function , i.e. $h(t) = h(-t)$). $C(\theta)$, backprojected from 0 to π, creates an "almost" constant term. To illustrate this statement, consider the sketch in Fig. 6.3: 4 constant projections at angles 0, $\pi/4$, $\pi/2$ and $3\pi/4$, each identified by a different color, are backprojected onto the image grid. The sketch shows that, when the edge-padding extends from [-1,$-\rho$] and [ρ,1], the unit circle and the FOV of radius ρ are covered by a constant value after backprojection of $C(\theta)$ along the given directions. When the edge-padding extends up to a radius $\rho < \rho^{(edge)} < 1$, the FOV is still constant (not the unit circle, this time). The conclusion is that analytical reconstructions of INT datasets performed with edge-padding are exact up to a constant.

6.3 Reconstruction of full INT data

The following section represents a reprinted manuscript published as: F. Arcadu, F. Marone and M. Stampanoni, *"Fast iterative reconstruction of data in interior tomography without a priori knowledge"*, to appear in Journal of Synchrotron Radiation, Novemeber 2016 [31].

6.3.1 Abstract

This paper introduces two novel strategies for iterative reconstruction of full interior tomography (FINT) data, i.e., when the field-of-view is entirely inside the object support and the knowledge of the object support itself or the attenuation coefficients inside specific regions-of-interest are not available. The first approach is based on data edge-padding. The second technique creates an intermediate virtual sinogram, which is, then, reconstructed by a standard iterative algorithm. Both strategies are validated in the framework of the alternate direction method of multipliers plug-and-play with gridding projectors that provide a speed-up of 3 orders of magnitude with respect to standard operators implemented in real space. The proposed methods are benchmarked on synchrotron-based X-ray microtomography datasets of mouse lung alveoli. Compared to analytical techniques, the proposed methods substantially improve the reconstruction quality for FINT underconstrained datasets, facilitating subsequent postprocessing steps.

6.3.2 Introduction

The term *interior tomography* (INT) refers to the problem of reconstructing an object function, when its support (S) is not a subset of the field-of-view (FOV) [92]. This imaging modality is broadly applied in medical screening, material science and biology, as it allows high-resolution investigations of small regions of interest (ROIs).

Since INT projections contain information $\notin$ FOV, reconstructions with filtered backprojection (FBP) suffer from a DC shift and low frequency artifacts [92], that compromise the quantitativeness of the results and make further postprocessing, visualization and rendering of the investigated object complicated [263].

Two different analytical reconstruction methods can address this problem: the differentiated backprojection (DBP) and FBP of edge-padded projections (FBP-E). The idea of DBP [269] is to backproject

the derivative of the data and, then, recover the original object by Hilbert transform techniques. This method provides quantitative reconstructions, when S is known and only specific geometries are involved: two opposite boundaries of the object are inside the FOV [269]; the FOV exceeds the object only from one side [118]; the FOV $\subset S$ and the attenuation coefficients are known in a subregion R $\subset$ FOV [271–273]. FBP-E [268, 275, 276] corresponds to FBP of projections, that have been padded with the edge pixels. This strategy prevents the appearance of low-frequency artifacts within the reconstructed FOV, although results are not fully quantitative.
Iterative methods have also been proposed for INT datasets, characterized by low photon statistics and/or with a low number of views(e.g. low-dose scan). Usually, these underconstrained datasets cannot be reconstructed with sufficient accuracy with analytical methods. The separable paraboloidal surrogate (SPS) [277], the maximum-likelihood expectation-maximization (EM-ML) [278] and the penalized weighted least square (PWLS) method [279] have been modified to deal with truncated projections. In [277], the regularized SPS is initialized through the projection-onto-convex-sets (POCS) [118]; the knowledge of S, the attenuation coefficients in one or multiple subregions and the position of the background pixels are needed for this reconstruction technique. In [278], the standard EM-ML is combined with thresholding, acting on pixels of selected ROIs. In [279], the PWLS is regularized with a complete high-quality scan of the same object. All these methods work for very specific (mostly medical) applications and require a priori knowledge obtained from previous scans. Iterative reconstruction of INT datasets without knowledge of either the object support or the attenuation coefficients inside specific ROIs (case of "full" interior tomography, abbreviated as FINT) has been initiated by [280]. In this work, the authors prove that an iterative scheme working with total variation regularization yields a unique reconstruction of a FINT dataset, in case the object under study is piecewise constant.
At synchrotron imaging beamlines, time resolved high resolution investigations of a large variety of dynamic systems spanning different fields (e.g. biology [2, 3], material sciences [4, 5], energy research [8, 9], earth and environmental sciences [6, 7]) are becoming routine. Due to the large variability of the examined samples and to the large datasets (several tens of TB) often associated to these experiments, efficient iterative reconstruction algorithms, not bounded to a specific application, are highly demanded.
This paper introduces two fast strategies for iterative reconstruction of FINT data, i.e., the FOV $\subset S$, S is unknown as well as the attenuation coefficients inside specific ROIs of the FOV. In this regard, the presented results can be considered a step further on the line of research launched by [280]. Differently from previous studies, this work shows the importance of data pre-processing (either edge-padding or differentiation) for analytical and iterative reconstruction of FINT datasets. The proposed iterative strategies are fast and provide high-quality, non quantitative reconstructions, best suited for subsequent post-processing and analysis steps (e.g. segmentation and morphological studies), independently of the imaged system.

Contributions

The contributions of this manuscript are summarized as follows:

- Analyze the importance of FINT data pre-processing through edge-padding and differentiation for analytical and iterative tomographic reconstruction.
- Introduction of two fast forward gridding projectors for iterative reconstruction of FINT data: one implements the derivative of the Radon transform, the other acts as Radon transform combined to edge-padding [28, 29].
- Introduction of the virtual strategy, as a more efficient alternative to specific forward, projectors for INT data. This strategy is independent from the chosen iterative scheme and regularization.
- Validation of the proposed methods within the framework of the alternate direction method of multipliers plug-and-play (ADMP) [230], on simulated and real datasets of full interior X-ray microtomography.

The edge-padding and virtual strategies studied in this work are neither bounded to the ADMP nor applicable only to the case of piecewise constant objects, but they can be utilized within any kind of iterative reconstruction algorithm and regularization scheme.

6.3.3 Reconstruction artifacts in interior tomography

Preliminaries

This work focuses on the reconstruction of a 2D slice from line projections acquired in parallel beam geometry. However, results can be generalized to more complex tomographic configurations (e.g., fan- and cone-beam).

The collection of line projections, measured at different angles $\theta \in [0, \pi)$, is called sinogram. The object to be reconstructed is a finite integrable real function, whose support, S, is assumed to be compact.

In interior tomography, the object support lies outside the field of view, i.e., $S \not\subseteq$ FOV and $S \cap$ FOV $\neq \emptyset$. Different INT configurations, listed in [118] and displayed in Fig.6.4, have been studied in literature. The case of full interior tomography (FINT), often characterizing microtomographic scans (Fig.6.4d, FOV $\subset S$), is considered in this work.

Case of analytical reconstruction methods

FBP reconstruction of a FINT sinogram is affected by a DC shift and low-frequency artifacts [268, 275, 276], due to the effect of the ramp filter on truncated projections [268]. To clarify this point, the example discussed in [268] is reproposed here.

Fig.6.5 shows a complete and truncated projection of a homogeneous circle before and after ramp filtering. In standard tomography (Fig.6.5 on the left), the filtered projection has a constant, positive profile in the middle with symmetric negative tails. Once all projections in $[0, \pi)$ are smeared back to the image grid, the circle is filled with constant pixel values and the negative tails of each projection are zeroed-out by the positive contributions from all other projections (Fig.6.6b). In interior tomography, the filtered projection (Fig.6.5 on the right) is instead everywhere positive, characterized by a bowl-shaped profile and tails with very high values. After backprojection, these positive tails are still present, because no negative compensation in the reconstruction process is possible, yielding the bowl artifact displayed in Fig.6.6c. The DC shift and bowl-shaped profile, shown in Fig.6.6d, can severely compromise the interpretation/morphological analysis of reconstructed FINT datasets.

Case of iterative reconstruction methods

Although iterative algorithms do not involve explicit ramp filtering, reconstructions of FINT data are affected by the same artifacts as for analytical methods.
Fig.6.7 shows an iterative reconstruction with SPS of the same sinogram used for Fig.6.6(c) and the corresponding line profile along the central row. The image is once again affected by a DC shift and a bowl artifact (Fig.6.7(b)). The reconstruction with any other iterative algorithm would present the

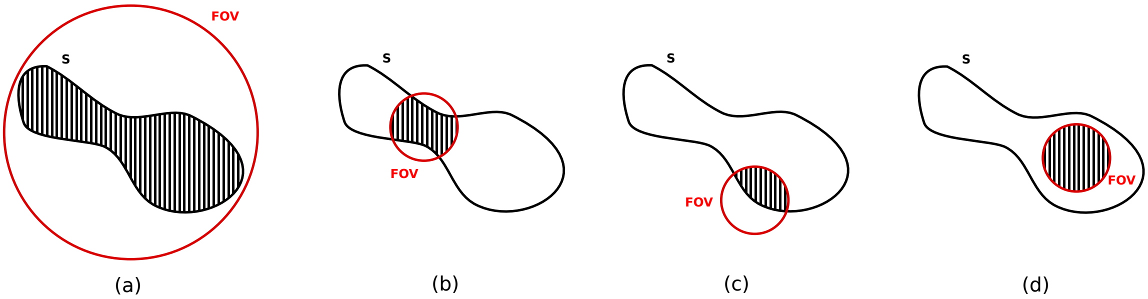


Figure 6.4: Different interior tomography configurations. The black line represents the boundary of the object support S; the red line delimits the FOV. The striped area identifies $S \cap$ FOV. (a): Standard tomography. (b): INT with two opposite sides of $\partial S \subset$ FOV [269]; [279]. (c): INT with the FOV exceeding S only from one side [118]. (d): Full INT (FINT), where FOV $\subset S$ [273, 277, 278, 280]. This work focuses on the latter configuration.

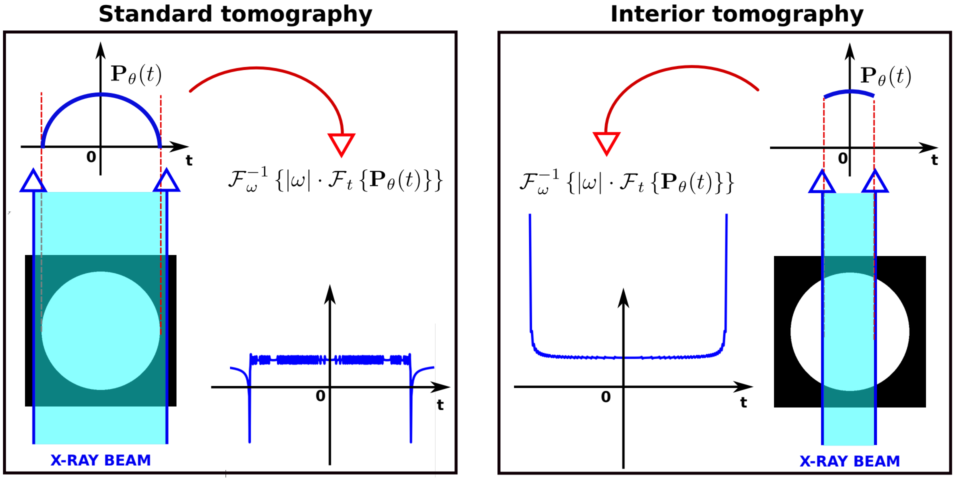
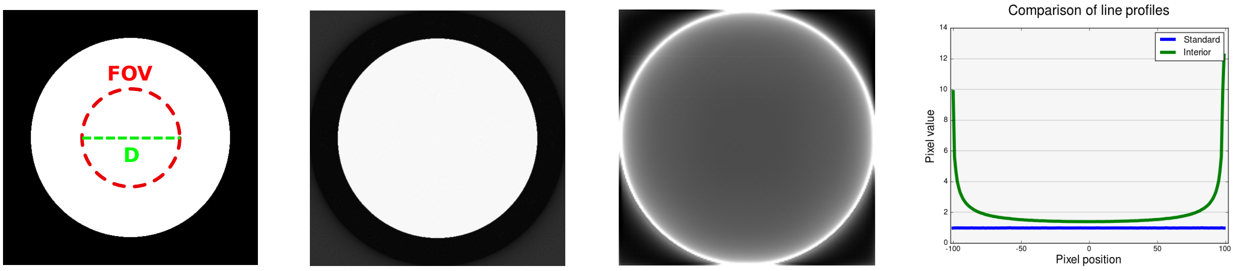


Figure 6.5: Illustrative example of the effect of the ramp filter on a single projection in case of standard and interior tomography. Notation: $\mathcal{F}_t$ ($\mathcal{F}_\omega^{-1}$) is the (inverse) Fourier transform along the variable t (ω); $|\omega|$ is the ramp filter; $P_\theta(t)$ is the projection in parallel beam geometry acquired at angle θ.



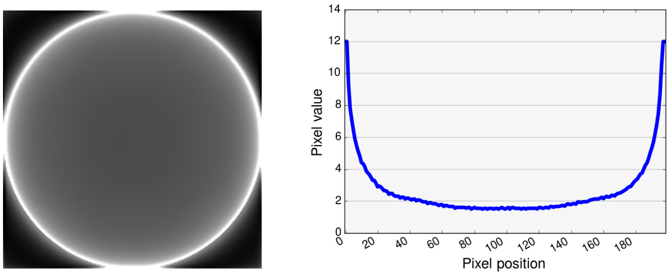
(a) (b) (c) (d)

Figure 6.6: (a) Simulated homogeneous circle: the background pixels are set to 0.0, the pixels inside the circle are set to 1.0. The image size is 512x512 pixels. The dashed red circle identifies the FOV for the interior tomography case. (b) FBP reconstruction of a sinogram with 800 views × 512 pixels in standard tomography; (c) FBP reconstruction of a sinogram with 800 views × 200 pixels in interior tomography with FOV depicted in (a). The image is zoomed to have the same dimensions of (a) and (b). (d) Comparison of the line profiles along the segment D, indicated in (a), for the reconstructions shown in (b) (blue line) and (c) (green line).

same problems.
This observation can be explained by the fact that iterative methods "mimic" the effect of the ramp filter: the reconstruction after the first iteration corresponds to a blurred version of the object, similar to the result of backprojection when omitting the filtering step; after several iterations, the object becomes sharper, until a similar spatial resolution as for the FBP reconstruction is reached. This "mimicking" behavior leads to the same problems characterizing FBP, when dealing with FINT data.

6.3.4 Proposed approach

The proposed approach builds upon existing methods, extended and combined to address the FINT problem. Although here for illustration purposes specific projectors, iterative scheme and regularization have been chosen, the proposed strategy is more general and not bounded to these choices.



(a) (b)

Figure 6.7: (a) Iterative reconstruction with SPS of the FINT sinogram used also for Fig.6.6(c). (b) Profile of the central line in the iterative reconstruction (a).

Flexible iterative reconstruction scheme

The alternate direction method of multipliers plug-and-play (ADMP) [230] offers an optimization framework, where the reconstructive and the regularization tasks are neatly separated in two subproblems. This structure allows the direct use of any forward projector for the reconstructive subproblem and any denoising method for the regularization subproblem. The ADMP can be viewed as a particular case of the classical formulation of the alternate direction method of multipliers [27].

In parallel beam geometry, the tomographic problem for a piecewise constant object has the following form:

$$\begin{gathered}\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\, R(\mathbf{x}) \;\Big|\; \mathbf{A}\mathbf{x}_i = \mathbf{b}_i \quad \text{for } i = 0, \cdots, n_z - 1 \\ \mathbf{x} = \left(\mathbf{x}_0^T, \cdots, \mathbf{x}_{n_z-1}^T\right)^T \in \mathbb{R}^{n_z \times n \times n} \quad \mathbf{x}_i \in \mathbb{R}^{n \times n} \quad \mathbf{A} \in \mathbb{R}^{m \times n, n \times n} \quad \mathbf{b}_i \in \mathbb{R}^{m \times n}\end{gathered} \tag{6.11}$$

where m is the number of views, n_z is the number of slices, n is the number of pixels along a row or a column (assuming the image grid to be square), $\mathbf{x}$ is the unknown 3D object, $\mathbf{x}_i$ is the unknown i-th slice, $\mathbf{A}$ is the forward projection operator ($\mathbf{A}^\dagger$ is the adjoint operator or backprojector), $\mathbf{b}$ is the sinogram and R is the functional enforcing the object to be piecewise constant. The dual variable $\mathbf{u}$ is introduced to transform the constrained optimization problem (6.11) into:

$$\underset{\mathbf{x},\mathbf{u}}{\operatorname{argmin}}\, F(\mathbf{x}, \mathbf{u}) = \underset{\mathbf{x},\mathbf{u}}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=0}^{n_z-1} \|\mathbf{A}\mathbf{x}_i - \mathbf{b}_i\|_2^2 + \lambda R(\mathbf{u}) \right\} \quad \text{subject to } \mathbf{u} = \mathbf{x} \tag{6.12}$$

The constraint $\mathbf{u} = \mathbf{x}$ is incorporated in the functional by augmenting F with an array of multipliers $\boldsymbol{\gamma}$. In this way, the following Lagrangian is obtained:

$$\underset{\mathbf{x},\mathbf{u},\boldsymbol{\gamma}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\gamma}) = \underset{\mathbf{x},\mathbf{u},\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=0}^{n_z-1} \|\mathbf{A}\mathbf{x}_i - \mathbf{b}_i\|_2^2 + \lambda R(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{x} - \mathbf{u} + \boldsymbol{\gamma}\|_2^2 \right\} . \tag{6.13}$$

The original problem (6.11) is, therefore, mapped into the minimization of the Lagrangian (6.13). The ADMP solves (6.13) by cyclically minimizing two subproblems with respect to the variable $\mathbf{x}$ and $\mathbf{u}$ and by updating the multipliers $\boldsymbol{\gamma}$ until convergence:

$$\begin{cases} \mathbf{x}^{(k+1)} \longleftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}\left(\mathbf{x}, \mathbf{u}^{(k)}, \boldsymbol{\gamma}^{(k)}\right) \\ \mathbf{u}^{(k+1)} \longleftarrow \underset{\mathbf{u}}{\operatorname{argmin}}\, \mathcal{L}\left(\mathbf{x}^{(k+1)}, \mathbf{u}, \boldsymbol{\gamma}^{(k)}\right) \\ \boldsymbol{\gamma}^{(k+1)} \longleftarrow \boldsymbol{\gamma}^{(k)} + \left(\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)}\right) \end{cases} \tag{6.14}$$

where the superscript $^{(k)}$ refers to the k-th iteration of a selected variable. The 1st and 3rd term of (6.13) will contribute to the $\mathbf{x}$-subproblem, whereas the 2nd and the 3rd term to the $\mathbf{u}$-subproblem. To explicitly derive the form of the $\mathbf{x}$-subproblem, the gradient of $\mathcal{L}$ with respect to $\mathbf{x}$ is calculated:

$$\mathbb{R}^{n \times n} \ni \mathbf{0} = \nabla_{\mathbf{x}_i} \mathcal{L}\left(\mathbf{x}, \mathbf{u}^{(k)}, \boldsymbol{\gamma}^{(k)}\right) = \nabla_{\mathbf{x}_i} \left\{ \frac{1}{2} \sum_{i=0}^{n_z-1} \left(\|\mathbf{A}\mathbf{x}_i - \mathbf{b}_i\|_2^2 + \mu \|\mathbf{x}_i - \mathbf{u}_i^{(k)} + \boldsymbol{\gamma}_i^{(k)}\|_2^2 \right) \right\}$$

$$
\begin{aligned}
&= \mathbf{A}^{\dagger}\left(\mathbf{A}\mathbf{x}_i - \mathbf{b}_i\right) + \mu\left(\mathbf{x}_i - \mathbf{u}_i^{(k)} + \boldsymbol{\gamma}_i^{(k)}\right) \\
&= \left(\mathbf{A}^{\dagger}\mathbf{A} + \mu\mathbf{I}\right)\mathbf{x} - \left(\mathbf{A}^{\dagger}\mathbf{b} + \mu\left(\mathbf{u}_i^{(k)} - \boldsymbol{\gamma}_i^{(k)}\right)\right) \\
&= \tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}} \quad .
\end{aligned}
\tag{6.15}
$$

Since $\tilde{\mathbf{A}}$ is symmetric and positive-definite, the conjugate gradient (CG) technique [224] is adopted to iteratively find the solution of the system $\tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{b}}$. For the $\mathbf{u}$-subproblem, it results that:

$$
\begin{aligned}
\underset{\mathbf{u}}{\operatorname{argmin}}\ \mathcal{L}\left(\mathbf{x}^{(k+1)}, \mathbf{u}, \boldsymbol{\gamma}^{(k)}\right) &= \underset{\mathbf{u}}{\operatorname{argmin}}\left\{\lambda R(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{x}^{(k+1)} - \mathbf{u} + \boldsymbol{\gamma}^{(k)}\|_2^2\right\} \\
&= \underset{\mathbf{u}}{\operatorname{argmin}}\left\{\frac{1}{2}\|\mathbf{u} - \tilde{\mathbf{u}}\|_2^2 + \frac{\lambda}{\mu}R(\mathbf{u})\right\} \quad .
\end{aligned}
\tag{6.16}
$$

The last term in (6.16) corresponds exactly to a denoising problem, where $\tilde{\mathbf{u}}$ represents the input noisy image and $\tau = \lambda/\mu$ is the regularization strength. The form of the $\mathbf{u}$-subproblem gives the "plug-and-play" qualification to the ADMP, as the type of denoising can be changed without altering the structure of the entire algorithm.

The iterative procedure is stopped when the L_2-norm of the relative difference between reconstructions of subsequent iterations, $\mathbf{x}^{(k)}$ and $\mathbf{x}^{(k+1)}$, is smaller than a certain threshold, i.e., $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2^2/\|\mathbf{x}^{(k)}\|_2^2 < \epsilon = 0.01$. One of the main advantages of ADMM-type methods is that satisfactory results in terms of image quality can be achieved after very few iterations.

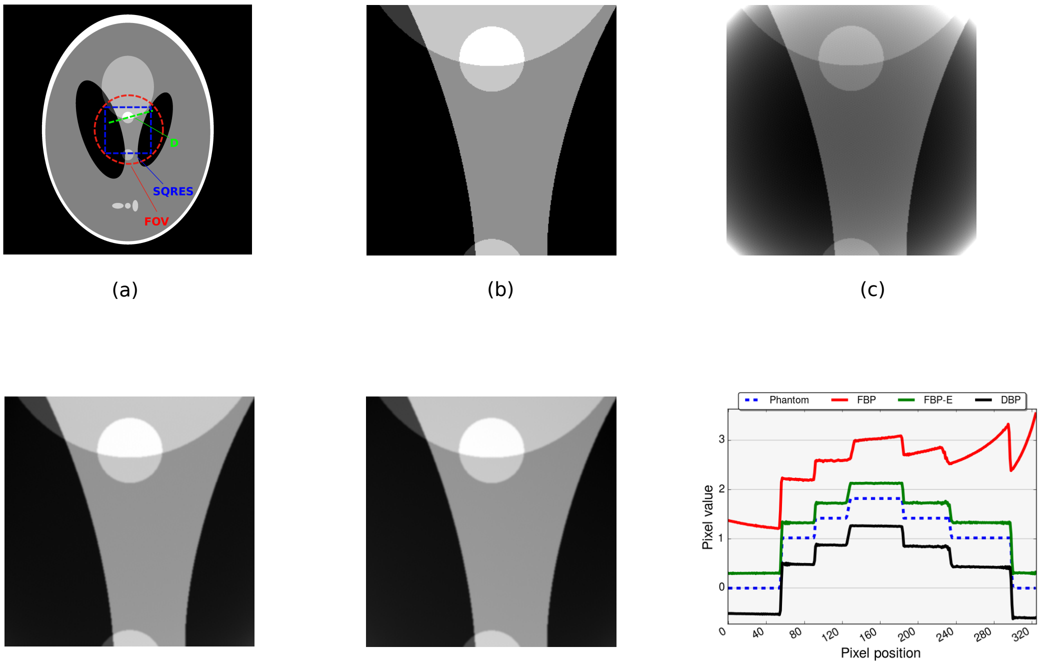
For the experiments presented in Sec.6.3.5, ADMP with split Bregman total variation (SBTV) (Goldstein & Osher, 2009) as plug-and-play regularization has been used. τ strongly determines the quality of the iterative reconstruction since it controls the trade-off between spatial resolution and noise removal.

Fast forward projectors for interior tomography

The DC shift compromises the quantitativity of FINT reconstructions, but, except for an offset constant throughout the tomographic slice, the information is preserved. The bowl artifact leads, instead, to non-quantitative reconstructions characterized by varying gray level values within homogeneous regions and prevents, therefore, any reliable morphological analysis, unless "decupping" algorithms or sophisticated segmentation approaches are utilized.

Two analytical methods can be adopted to avoid this bowl artifact: DBP and FBP-E. Reconstructions with these techniques are non-quantitative, i.e., exact up to an unknown constant, but can be safely used for image analysis if the knowledge of the attenuation coefficients in an absolute sense is not needed, as it is often the case. Fig.6.8 shows the reconstruction of an FINT sinogram created from a Shepp-Logan (SL) phantom using FBP, DBP and FBP-E. The original SL with 2048 × 2048 pixels (Fig.6.8a) is forward projected over 800 angles in $[0, \pi)$ and only the central 512 pixels of the sinogram are retained. The bowl artifact is visible at the corners of the FBP reconstruction (Fig.6.8c) and in a line profile (Fig.6.8f red). This artifact is instead not present in the DBP and FBP-E reconstructions (Fig.6.8d,e and Fig.6.8f green, black).

For the iterative reconstruction of FINT data, the proposed idea is to use tomographic forward ope-



(a) (b) (c)

(d) (e) (f)

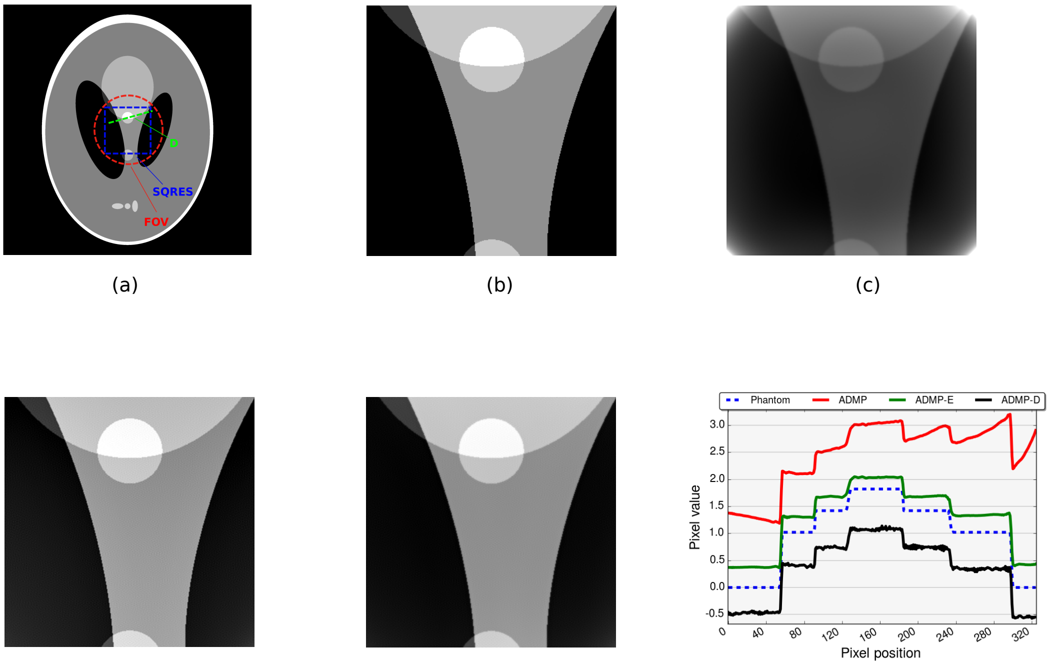
Figure 6.8: (a) Complete SL phantom. (b) SQRES of the SL phantom. (c) SQRES of the FBP reconstruction. (d) SQRES of the DBP reconstruction. (e) SQRES of the FBP-E reconstruction. (f) Line profiles along segment D. The line profiles in (f) show that DBP and FBP-E successfully remove the bowl artifact. The line profiles are manually shifted along the vertical axis to better distinguish one from another.

rators based on the principles of DBP and FBP-E, i.e., a projector that implements the derivative of the Radon transform and a Radon transform acting on an edge-padded projections, respectively.
To guarantee fast reconstructions, the forward gridding projector (FGP) [28], that works in the Fourier space and has a complexity of $O(N^2 \log_2 N)$, has been selected for this work. Studies conducted in [28] showed that the FGP is significantly faster than standard space-based projectors (complexity of $O(N^3)$) [93] and its accuracy results comparable to that of very sophisticated operators, when used in iterative schemes.
In standard tomography, the FGP works on an oversampled grid created with zero-padding. For FINT data, the following modifications of the original forward gridding projector are proposed: (i) the differentiated FGP (DFGP) [29], that implements the derivative of the Radon transform (still works on an oversampled grid created with zero-padding) and computes a differential sinogram; (ii) the FGP combined with edge-padding (FGP-E), i.e., the oversampled grid is created by edge-padding of the object. If these 2 operators are used within the ADMP framework, two iterative schemes can be defined: the ADMP-D implementing the DFGP and the ADMP-E implementing the FGP-E.
As a proof of principle, the experiment in Fig.6.8 is repeated with the standard ADMP, ADMP-D and ADMP-E. The bowl artifact, appearing in the reconstruction with the standard ADMP (Fig.6.9(c)), is not present in the images computed with ADMP-D (Fig.6.9(d)) and ADMP-E (Fig.6.9(e)).

Edge-padding has two main advantages with respect to the differentiated operator. First, the computation of the differential sinogram by finite differentiation is very likely to enhance the noise affecting the original sinogram; if a sophisticate technique is used instead (like the Savitzky-Golay filter [281]), additional parameters, ruling the trade-off between noise and spatial resolution, are added to the reconstruction problem. This argument is valid for both analytical and iterative reconstructions. Second, the number of sub-iterations required by the CG step in the ADMP is related to the conditioning number (CN) of the operator $\mathbf{A}$: for the ADMP-D, $\mathbf{A}$ corresponds to the derivative of the Radon transform; for the ADMP-E, $\mathbf{A}$ is the Radon transform itself, having a lower CN than its derivative. It has been experimentally observed that to reach convergence with the same number of iterations, the CG loop of the ADMP-D requires at least 15 sub-iterations, whereas 4 are enough for the CG loop of the ADMP-E. For these reasons, only the ADMP-E is utilized for the experiments discussed in Sec.6.3.5.

Iterative virtual method

For the iterative reconstruction of FINT data without a priori knowledge on the object support or on the attenuation coefficients inside specific ROIs, an alternative four-step strategy is also proposed: the iterative virtual method (abbreviated as ADMP-V). The steps of the ADMP-V are (Fig.6.10):



(a) (b) (c)

(d) (e) (f)

Figure 6.9: (a) Complete SL phantom. (b) SQRES of the SL phantom. (c) SQRES of the ADMP reconstruction. (d) SQRES of the ADMP-D reconstruction. (e) SQRES of the ADMP-E reconstruction. (f) Line profiles along segment D. The line profiles in (f) show that ADMP-D and ADMP-E successfully remove the bowl artifact. The line profiles are manually shifted along the vertical axis to better distinguish one from another.

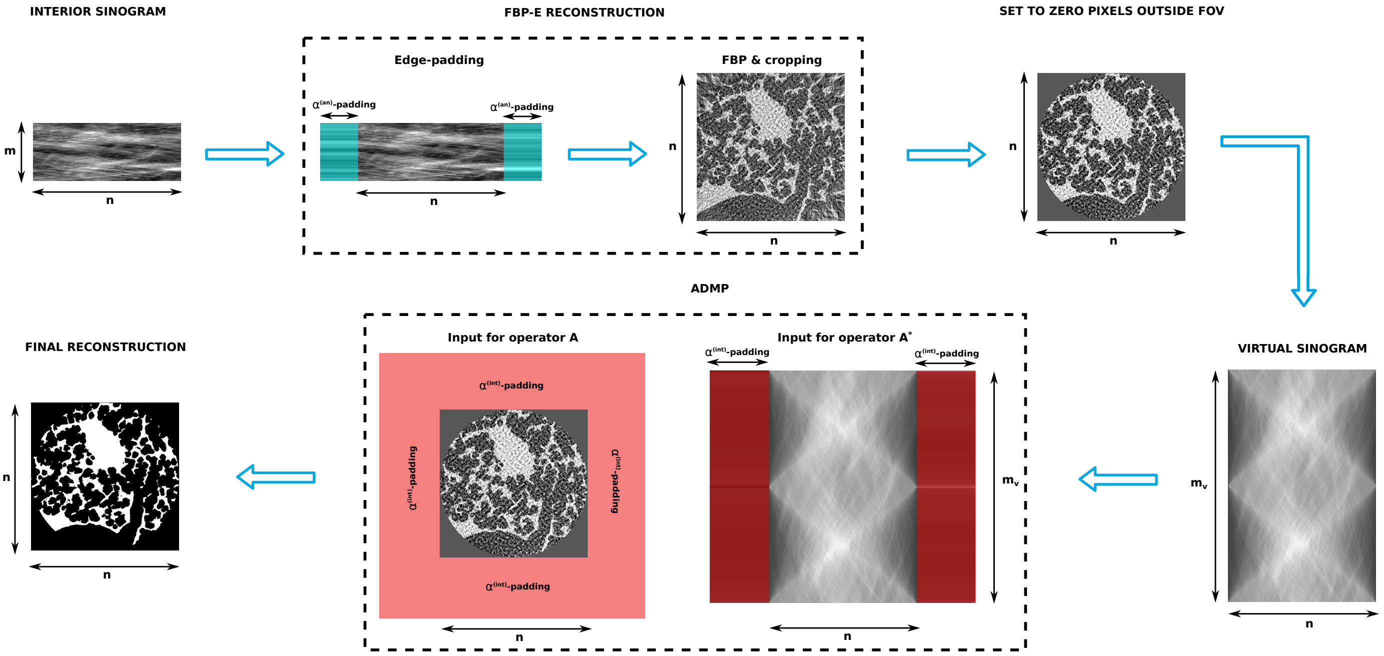


Figure 6.10: Diagram showing the steps of the virtual strategy (ADMP-V). (1) Reconstruction with FBP-E. (2) Zero-out all pixels $\notin$ FOV. (3) Forward projection to obtain the virtual sinogram ($m_v = n\pi/2$). (3) The virtual sinogram is used as input for the ADMP. The $\alpha^{((an))}$-padding used for FBP-E is colored in light blue, whereas the $\alpha^{((int))}$-padding inside ADMP is colored in light red.

(i) data reconstruction with an analytical method, like DBP or FBP-E; (ii) zeroing of all pixels outside the reconstruction circle; (iii) forward projection of this newly computed image to obtain a "virtual" sinogram, simulating a non-FINT dataset; (iv) reconstruction with ADMP using projectors for standard tomography (FGP, in this case) and using physical constraints (e.g. zeroing all pixels outside the reconstruction circle at each iteration).
After steps (i), (ii) and (iii), the initial FINT dataset is transformed into a standard tomographic dataset with known support, that can be reconstructed with any analytical or iterative algorithm.

6.3.5 Experiments

Analysis framework

Metrics to assess the reconstruction accuracy are computed over the square inscribed inside the reconstruction circle (SQRES).
The contrast-to-noise ratio (CNR) is defined as:

$$\mathrm{CNR} = \frac{|S_{r_{i1}} - S_{r_{i2}}|}{\sigma_{r_{i1}} + \sigma_{r_{i2}}} \quad , \tag{6.17}$$

where S_{r_i} and σ_{r_i} are the mean and standard deviation of the ROI r_i, assumed to be homogeneous. For the computation of the CNR, r_{i1} and r_{i2} must be neighboring. The values of CNR, reported in the tables of Sec.6.3.5, represent averages over multiple ROIs at different distances from the image center.

In case of simulated data, the peak signal-to-noise ratio (PSNR) [144] and the mean structural similarity index (MSSIM) [145] are also used as figures of merit. Since results are not quantitative due to the DC shift, a linearly regressed version, I_{regr}, of the reconstruction, I, is used as input for the analysis. Calling O the phantom, I_{regr} is computed as:

$$I_{\text{regr}} = a \cdot I + b \quad \text{such that} \quad \|I_{\text{regr}} - O\|_2^2 \quad \text{is minimized} \tag{6.18}$$

where $a, b \in \mathbb{R}$. In this way, PSNR and MSSIM scores are not biased by the DC shift.

Edge profiles and histograms are also displayed to provide additional information about the spatial resolution and the segmentability of the images. Reconstructions are mapped to the interval [0,1] to facilitate the display and direct comparison of profiles and histograms.

6.3.6 Data and settings

The phantoms used to generate the simulated data are displayed in Fig.6.11. The first dataset is the SL phantom [129], which is often utilized to validate tomographic reconstruction algorithms. The second dataset is a segmented reconstruction of mouse lung tissue and it is labeled MLT. This phantom is evidently characterized by a high structural complexity and is related to the real data used in Sec.6.3.10. The procedure to create FINT sinograms from a reference image has been described in Sec.6.3.4.

The real datasets have been acquired at the TOMCAT beamline of the Swiss Light Source in the framework of the SNF project "In vivo study of lung physiology with fast X-ray tomographic microscopy" [1]. The three sinograms correspond to scans of mouse lung tissue at micrometer scale with an effective detector pixel size of 2.9 (MOUSE-1, MOUSE-2) and 1.1 (MOUSE-3) μm. The first two datasets consist of 901 Paganin phase-retrieved [77] projections and 2016 pixels, whereas the third one has 501 projections × 2016 pixels.

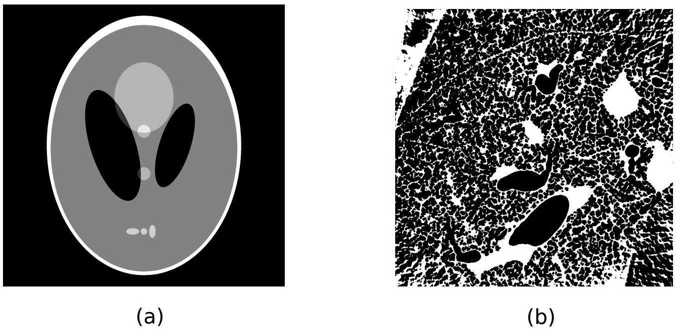

(a) (b)

Figure 6.11: Phantoms used to generate the simulated FINT sinograms. (a) SL phantom. (b) MLT phantom.

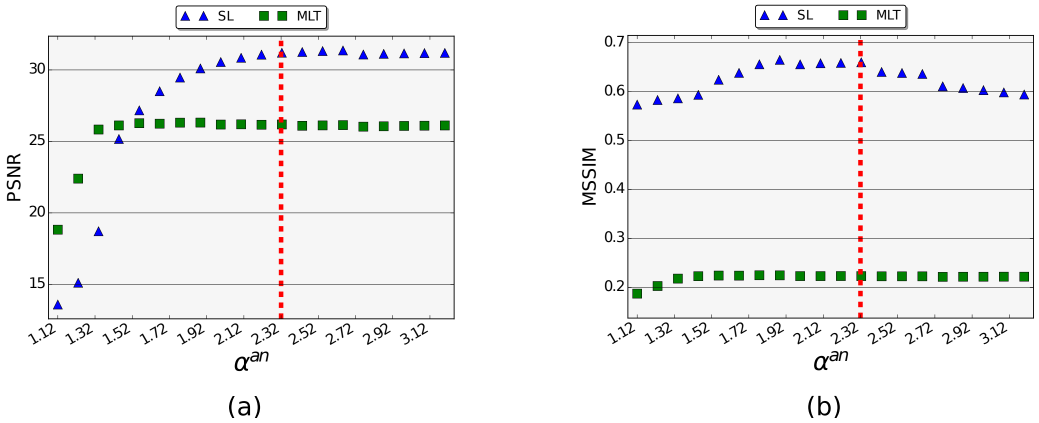


Figure 6.12: PSNR and MSSIM of the reconstructions of an FINT sinogram of SL and MLT phantoms computed by GRID-E as a function of $\alpha^{(\mathrm{an})}$, the edge-padding factor. The dashed red line locates the optimal oversampling factor, $\alpha^{(\mathrm{an})} = 2.32$.

The regularization strength τ (the only free parameter in the ADMP) is tuned so that reconstructions with ADMP-E and ADMP-V look visually as similar as possible.

6.3.7 Optimal edge-padding length for analytical and iterative reconstructions

The edge-padding length rules the trade-off between the reconstruction accuracy and computational efficiency for both analytical and iterative reconstruction methods. An insufficient amount of padding fails at removing the bowl artifact, whereas an excessive padding substantially increases the computational cost with insignificant gain in the reconstruction quality.

In this work, analytical reconstructions of FINT data are performed by means of the ramp-filtered adjoint FGP-E. The method, labeled GRID-E, is equivalent to FBP-E. To improve the signal-to-noise ratio (SNR) of the retrieved image, projections are additionally windowed with a Hamming function superimposed on the ramp filter. GRID-E is also used in the first step of the ADMP-V.

When performing analytical reconstructions with GRID-E, the edge-padding required for the INT dataset and the oversampling needed by the gridding backprojector coincide. The edge-padding or oversampling factor, i.e. the ratio between the number of pixels of edge-padded and original projections, used for GRID-E is indicated with $\alpha^{(\mathrm{an})}$ (the superscript stands for "analytical"). Fig.6.12 shows the PSNR and MSSIM scores as a function of $\alpha^{(\mathrm{an})}$ for the reconstructions of a SL and MLT sinograms with 1500 views × 512 pixels created from initial phantoms with 4096 × 4096 pixels. The smallest $\alpha^{(\mathrm{an})}$ corresponding to the highest values of both PSNR and MSSIM in Fig.6.12 is 2.32 (marked with a dashed red line).

An iterative reconstruction method utilizing the FGP-E as forward projector (like ADMP-E) depends on two distinct edge-padding factors: $\alpha^{(\mathrm{int})}$ (the superscript stands for "internal"), corresponding to the oversampling factor required by each call of the gridding implementations of $\mathbf{A}$ and $\mathbf{A}^{\dagger}$ for both standard and interior tomography; $\alpha^{(\mathrm{ext})}$ (the superscript stands for "external"), which is the edge-padding factor required to address the reconstruction of FINT datasets, can be considered a simple data extrapolation approach and is used for the entire duration of the iterative proce-

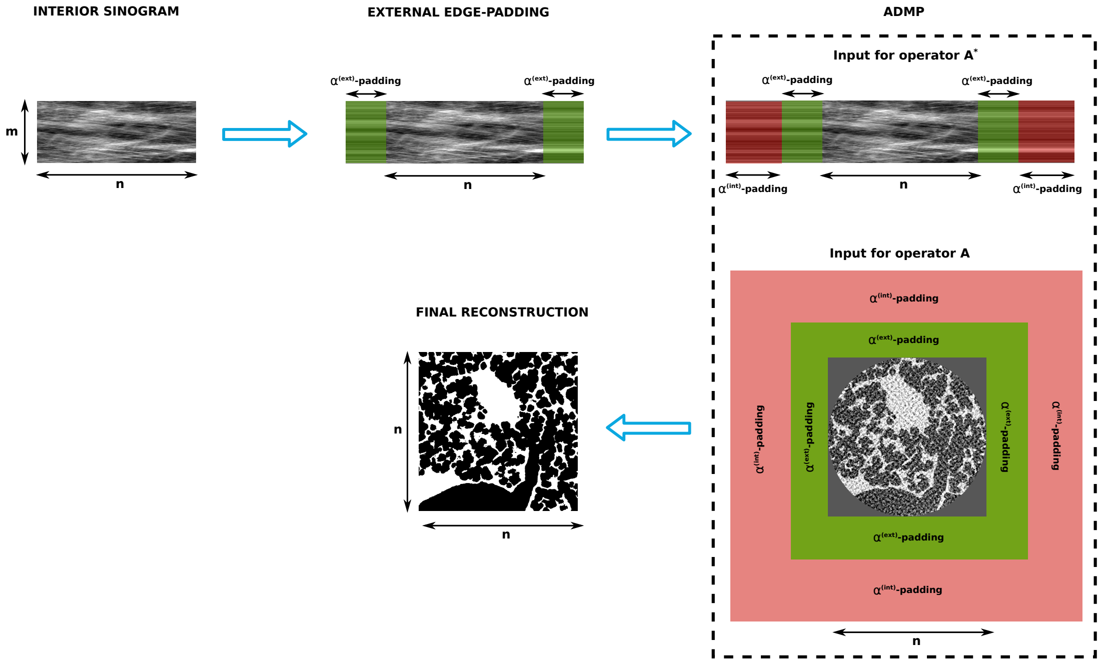
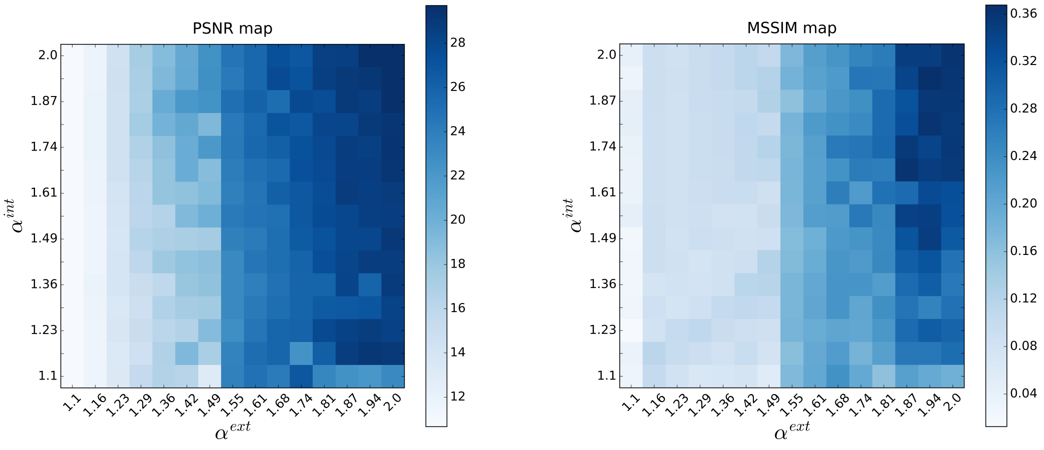


Figure 6.13: Diagram showing the difference between the two edge-paddings required by ADMP-E. The edge-padding by a factor $\alpha^{(\text{ext})}$ ($\alpha^{(\text{ext})}$-padding, indicated in the scheme with a light green color) is done before the start of the ADMP-E and remains for the entire duration of the iterative reconstruction. The edge-padding by a factor $\alpha^{(\text{int})}$ ($\alpha^{(\text{int})}$-padding, indicated in the scheme with a light red color) is required every time the forward gridding projector, $\mathbf{A}$, and its adjoint operator, $\mathbf{A}^\dagger$, are called inside the ADMP-E. The $\alpha^{(\text{int})}$-padding is summed up to the $\alpha^{(\text{ext})}$-padding when making use of $\mathbf{A}$ and $\mathbf{A}^\dagger$.

dure. A sinogram $\mathbf{b}$, reconstructed by ADMP-E, is first padded by a factor $\alpha^{(\text{ext})}$, becoming $\mathbf{b}'$ (β = num. pixels $\mathbf{b}'$/num. pixels $\mathbf{b}$). Then, every time $\mathbf{A}^\dagger$ (analogously it works for $\mathbf{A}$) is invoked inside the CG, $\mathbf{b}'$ is padded again by a factor $\alpha^{(\text{int})}$, becoming $\mathbf{b}''$. The second padding is only temporary (since it is a requirement of the gridding operators) and, once $\mathbf{A}^\dagger$ ($\mathbf{A}$) has completed its own calculation, the resulting image (sinogram) is cropped to remove the additional $\alpha^{(\text{int})}$-padding. The $\alpha^{(\text{ext})}$-padding, instead, remains for the entire run of the ADMP-E. The double-padding mechanism required by ADMP-E is showed in Fig.6.13.

Fig.6.14 shows the PSNR and MSSIM maps as a function of $\alpha^{(\text{int})}$ and $\alpha^{(\text{ext})}$ for the reconstruction of the same SL sinogram as in the experiment of Fig.6.12. The maps clearly point out that the reconstruction accuracy strongly varies with $\alpha^{(\text{ext})}$ and only in a weaker way with $\alpha^{(\text{int})}$. We selected $\alpha^{(\text{int})} = 1.7$ and $\alpha^{(\text{ext})} = 1.87$ as optimal edge-padding parameters, because they are the smallest values (highest computational efficiency) guaranteeing simultaneously maximum accuracy in terms of both PSNR and MSSIM.

The optimal edge-padding factors determined here are used for the following reconstructions with ADMP-V and ADMP-E. For ADMP-V, $\alpha^{(\text{an})} = 2.32$ is utilized inside GRID-E and $\alpha^{(\text{int})} = 1.7$ inside ADMP. ADMP-E utilizes $\alpha^{(\text{ext})} = 1.87$ as external and $\alpha^{(\text{int})} = 1.7$ an internal edge-padding factors.



(a) (b)

Figure 6.14: PSNR and MSSIM maps of the reconstructions of an FINT sinogram of a SL phantom computed by ADMP-E as a function of $\alpha^{(\text{int})}$ and $\alpha^{(\text{ext})}$.

6.3.8 Validation for different zoom-in factors

The reconstruction accuracy of the ADMP-E and ADMP-V has been tested for different zoom-in factors (ZIFs), defined as the ratio between the number of pixels of the sinogram in standard tomography and the corresponding FINT one. For example, given a sinogram in standard tomography with 4096 pixels, a ZIF=4 describes a FINT sinogram consisting of the central 1024 pixels of the original dataset.

Fig.6.15 shows reconstructions computed by the two iterative methods for increasing ZIFs. The original MLT sinogram has 1500 views × 4096 pixels. At visual inspection the reconstructions are not affected by FINT artifacts and look almost identical, showing that both methods can satisfactorily tackle the reconstruction of FINT data for different ZIFs.

The reconstructions with ZIF=8 (Fig.6.15(c),(f)) are overall characterized by a sort of background pattern. For the simulated data used here, this problem, starting to be barely visible with ZIF=8, can be neglected up to ZIF~32. This does not represent a problem for the real datasets used in subsection 6.3.10, as the highest ZIF is roughly 9.

6.3.9 Validation for different conditions of asymmetry

The accuracy of the ADMP-E and ADMP-V has been tested also for the reconstruction of FOVs placed at various distances from the phantom center. In this way, the iterative methods can be validated for different conditions of asymmetry, i.e., not symmetrical distribution of the attenuation coefficients around the selected FOV.

An example of such test, performed with the SL phantom, is shown in Fig.6.16. The sinograms for FOV-1 and FOV-2 have both 1500 views × 512 pixels. Once again, the reconstructions are not affected by FINT artifacts and look practically identical, proving that both methods can handle FINT cases with pronounced feature asymmetry around the FOV.

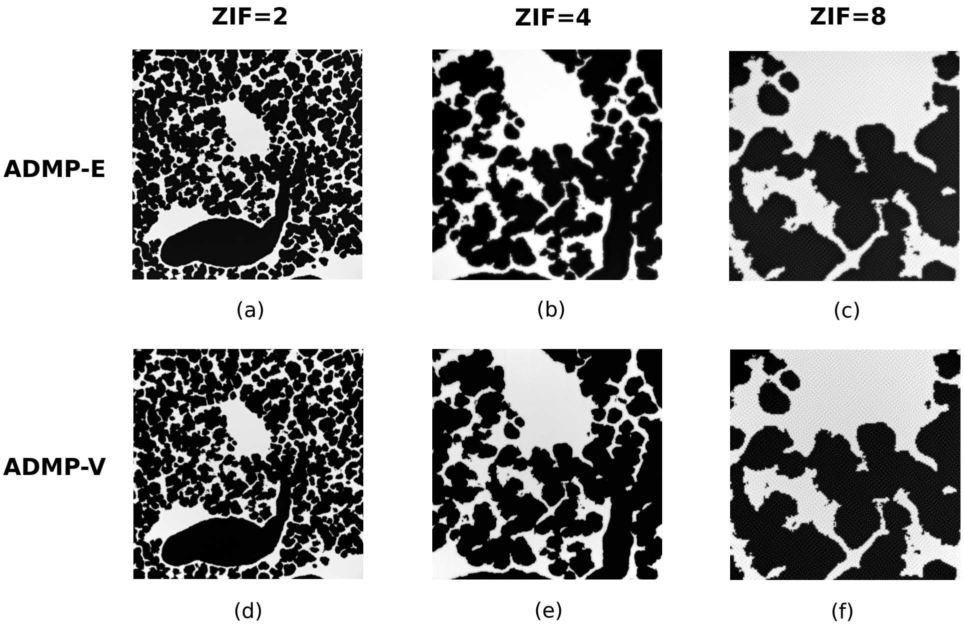


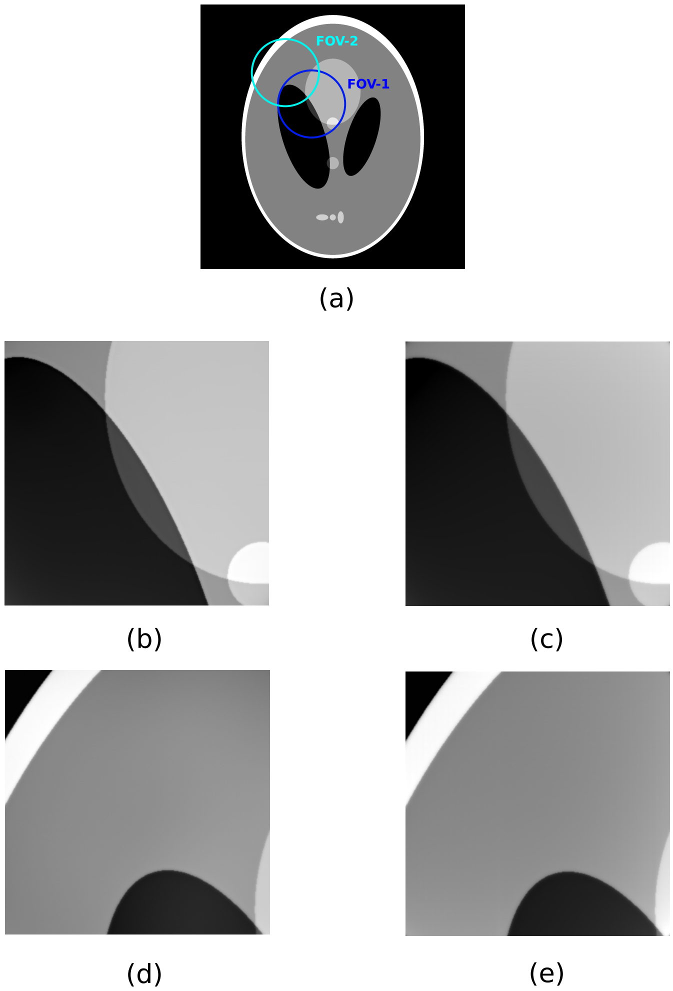
Figure 6.15: Reconstructions computed by ADMP-E and ADMP-V of MLT sinograms with increasing ZIF: 2, 4 and 8. (a) ADMP-E reconstruction: ZIF=2. (b) ADMP-E reconstruction: ZIF=4. (c) ADMP-E reconstruction: ZIF=8. (d) ADMP-V reconstruction: ZIF=2. (e) ADMP-V reconstruction: ZIF=4. (f) ADMP-V reconstruction: ZIF=8.

	GRID-E	ADMP-E	ADMP-V
MSSIM	0.026	0.047	0.045
PSNR	14.74	24.69	24.43
CNR	0.66	2.92	2.85

Table 22: PSNR, MSSIM and CNR computed for the reconstructions of SL (Fig.6.17) phantom.

6.3.10 Reconstruction of underconstrained sinograms

Iterative tomographic algorithms are mostly designed to address underconstrained datasets, providing insufficient direct information for accurate reconstruction with analytical methods. An underconstrained sinogram is either undersampled, noisy or a combination of both factors. A sinogram in standard tomography is considered undersampled if $m < n \cdot \pi/2$, where m is the number of views and n the number of detector pixels [12]. An FINT sinogram is undersampled when $m < n_s \cdot \pi/2$ with $n_s = n \cdot$ ZIF: n_s represents the number of detector cells required to "cover" the entire object in standard tomography and n the number of available detector cells from the FINT scan [263]. Since low-dose, fast scans of FINT microtomography are usually characterized by $m \ll n \ll n_s$, undersampling combined to local tomography artifacts result in a sort of "background texture", that can severely affect FINT reconstructions. Iterative reconstruction in standard tomography can greatly reduce the amount of this background texture through the usage of constraints and regularization. Since no



(a)

(b) (c)

(d) (e)

Figure 6.16: Reconstructions of two different FOVs computed by ADMP-E and ADMP-V. (a) SL phantom with FOV-1 and FOV-2. (b) ADMP-E reconstruction: FOV-1. (c) ADMP-V reconstruction: FOV-1. (d) ADMP-E reconstruction: FOV-2. (e) ADMP-V reconstruction: FOV-2.

constraints are available for the FINT datasets, the removal of the background texture relies entirely on the regularization: for the same amount of noise in the data, the regularization strength τ should be bigger, the higher is the undersampling factor, defined as UF $= (1 - m/(n \cdot \pi/2 \cdot \text{ZIF})) \cdot 100\%$. If τ is too low, the jump associated to the background texture can be regarded by the TV as a collection of faint edges to be preserved and the regularization will only remove the noise component.

The ADMP-E and ADMP-V are here tested for the reconstruction of underconstrained datasets against GRID-E. The regularization strength, τ, is separatly chosen for ADMP-E and ADMP-V to guarantee a higher quality reconstruction in terms of CNR (while maintaining a similar spatial resolution) compared to GRID-E.

In the first experiment, an FINT sinogram of the SL phantom with ZIF=4, 200 views × 512 pixels, UF $= 94\%$ and corrupted by Gaussian noise with $\sigma = 2.5\%$ of the sinogram mean is considered.

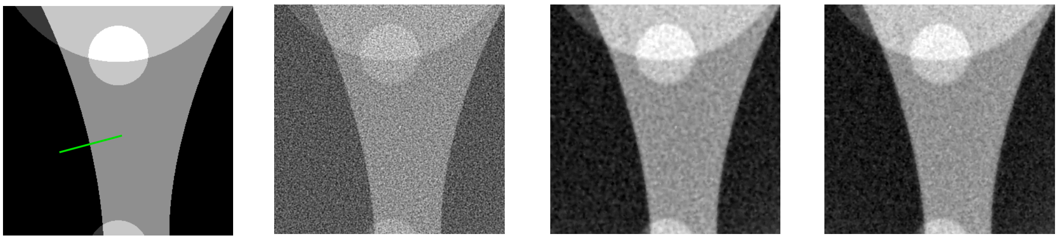
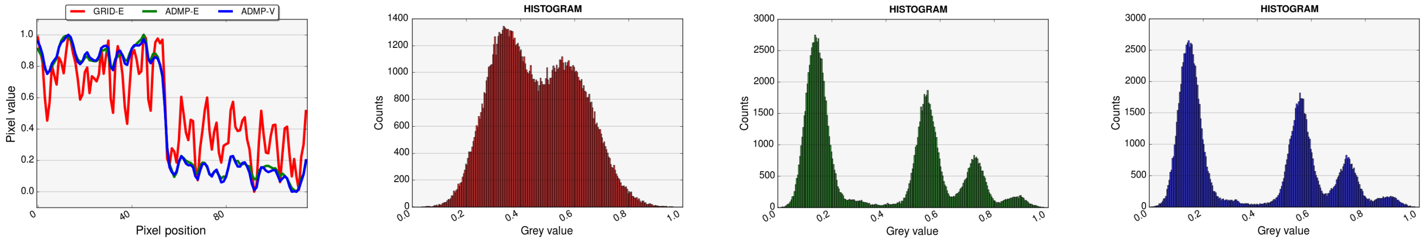

(a) (b) (c) (d)

Figure 6.17: Reconstructions computed by GRID-E (a), ADMP-E (b) and ADMP-V (c) of the SL sinogram with 200 views × 512 pixels, additional Gaussian noise with $\sigma = 2.5\%$ of the sinogram mean and ZIF=4. The green segment in (a) shows the position of the line profiles displayed in Fig.6.18(a). (a) SL phantom. (b) GRID-E reconstruction. (c) ADMP-E reconstruction. (d) ADMP-V reconstruction.



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

Figure 6.18: Edge profiles (a) and histograms (b,c,d) for the reconstructions with GRID-E, ADMP-E and ADMP-V shown in Fig.6.17.

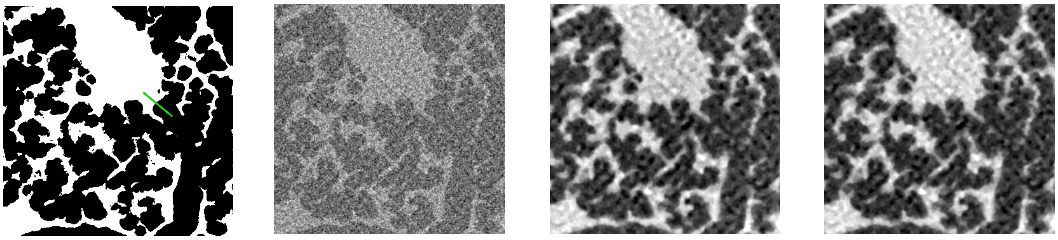
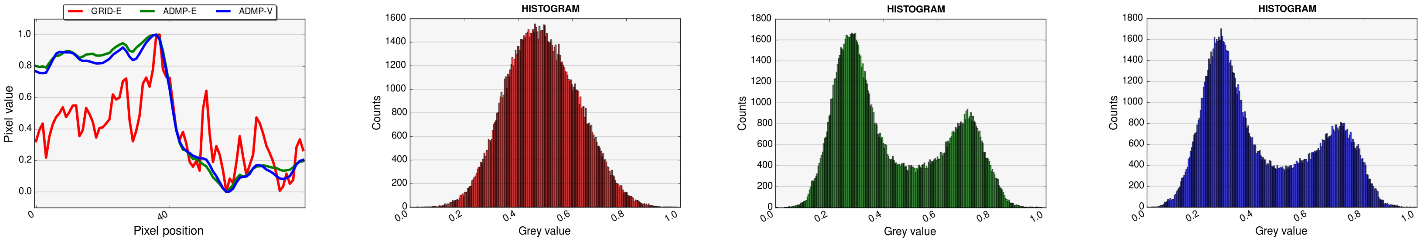

(a) (b) (c) (d)

Figure 6.19: Reconstructions computed by GRID-E (a), ADMP-E (b) and ADMP-V (c) of the MLT sinogram with 200 views × 512 pixels, additional Gaussian noise with $\sigma = 3.5\%$ of the sinogram mean and ZIF=4. The green segment in (a) shows the position of the line profiles displayed in Fig.6.20(a). (a) MLT phantom. (b) GRID-E reconstruction. (c) ADMP-E reconstruction. (d) ADMP-V reconstruction.

Fig.6.17 shows that the reconstructions computed by the ADMP-E and ADMP-V look more similar to the phantom compared to the GRID-E reconstruction. They have higher MSSIM, PSNR and CNR (increased by a factor of 4.4) scores, as reported in Tab.22. The edge profiles (Fig.6.18(a)) indicate



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

Figure 6.20: Edge profiles (a) and histograms (b,c,d) for the reconstructions with GRID-E, ADMP-E and ADMP-V shown in Fig.6.19.

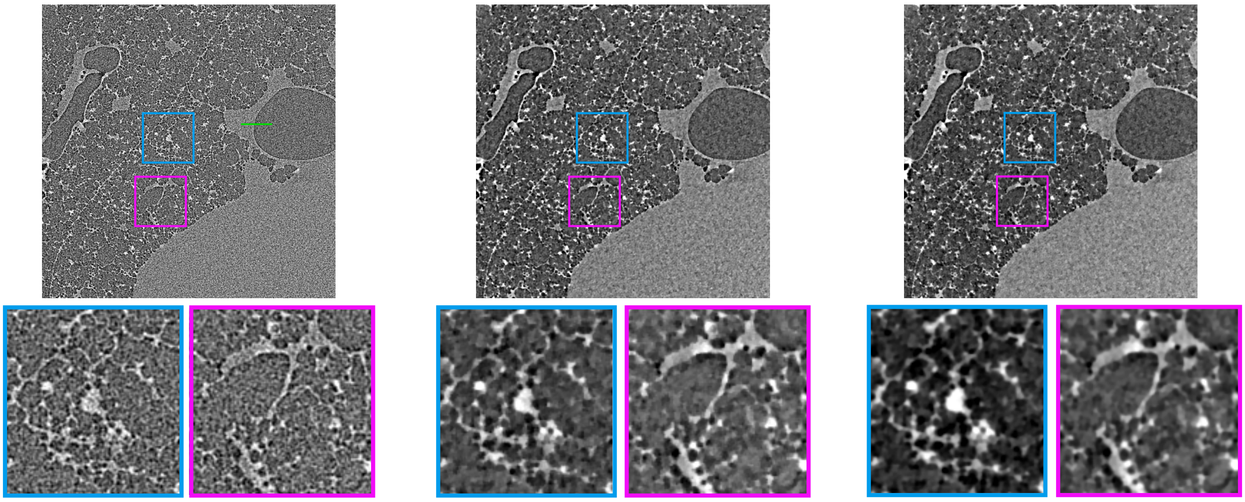
	GRID-E	ADMP-E	ADMP-V
MSSIM	0.101	0.135	0.135
PSNR	8.39	12.68	12.63
CNR	0.44	1.98	1.87

Table 23: PSNR, MSSIM and CNR computed for the reconstructions of MLT (Fig.6.19) phantom.

that the proposed iterative approaches lead to improved CNR without deterioration of the spatial resolution. Moreover, the peaks corresponding to the different phases of the SL phantom can be clearly identified in the histograms for the iterative reconstructions in Fig.6.18(c,d), whereas only two peaks appear in Fig.6.18(b) for the analytical reconstruction.

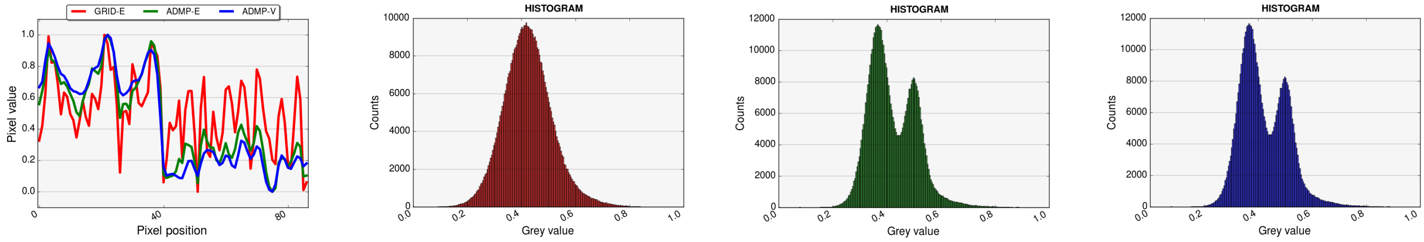
For the second experiment, a FINT sinogram of the MLT phantom with the same condition of undersampling considered in the previous case (200 views × 512 pixels, ZIF=4) and a larger amount of Gaussian noise (3.5% of the sinogram mean) is used. ADMP-E and ADMP-V substantially reduce the noise in the reconstructions displayed in Fig.6.19(c,d) compared to GRID-E (Fig.6.19(b)) making the light structures clearer against the dark ones. The metric scores in Tab.23 are largely higher for the iterative reconstructions, with an improvement of the CNR by a factor of nearly 4.4. The edge profiles (Fig.6.20(a)) for the three different reconstructions coincide almost perfectly at the edge position, demonstrating the efficacy of the split bregman total variation regularization in removing noise while preserving edges. The higher quality of the iterative reconstructions is also highlighted by the histograms in Fig.6.20(c,d) showing clear peaks corresponding to the two phases of the MLT phantom.

For the third experiment, the dataset of MOUSE-1, characterized by a ZIF$\approx$ 3.2, 901 views × 2016 pixels, UF = 91%, a pronounced asymmetry and highly absorbing structures (e.g. ribs) outside the FOV, is considered. Due to the feature size and the pattern complexity, the different quality of the reconstructions displayed in Fig.6.21 can be best appreciated in the blow-ups below the images. Features in the iterative reconstructions can be more easily identified thanks to the reduced noise and to the CNR increased by a factor of 5.3 (Tab.24). The line profiles at the edge position (Fig.6.22(a)) for the ADMP-E and ADMP-V results match that for the GRID-E reconstruction. Moreover, the double peak in Fig.6.22(c,d) shows that the two main phases of the lung tissue are better separated in



(a) (b) (c)

Figure 6.21: Reconstructions computed by GRID-E (a), ADMP-E (b) and ADMP-V (c) of the MOUSE-1 sinogram with 900 views × 2016 pixels. The green segment in (a) shows the position of the line profiles displayed in Fig.6.22(a). The blow-ups below each reconstruction are zoomins for two different ROIs.



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

Figure 6.22: Edge profiles (a) and histograms (b,c,d) for the reconstructions with GRID-E, ADMP-E and ADMP-V shown in Fig.6.21.

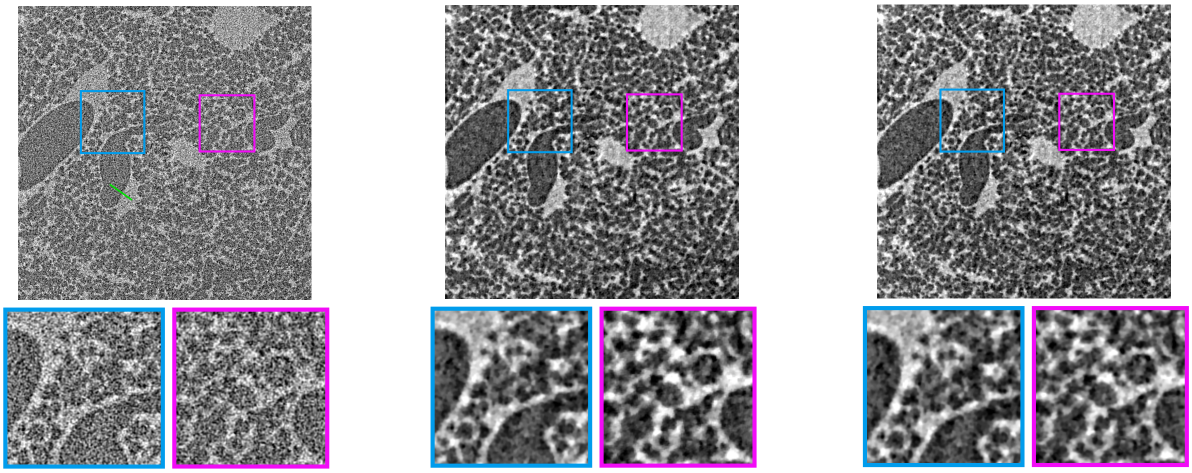
	GRID-E	ADMP-E	ADMP-V
CNR	0.16	0.83	0.85

Table 24: CNR scores computed for the reconstructions of the sinograms of MOUSE-1 (Fig.6.21).

	GRID-E	ADMP-E	ADMP-V
CNR	0.26	1.21	1.19

Table 25: CNR scores computed for the reconstructions of the sinograms of MOUSE-2 (Fig.6.23).

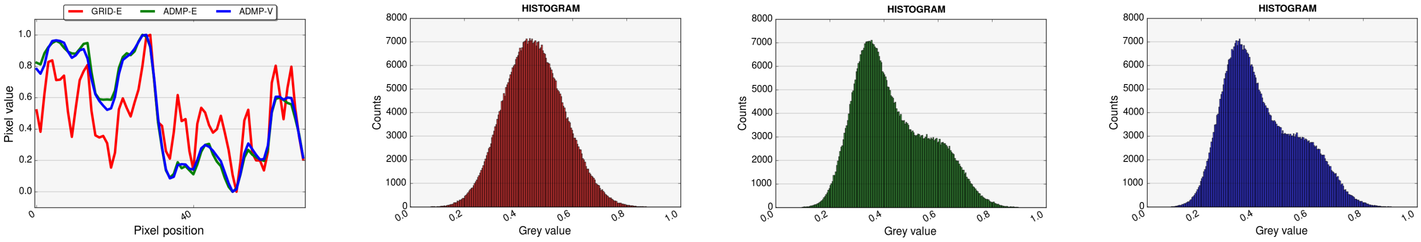
terms of gray level in the iterative reconstructions, whereas a single peak characterizes the histogram for the analytical reconstruction in Fig.6.22(b).



(a) (b) (c)

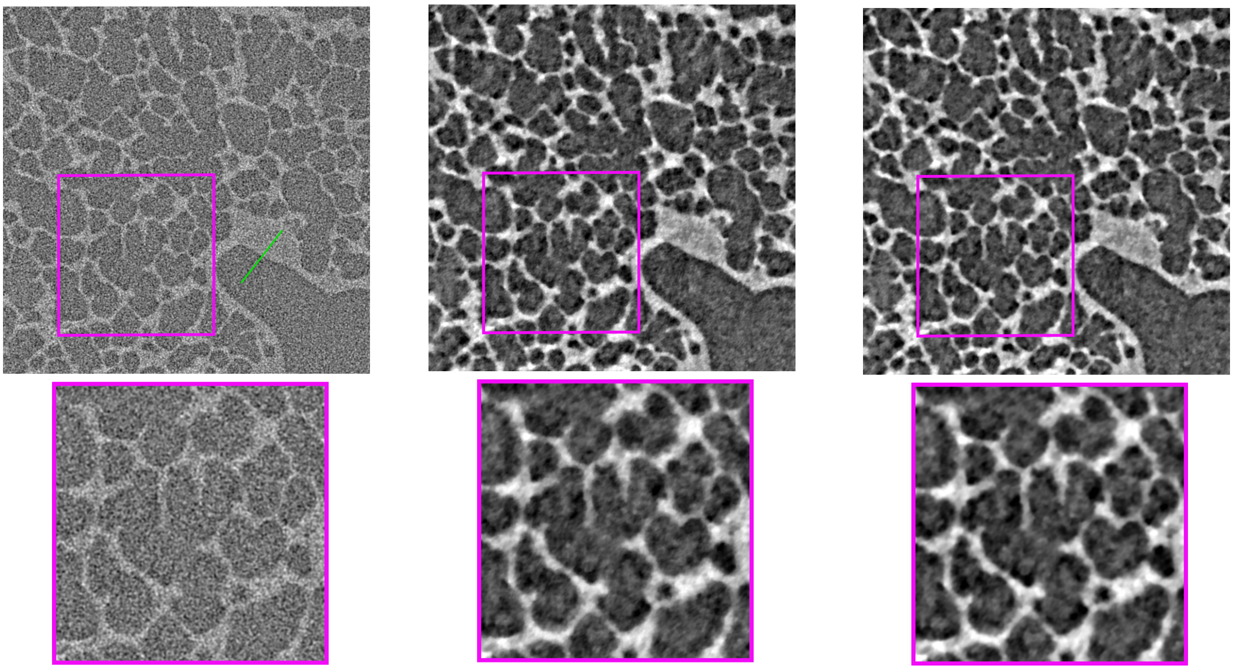
Figure 6.23: Reconstructions computed by GRID-E (a), ADMP-E (b) and ADMP-V (c) of the MOUSE-2 sinogram with 900 views × 2016 pixels. The green segment in (a) shows the position of the line profiles displayed in Fig.6.24(a). The blow-ups below each reconstruction are zoomins for two different ROIs.

In the fourth experiment, the reconstructed dataset of MOUSE-2 (901 views × 2016 pixels, ZIF≈ 3.2) has a very similar pattern complexity as MOUSE-1, whereas the morphology, e.g. the shape of the small structures, is quite different. Reconstructions with the ADMP methods in Fig.6.23 clearly show higher quality compared to the analytical one, thanks to the decreased noise level and the higher CNR (increased by a factor 4.6, Tab.25). Small dark features are well identifiable in Fig.6.23(b,c), whereas they are mainly covered by noise and undersampling/local tomography artifacts in Fig.6.23(a). The edge profiles in Fig.6.24 for the iterative reconstructions overlap almost exactly at the edge position with that for the analytical result, indicating also in this case that the total variation regularization operates with negligible loss in terms of spatial resolution. The histograms corresponding to the reconstructions with the ADMP-E and ADMP-V in Fig.6.24(c,d) show the presence of 2 phases. This is not the case in the histogram for GRID-E in Fig.6.24(b).



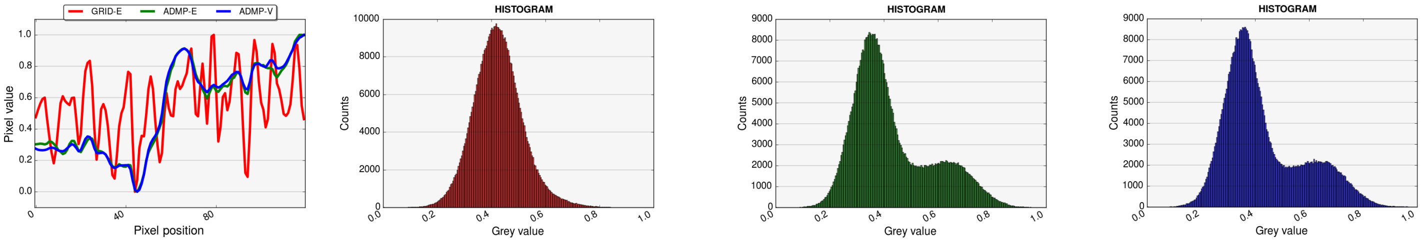
(a) (b) (c) (d)

Figure 6.24: Edge profiles (a) and histograms (b,c,d) for the reconstructions with GRID-E, ADMP-E and ADMP-V shown in Fig.6.23.



(a) (b) (c)

Figure 6.25: Reconstructions computed by GRID-E (a), ADMP-E (b) and ADMP-V (c) of the MOUSE-3 sinogram with 500 views × 2016 pixels. The green segment in (a) shows the position of the line profiles displayed in Fig.6.26(a). The blow-ups below each reconstruction are zoomins for a ROI.



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

Figure 6.26: Edge profiles (a) and histograms (b,c,d) for the reconstructions with GRID-E, ADMP-E and ADMP-V shown in Fig.6.25.

	GRID-E	ADMP-E	ADMP-V
CNR	0.24	0.88	0.89

Table 26: CNR scores computed for the reconstructions of the sinograms of MOUSE-3 (Fig.6.25).

For the fifth experiment, the dataset of MOUSE-3, characterized by 500 views × 2016 pixels and a ZIF$\approx$ 9.0, is used. In this case, UF = 98%, but features are in average much larger compared to MOUSE-1 and MOUSE-2, since projections were acquired with a higher magnification. The iterative reconstructions in Fig.6.25 offer a clearer vision of the object; nevertheless, all structures recognizable in Fig.6.25(b,c) can be visually identified in Fig.6.25(a) as well. The CNR is improved in this example by a factor of 3.7 (Tab.26(c)). In Fig.6.26(a) an edge can hardly be recognized for the reconstruction with GRID-E, whereas those corresponding to ADMP-E and ADMP-V, practically identical, do show a flank. Histograms in Fig.6.26(b,c) show a distinction between the two main phases of the lung tissue, whereas a single peak dominates in the histogram for the analytical reconstruction in Fig.6.26(a).

These five experiments with simulated and real sinograms show that: (i) ADMP-E it manages to substantially improve the image quality compared to an analytical method like GRID-E when tackling the reconstruction of underconstrained FINT datasets with different ZIF, noise level and feature complexity; (ii) the virtual strategy incorporated in the ADMP-V can provide comparable results to those achievable with the ADMP-E.

6.3.11 Computational cost

To give an idea of the superior computational performance of the ADMP-V compared to the ADMP-E, the time required for a single iteration, Δt, has been measured for two different datasets on an Intel(R) Core(TM) i7-3520M CPU 2.90GHz. For a sinogram with 800 views × 504 pixels (convergence reached after 8 iterations), Δt(ADMP-E) = 46.7s and Δt(ADMP-V) = 1.45s; for a sinogram with 1584 views × 1008 pixels (convergence reached after 9 iterations), Δt(ADMP-E) = 173.2s and Δt(ADMP-V) = 5.7s. Thus, the ADMP-V for small and medium datasets is approximately 30 times faster with respect to the ADMP-E.

6.3.12 Conclusion

This work introduces two novel strategies for iterative reconstructions of datasets in full interior tomography (FINT), when the FOV is completely inside the object support and no a priori knowledge regarding the support itself and the distribution of the attenuation coefficients in certain ROIs is available. FINT represents a very general case, frequently encountered in many tomographic applications like synchrotron-based X-ray microtomography.

One strategy works with an edge-padding forward projector. The second is a four-step procedure, requiring the creation of an intermediate virtual sinogram, simulating a full tomography dataset; this sinogram is, then, reconstructed by a standard iterative algorithm, while enforcing a tight constraint outside the reconstruction circle. Both strategies, although not bounded to a specific iterative scheme, have been implemented in this work inside the alternate direction method of multipliers plug-and-play (ADMP), that offers a versatile optimization framework, where different forward projectors and regularizers can be easily plugged in, without altering the structure of the iterative solver. The two resulting iterative methods have been labeled ADMP-E (edge-padding strategy) and ADMP-V (virtual strategy).

The forward gridding projector with minimal oversampling (FGP) is used as standard and edge-padding forward operator for the ADMP. The FGP guarantees fast iterative reconstructions, while keeping the same accuracy of the results achieved with more sophisticated, but much slower, implementations of the Radon transform.

The ADMP-E and ADMP-V have been, first, validated for the reconstruction of FINT datasets with different zoom-in factors and asymmetry conditions around the FOV. The methods have, then, been tested on underconstrained simulated and real FINT datasets. Results show that both iterative techniques yield reconstructions of higher quality compared to a standard analytical method: the CNR is greatly improved (in average 4 times higher), while preserving the spatial resolution, and small features can be more easily identified. The reconstruction quality achieved with the two proposed iterative strategies is comparable. ADMP-V provides, though, superior computational efficiency (about 30 times faster), since it requires a much smaller grid for the computations inside the iterative procedure.

Chapter 7

Sinogram inpainting

"Inpainting" is an image processing technique designed to replace a set of labeled pixels with the information provided by unlabeled neighboring pixels [282]. Inpanting is often used to remove specific features from an image without compromising its continuity, as illustrated by the example in Fig. 7.1 Zooming can be considered a special case of inpainting, since the geometry of the pixels to reconstruct is well-defined. Consider a zooming by a factor k and a low-resolution image $x^{(\mathrm{lr})}[i, j]$, the target high-resolution image, at first, is $x^{(\mathrm{hr})}[ki, kj] = x^{(\mathrm{lr})}[i, j]$ and 0 otherwise. The inpainting or zooming algorithm uses $x^{(\mathrm{hr})}[ki, kj]$ to infer the values of all the other pixels.

Sinogram inpainting is here referred as the procedure to upsample the sinogram along the view direction, creating new projections at angles between the ones already available. The inpaiting geometry is, therefore, a bundle of horizontal equidistant lines, whereas for image zooming it corresponds to

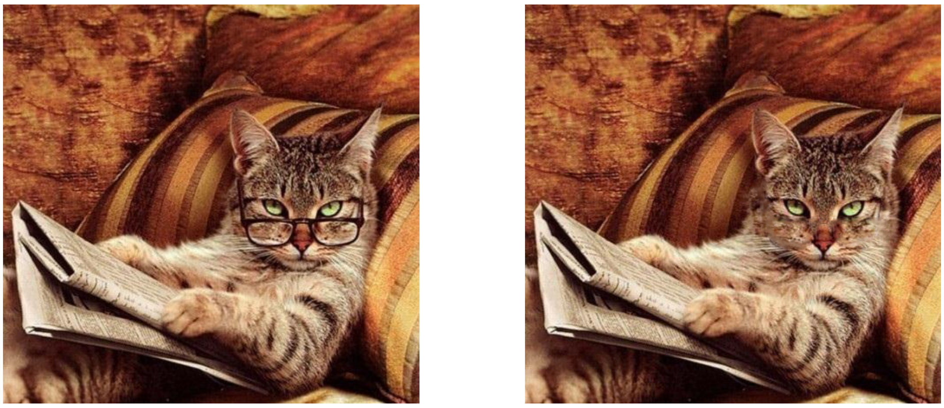

(a) Original image (b) Inpainted image

Figure 7.1: Example of image inpainting. (a) In the original image the cat wears glasses. (b) The pixels corresponding to the glasses are labeled and then replaced through inpainting with information coming from the surroundings. The performance of an inpanting algorithm is visually measured by the degree of compatibility between the inpainted pixels and the neighbors.

a pattern of both horizontal and vertical lines to be inferred. Increasing the number of views of an undersampled sinogram has the potential of reducing the streak artifacts affecting its FBP reconstruction.

The image processing literature offers a wide variety of different algorithms for inpanting or zooming, that theoretically could be directly applied to the specific case of sinogram inpanting. Many of such techniques work well with natural images, yielding satisfactory visual results, but they are not suited for sinograms: visually acceptable inpainted sinograms can unexpectedly lead to FBP reconstructions affected by severe artifacts. The problem is that neighboring pixels of a sinogram do not necessarily belong to the same local area, as it occurs for natural images, thus, the principle ruling almost any inpainting/zooming algorithm cannot be applied to this case. Here is a list of different methods for image inpainting/zooming that have been tested in this work for tomographic datasets: Papoulis-Gerchberg [283, 284]; matrix extrapolation [285]; frequency-selective extrapolation [286]; Criminisi inpainting [287]; Navier-Stokes based inpainting [288]; fast marching inpainting [289]; image zooming through directional cubic convolution interpolation [290].

Specific techniques for sinogram interpolation/extrapolation have been proposed since the 90s, but are usually designed to address the "missing-wedge" problem or the extrapolation of projections out of the FOV. These algorithms are also iterative and require several input parameters. Using an iterative technique purely working in the Radon domain has the advantage of being much more computationally efficient than iterative reconstruction algorithms making use of tomographic operators. The drawback is that iterative algorithms for sinogram restoration combined to FBP cannot outperform iterative reconstruction algorithms in terms of accuracy. It is for this reason that iterative algorithms for sinogram restoration have never taken over a leading position in the research on tomographic reconstruction.

7.1 Sinogram filter based on the consistency conditions

The following section represents a reprinted manuscript published as: F. Arcadu, J. Vogel, M. Stampanoni and F. Marone, submitted to Fundamenta Informaticae, 2016 [32].

7.1.1 Abstract

This work introduces and characterizes a fast parameterless filter based on the Helgason-Ludwig consistency conditions, used to improve the accuracy of analytical reconstructions of tomographic undersampled datasets. The filter, acting in the Radon domain, extrapolates intermediate projections between those existing. The resulting sinogram, doubled in views, is then reconstructed by a standard analytical method. Experiments with simulated and real data prove that the peak-signal-to-noise ratio of the results computed by filtered backprojection is improved up to 5–6 dB, if the filter is used prior to reconstruction.

7.1.2 Introduction

The Radon transform [291] of a function belonging to the space of rapidly decreasing C^∞ functions on $\mathbb{R}^2$ satisfies the Helgason-Ludwig consistency conditions [292, 293] (HLCC). These properties

characterize any k-th moment of the Radon transform. The consistency conditions, known since the 1960s [292], have been mainly exploited in iterative pre-processing algorithms (prior to reconstruction) to restore limited-angle tomographic datasets [277, 294–296], to reduce translational motion artifacts in fan-beam computed tomography (CT) reconstructions [297] or metal artifacts [298], and for alignment in cardiac position emission tomography (PET) and CT [299].
The algorithm presented in this work utilizes the HLCC for a different purpose, namely, augmenting the number of views of a sinogram with projections homogeneously acquired in $[0, \pi)$ to improve its reconstruction with an analytical technique like filtered backprojection [300] (FBP). This augmentation strategy should provide better analytical reconstructions especially for undersampled datasets, i.e., $m \ll n\pi/2$, where m is the number of views and n the number of detector pixels [12]. Iterative algorithms, often utilized for the reconstruction of strongly undersampled datasets, are characterized by a high computational cost, a rather large hyper-parameter space and need for sample-specific constraints and/or a priori knowledge. The proposed algorithm, instead, acting as a filter that preprocesses the sinogram before the actual reconstruction, is parameterless, sample-independent and fast (only slightly impacting the total reconstruction speed).
The use of this HLCC-based filter substantially improves the quality of FBP reconstructions of undersampled datasets. However, the reconstruction quality provided by highly optimized iterative algorithms, set with finely tuned hyper-parameters and sample-specific a priori knowledge, is not yet achieved. Nonetheless, the proposed method represents an effective strategy when sufficient computational power (e.g. access to high performance computing facility) is not available or high variability of the samples requires time consuming integration of a priori knowledge and hyper-parameters optimization for each single investigated object.

7.1.3 Helgason-Ludwig consistency conditions

The HLCC are properties characterizing the Radon transform, $\mathcal{R}$, of a 2D function $f(\mathbf{x})$ with bounded support, Ω, entirely placed within the field-of-view (FOV), i.e., $\Omega \subset$ FOV. In parallel beam geometry, the Radon transform is defined as [12]:

$$p(\theta, t) = \mathcal{R}\{f(\mathbf{x})\}(\theta, t) = \int_{\Omega} d\mathbf{x}\, f(\mathbf{x})\, \delta(\mathbf{x} \cdot \hat{\mathbf{n}} - t) \quad , \tag{7.1}$$

where $\mathbf{x} = (x_1, x_2)$, $\hat{\mathbf{n}} = (\cos\theta, \sin\theta)$, θ is the projection angle formed with the x_1-axis and t is the distance of the X-ray line from the origin of the reference frame.
The HLCC state that the integral:

$$\int_{-1}^{+1} dt\, t^k\, p(\theta, t) \tag{7.2}$$

is a homogeneous polynomial of degree k in $\cos\theta$ and $\sin\theta$ for $k \geqslant 0$ [292, 293]. The zeroth-order condition ($k = 0$), e.g, corresponds to [294]:

$$\mu = \int_{-1}^{+1} dt\, p(\theta, t) = \int_{\mathbb{R}^2} d\mathbf{x}\, f(\mathbf{x}) \quad \forall\, \theta\,, \tag{7.3}$$

meaning that the integral of any projection along t is constant and equals the integral of f over $\mathbb{R}^2$. For $k \geqslant 1$, conditions related to higher moments of $p(\theta, t)$ are obtained.
An effective way to enforce the HLCC is by expanding $p(\theta, t)$ onto a Fourier-Chebyshev basis [295]:

$$p(\theta, t) = \frac{1}{\pi} \sum_{k=0}^{\infty} \sum_{l=-\infty}^{+\infty} b_{kl}\, (1 - t^2)^{1/2}\, U_k(t)\, e^{il\theta} \quad , \tag{7.4}$$

where $U_k(t)$ represents the k-th order Chebyshev polynomial of second kind, the b_{kl}'s are the Fourier Chebyshev coefficients and $\theta \in [0, 2\pi]$ and i is the imaginary unit. On this orthogonal basis, the HLCC become [295]:

$$b_{kl} = 0 \quad \text{for} \quad \begin{cases} |l| > k \\ k + |l| = 2z + 1 \quad \forall\, z \in \mathbb{N} \end{cases} . \tag{7.5}$$

If the Radon transform is sampled over 2π, the positions of the non-zero b_{kl} form in the (k, l)-space a checkboarded wedge as shown in Fig. 7.2.
Consistency conditions have been extended to fan-beam [301] and cone-beam [302] geometry as well.

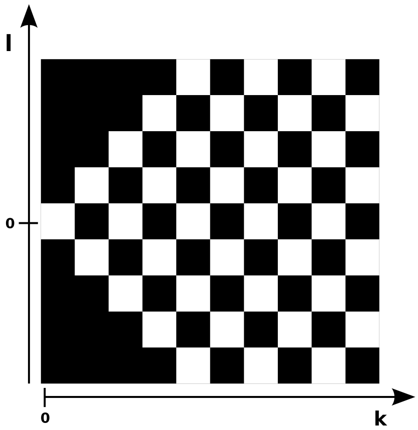


Figure 7.2: Checkboarded wedge pattern of the b_{kl} coefficients in the (k, l)-space. The non-zero b_{kl}'s are placed in the white squares, the rest is zero.

(a) (b)

Figure 7.3: Example of a sinogram interleaved with 0-valued lines (a) and the binarized real part of its Fourier-Chebyshev decomposition (b), showing that inconsistencies give rise to non-zero b_{kl}'s at the locations indicated in (7.5).

7.1.4 Proposed method

The discrete version of $p(\theta, t)$ (also called sinogram) is indicated with $p[\theta_h, t_j]$, where $\theta_h = h\pi/m \in [0, \pi)$ for $h = 0, 1, ..., m-1$ and $t_j = -1 + 2j/(n-1)$ for $j = 0, 1, ..., n-1$, considering $t_j \in [-1, 1]$. The proposed Helgason-Ludwig sinogram filter (HLSF) doubles the number of views of a sinogram $\in \mathbb{R}^{m \times n}$ by extrapolating m missing projections at intermediate angles $\theta_{h+1/2} = (\theta_h + \theta_{h+1})/2$. The input sinogram is, first, interleaved with 0-valued projections, as shown in Fig. 7.3(a). In this way, it becomes an inconsistent dataset characterized by several non-zero b_{kl}'s at the locations indicated in (7.5) (Fig. 7.3(b)). By enforcing (7.5), consistency is recovered, the 0-valued lines are filled with extrapolated values and the resulting sinogram $\in \mathbb{R}^{2m \times n}$ can be reconstructed by means of an analytical method like FBP.
HLSF consists of the following four steps:

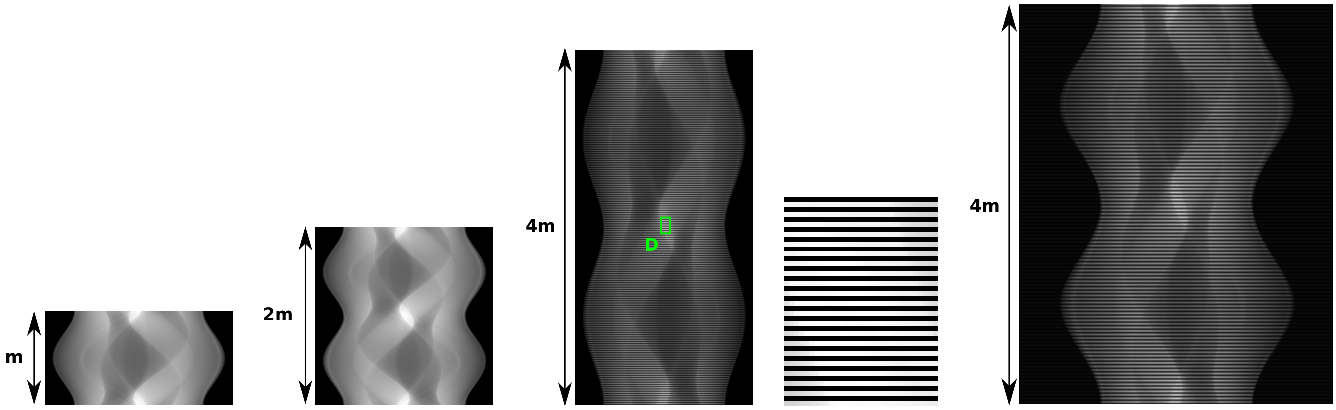
1. extension of the data to $[0, 2\pi]$;

2. creation of an intermediate sinogram with $4m$ views; the $2m$ original projections are interleaved with $2m$ zero lines;

3. imposition of the HLCC on the intermediate inconsistent sinogram;

4. crop of the interval $[0, \pi)$ and reassignment of the original m views.

"Imposing HLCC" means to enforce the property expressed by (7.5), i.e. zero-out all $b_{kl} \neq 0$, that are supposed to be zero.



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

Figure 7.4: Succession of steps to compute the Fourier-Chebyshev decomposition of a sinogram. (a) Input sinogram of a Shepp-Logan phantom, $\theta \in [0, \pi)$. (b) Extended sinogram, $\theta \in [0, 2\pi]$. (c) Interleaved sinogram. (d) Zoom-in of the region D in (c). (e) Sinogram with cosine-resampled projections.

7.1.5 Step (1): extend sinogram to $[0, 2\pi]$

First, the input sinogram, $p[\theta_h, t_j]$, has to be extended to $[0, 2\pi]$ in order to use (7.4):

$$p^{(1)}[\theta_{h'}, t_j] = \begin{cases} p[\theta_h, t_j] & \text{for } \theta_{h'} \in [0, \pi) \\ \\ p[\theta_h, n - t_j] & \text{for } \theta_{h'} \in [\pi, 2\pi) \end{cases} \in \mathbb{R}^{2m \times n}$$

$$\theta_{h'} = \frac{\pi h'}{m} \qquad h' = 0, 1, ..., 2m-1 \quad . \tag{7.6}$$

Fig. ??(b) shows the result of extending the example sinogram in Fig. ??(a).

7.1.6 Step (2): interleaved sinogram

The interleaved sinogram $p^{(2)}[\theta_{h''}, t_j]$ has $\theta_{h''} = h''\pi/2m$ for $h'' = 0, 1, ..., 4m-1$ and corresponds to:

$$p^{(2)}[\theta_{h''}, t_j] = \begin{cases} p^{(1)}[\theta_{h'}, t_j] & \text{for } h'' = 2h' \\ \\ 0 & \text{for } h'' = 2h'+1 \end{cases} \in \mathbb{R}^{4m \times n} \quad . \tag{7.7}$$

An example of interleaved sinogram is shown in Fig. ??(c).

7.1.7 Step (3): imposing HLCC

By introducing the following quantities:

$$c_k(\theta) = \sum_{l=-\infty}^{+\infty} b_{kl}\, e^{il\theta} \quad , \tag{7.8}$$

(??) can be rewritten as:

$$p(\theta, t) = \frac{1}{\pi} \sum_{k=0}^{+\infty} c_k(\theta)\, (1 - t^2)^{1/2}\, U_k(t) \tag{7.9}$$

The $c_k(\theta)$'s are the Chebyshev coefficients of $[p(\theta, t)(1 - t^2)^{-1/2}]$; the b_{kl}'s are the Fourier series coefficients of the c_{kl}'s. The strategy consists in computing the b_{kl} coefficients in two steps: first $p(\theta, t)$ is expanded onto the Chebyshev basis, then, the b_{kl}'s are retrieved from the $c_k(\theta)$'s.
The $c_k(\theta)$'s can be easily computed, by resampling the projections at points $t'_j = \cos(\pi(j + 1)/(n + 1))$ [277, 303]. The cosine-resampling simplifies $U_k(t)$ into a sine function. Thus, the first term of (7.9) becomes:

$$p^{(2)}[\theta_{h''}, t'_j] = \sum_{0}^{n-1} c_k[\theta_{h''}] \sin\left(\frac{\pi(j + 1)(k + 1)}{n + 1}\right) \quad . \tag{7.10}$$

(7.10) shows that the $c_k[\theta_{h''}]$'s correspond to the type-1 discrete sine transform (DST-1) coefficients of the cosine-resampled projections (an example is provided by Fig. 7.4(e)).
After running the DST-1 along the channel direction of the sinogram, the b_{kl}'s are finally yielded by the FFT along the view direction:

$$b_{kl} = \frac{1}{4m} \mathrm{DFT}\{c_k[\theta_{h''}]\} \quad . \tag{7.11}$$

Altogether, imposing the HLCC on $p^{(2)}[\theta_{h''}, t_j]$ requires the cosine resampling, a DST-1 along the channel direction, a FFT along the view direction, setting to zero the b_{kl}'s according to (7.5) (Fig. 7.2) and reversing the process, i.e. an IFFT along the view direction, an IDSF-1 along the channel direction and resampling at positions t_j. The resulting sinogram is $p^{(3)}[\theta_{h''}, t_j] \in \mathbb{R}^{4m \times n}$.

7.1.8 Step (4): crop in $[0, \pi)$ and reassign original data

The sinogram, $p^{(4)}[\theta_{h'''}, t_j]$, is finally cropped in the interval $[0, \pi)$, i.e. $p^{(4)}[\theta_{h'''}, t_j] \in \mathbb{R}^{2m \times n}$ and the original projections are reassigned. Therefore, we obtain:

$$p^{(4)}[\theta_{h'''}, t_j] = \begin{cases} p[\theta_h, t_j] & \text{for } h''' = 2h \\ p^{(3)}[\theta_h, t_j] & \text{for } h''' = 2h + 1 \end{cases} \in \mathbb{R}^{2m \times n} \quad , \quad h = 0, 1, \ldots, m - 1 \quad . \tag{7.12}$$

Experiments have shown that reassigning the original projections at the end of the procedure slightly increases the quality of the FBP reconstruction in terms of peak-signal-to-noise ratio, even though the consistency of the obtained sinogram marginally decreases.

7.1.9 Complexity and efficiency

The computational cost of the HLSF lies in the calls of the DST-1 along the channel direction and the FFT along the view direction. The computation of the DST can be factorized similarly to the FFT ($O(N \log_2 N)$ complexity) plus few pre- and post-processing steps with $O(N)$ complexity [304]. For an input sinogram with m views × n pixels, this yields to approximately $8mn(\log_2 n + \log_2 4m)$ floating operations and a resulting $O(mn(\log_2 n + \log_2 4m))$ complexity.

To show that the proposed filter only slightly impacts the total reconstruction speed, a pure Matlab implementation of the HLSF has been compared to the well known Matlab function *iradon*. Results for sinograms of different sizes, collected in Tab. 27, prove that the HLSF requires smaller runtimes than a standard non-GPU implementation of FBP, especially for real datasets where $m, n > 10^3$.

	HLSF	FBP
805 views × 512 pix.	0.22 s	0.30 s
1608 views × 1024 pix.	0.63 s	2.34 s
2500 views × 2048 pix.	1.65 s	14.50 s

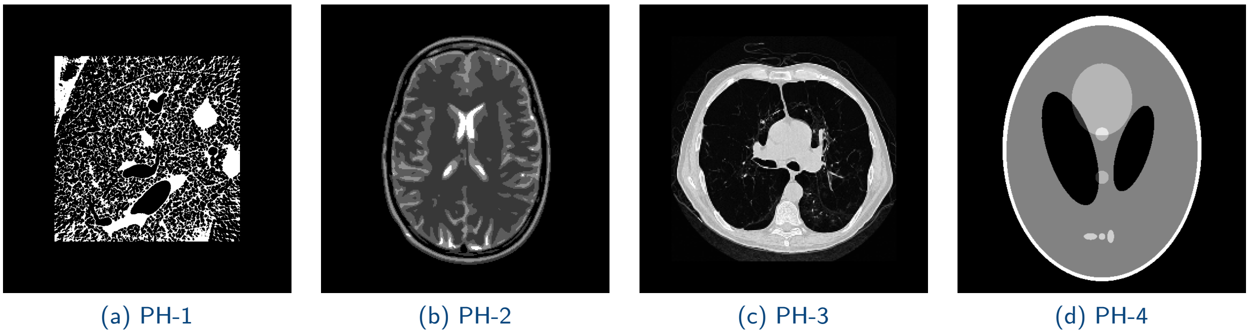
Table 27: Comparison between the time elapsed to run the HLSF and the FBP reconstruction for sinograms of different sizes. HLSF is implemented in pure Matlab. FBP is performed by the Matlab function *iradon*.

7.1.10 Benchmark procedure

To assess the performance of the HLSF, four phantoms with structural patterns of different complexity have been considered (Fig. 7.5). PH-1 in Fig. 7.5(a) is the segmentation of a reconstructed slice of mouse lung tissue at micrometer scale. PH-2 in Fig. 7.5(b) is a multilevel segmentation of a MRI scan of a human brain. PH-3 in Fig. 7.5(c) is a multilevel segmentation of a CT slice of a human lung. PH-4 is the well-known Shepp-Logan phantom [129].

The simulated sinograms are computed by a standard space-based implementation of the Radon transform based on slant-stacking with linear interpolation [93].

Reconstructions are performed by means of FBP. The tradeoff between signal-to-noise ratio (SNR) and spatial resolution of FBP reconstructions is highly dependent on the choice of the filter function. For this reason, the dataset has been reconstructed with 3 different filters: a pure ramp or Ram-Lak filter and a ramp combined to a Hanning or Parzen [106] window to damp the high frequency components of the projections [12]. These filters are indicated in the following as Ram-Lak, Hanning and Parzen, respectively. The Ram-Lak filter provides the highest spatial resolution and the worst



(a) PH-1 (b) PH-2 (c) PH-3 (d) PH-4

Figure 7.5: Set of simulated data used to benchmark the HLSF. PH-1 has 784 × 784 pixels; PH-2 has 592 × 592 pixels; PH-3 has 500 × 500 pixels; PH-4 has 512 × 512 pixels.

SNR, the opposite occurs for Parzen; Hanning is placed in the middle. Reconstructions labelled "CFBP" (Consistent FBP) were computed with FBP on a sinogram pre-filtered by HLSF with an upsampling factor of 2; reconstructions labelled "IFBP" (Interpolation FBP) were computed with FBP on a sinogram upsampled by a factor of 2 by means of 1D cubic spline interpolation along the view direction; reconstructions labelled "TFBP" (Telea FBP) were computed with FBP on a sinogram upsampled by a factor of 2 after having applied the Telea inpainting algorithm [289]; otherwise they are simply labelled "FBP".

The standard peak-signal-to-noise ratio (PSNR) [144], calculated within the reconstruction circle, is used to score each reconstruction with respect to the corresponding phantom.

The "sampling factor" (SF) is defined as the ratio between the number of projections of the considered sinogram and the number of projections of an optimally sampled sinogram. For parallel beam geometry, a sinogram is optimally sampled if $m \geqslant n\,\pi/2$, with m the number of views and n the number of detector pixels [12]. A sinogram with 100 views × 512 pixels, for example, considered well-sampled with $805 \backsimeq 512 \cdot \pi/2$ views, has SF $= 0.12$.

A metric, called *power of wrong coefficients (PWC)*, is here introduced to quantify the inconsistency of a given sinogram on the basis of its Fourier-Chebyshev coefficients. It is defined as follows:

$$\mathrm{PWC} = \frac{\sum\limits_{(k,l)=(k_w,l_w)} |b_{kl}|^2}{\sum\limits_{\forall k,l} |b_{kl}|^2} \cdot 100\% \quad , \tag{7.13}$$

where k_w, l_w are the indices of all $b_{kl} \neq 0$, that are supposed to be zero according to property (7.5).

7.1.11 Experiments

First, the performance of HLSF is tested for the noiseless case: noise-free sinograms with different SFs are reconstructed with FBP after pre-filtering with HLSF. The plots in Fig. 7.6 show that HLSF improves the reconstruction quality for small SF, while for higher SF the results for FBP and CFBP are comparable. The exact boundary between these 2 regimes depends on the filter choice. It is marked in Fig. 7.6 with a dashed black line and corresponds roughly to 0.47 for the Ram-Lak filter, 0.30 for Hanning and 0.15 for Parzen, regardless of the reconstructed object. The stronger the action of the filter, the smaller the SF interval where the HLSF increases the PSNR of the reconstruction.

Subsequently, HLSF is also tested for noisy sinograms. The standard deviation of the added Poisson noise is expressed as percentage of the original sinogram mean value and is indicated with σ. Fig. 7.7 presents two-dimensional maps showing the difference between the PSNR score of CFBP and FBP reconstructions. Positive values indicate that CFBP outperforms FBP and viceversa. Each map corresponds to a specific choice of reconstruction filter and phantom.

The differential maps, shown in Fig. 7.7, are characterized by the same trend observed for the PSNR in Fig. 7.6: the stronger the action of the filter, the smaller the area, where HLSF provides a substantial improvement. However, CFBP reconstructions have always a higher PSNR compared to the FBP ones, since values on these maps are everywhere positive (Fig. 7.7).

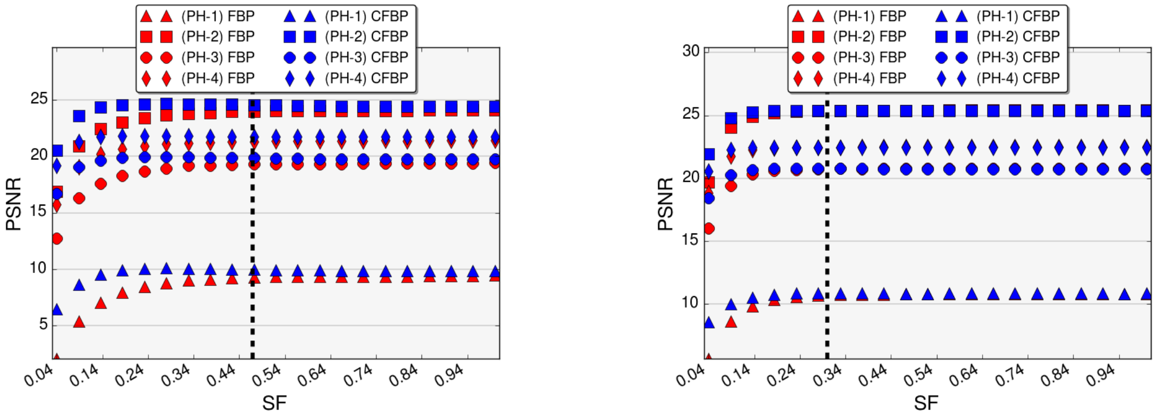
Reconstructions with FBP and CFBP for each phantom are displayed in Fig. 7.8, 7.9, 7.10 and 7.11. At visual inspection, the reconstructions with CFBP show better quality and details can be more easily identified.

CFBP has also been compared to 1D cubic spline interpolation along the view direction (upsampling factor of 2) followed by FBP (IFBP). Simple 2D interpolation schemes used to double the number of views of a sinogram can yield visible artifacts on the FBP reconstruction and are, therefore, not considered here for comparison with HLSF. The algorithms are tested, first, on noiseless, then, on noisy datasets. Figure 7.12(a) shows that CFBP outperforms IFBP for any value of SF, when using the Ram-Lak filter, whereas differences vanish, when using the Hanning filter and $\mathrm{SF} \gtrsim 0.30$, as illustrated by Fig. 7.12(b). Analogously to the results of Fig. 7.7, the stronger the action of the filter, the smaller the SF interval where CFBP can outperform IFBP for the reconstruction of noiseless undersampled datasets. The differential maps for PH-4 in Fig. 7.13 are everywhere positive (positive values correspond to $\mathrm{PSNR}_{\mathrm{CFBP}}(\sigma, \mathrm{SF}) > \mathrm{PSNR}_{\mathrm{IFBP}}(\sigma, \mathrm{SF})$) and show that CFBP yields a better reconstruction quality than IFBP, especially for $\sigma > 1.73$ and $\mathrm{SF} < 0.33$. Very similar results have also been obtained for PH-1, PH-2 and PH-3.

If an interleaved sinogram (as the one shown in Fig.7.3(a)) is simply regarded as an image, the extrapolation of missing projections can be performed, in principle, through inpainting. Among the methods tested for image inpainting, the Telea algorithm [289] was the only one filling the interleaved sinogram, providing FBP reconstructions, that were not affected by severe visible artifacts. CFBP was, therefore, compared to the Telea algorithm followed by FBP (TFBP). The differential maps for PH-2 displayed in Fig.7.14 are, once again, everywhere positive (positive values correspond to $\mathrm{PSNR}_{\mathrm{CFBP}}(\sigma, \mathrm{SF}) > \mathrm{PSNR}_{\mathrm{TFBP}}(\sigma, \mathrm{SF})$). Results showing CFBP outperforming TFBP have been also obtained for PH-1, PH-3 and PH-4.

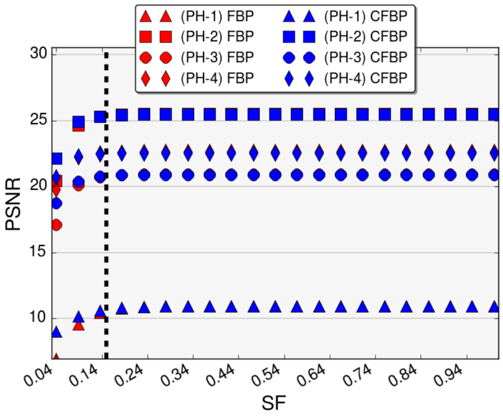
We finally present an experiment dealing with a synchrotron-radiation-based X-ray tomographic microscopy dataset of a modern seed acquired at the TOMCAT beamline of the Swiss Light Source at the Paul Scherrer Institute. The original sinogram consisting of 900 views × 2560 pixels has been downsampled to 150 views to be sufficiently undersampled. Figure 7.15 shows the FBP reconstruction of the original sinogram (Fig. 7.15(a)) and the FBP, IFBP, TFBP and CFBP reconstructions of the downsampled sinogram (Fig. 7.15(b-e)), respectively. The Hanning filter was used for all reconstructions. The undersampling artifacts appear quite reduced in the CFBP reconstruction compared to the results in Fig.7.15(b,c,d). The improvement in reconstruction quality is confirmed by the PSNR scores of the reconstructions (b-e) computed with respect to the FBP reconstruction of the original sinogram, used in this experiment as ground truth.

Other experiments (not shown here) have indicated that no additional improvement can be obtained from either applying HLSF multiple times sequentially or using HLSF to triple or quadruple at once the number of views.



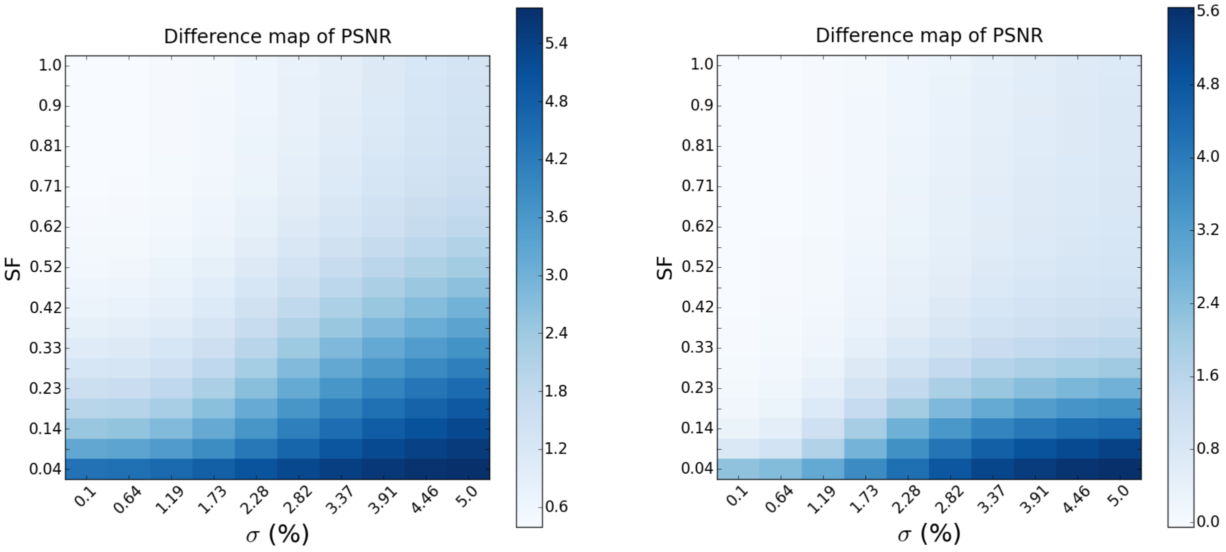
(a) Ram-Lak

(b) Hanning



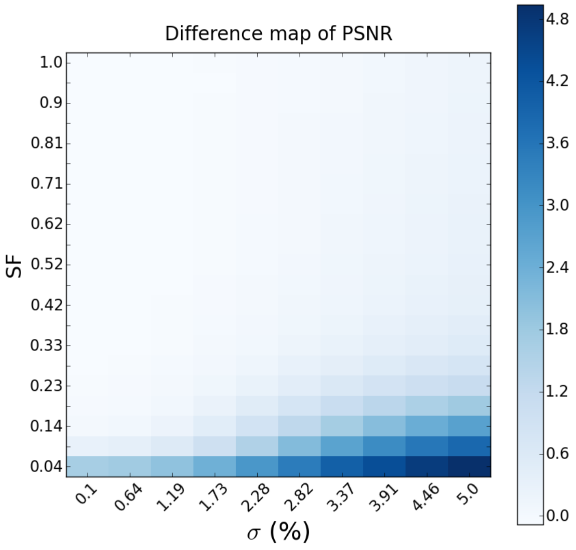
(c) Parzen

Figure 7.6: PSNR scores as a function of the sampling factor, SF, of the FBP reconstructions performed with Ram-Lak, Hanning and Parzen filters. The red markers correspond to FBP reconstructions, the blue ones to CFBP reconstructions. The marker shape is related to the phantom used to create the simulated data.



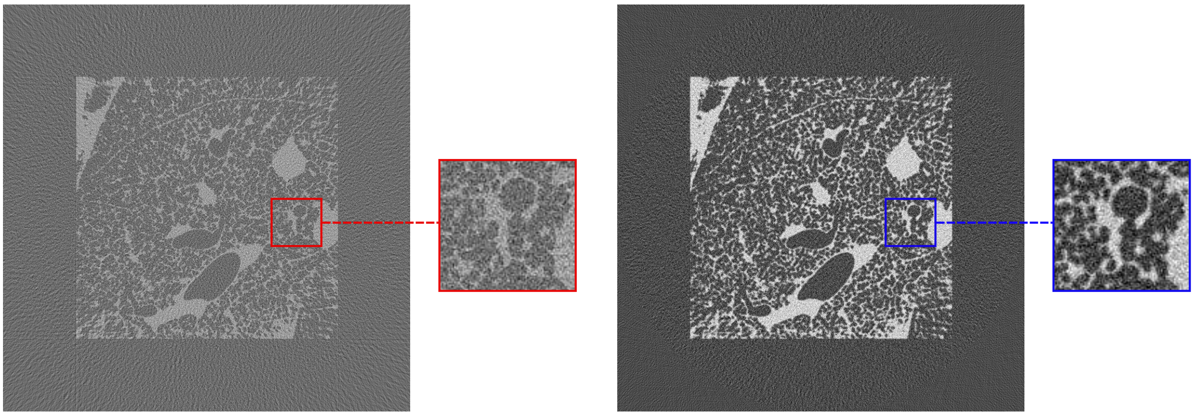
(a) Ram-Lak (PH-1)

(b) Hanning (PH-2)



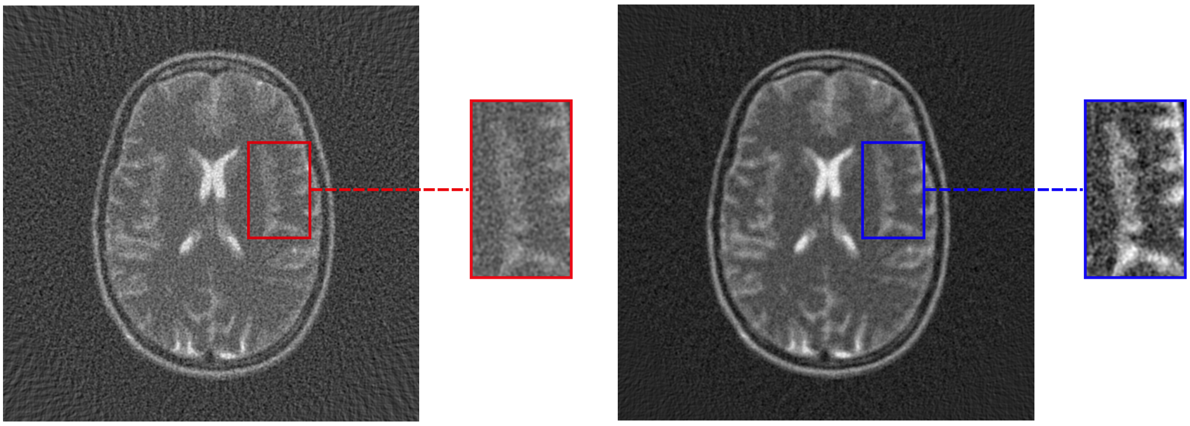
(c) Parzen (PH-3)

Figure 7.7: Maps of difference between PSNR scores of the CFBP and the FBP reconstructions as a function of the sampling factor, SF, and the variance, σ, of the additional Poisson noise. σ is expressed as percentage of the original sinogram mean value. The caption of each map specifies what filter and phantom were used for the reconstructions.



(a) FBP (b) CFBP

Figure 7.8: Reconstructions performed by FBP and CFBP with Ram-Lak filter of a PH-1 sinogram 168 views × 784 pixels (SF = 14%) + Poisson noise with $\sigma = 2.2\%$ of the original sinogram mean value.



(a) FBP (b) CFBP

Figure 7.9: Reconstructions performed by FBP and CFBP with Parzen filter of a PH-2 sinogram 82 views × 592 pixels (SF = 9%) + Poisson noise with $\sigma = 2.8\%$ of the original sinogram mean value.

7.2 Relationship between reconstruction quality and sinogram consistency

The experiment in Fig.7.16, involving a simulated sinogram of Shepp-Logan, shows how the PWC varies at the different stages of the HLSF procedure and highlights a fact, that was only briefly mentioned at the end of Section 7.1.4: the PWC does not always and necessarily correlate with the reconstruction quality expressed in terms of PSNR.

The original undersampled sinogram with 110 views × 512 pixels (Fig.7.16(a)) is consistent by con-

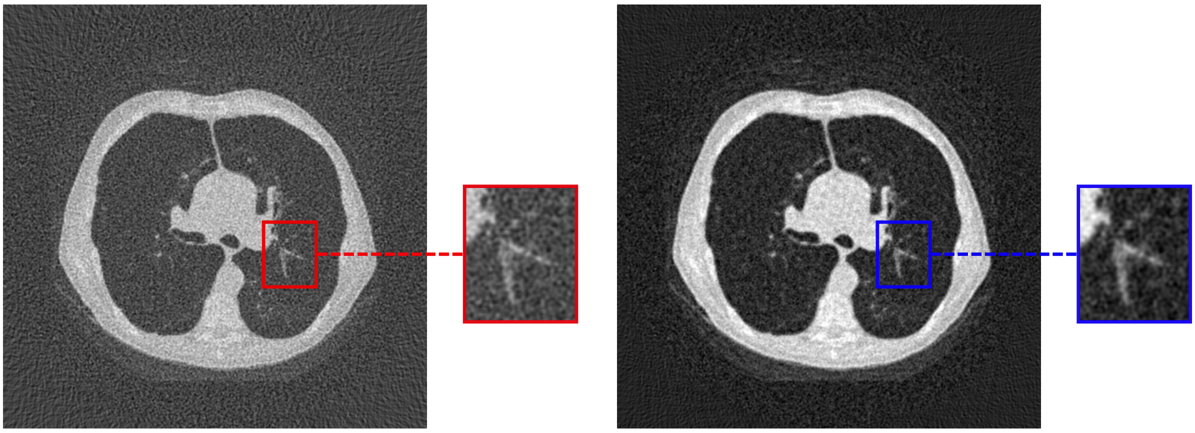

(a) FBP (b) CFBP

Figure 7.10: Reconstructions performed by FBP and CFBP with Hanning filter of a PH-3 sinogram 108 views × 500 pixels (SF $= 14\%$) + Poisson noise with $\sigma = 2.2\%$ of the original sinogram mean value.

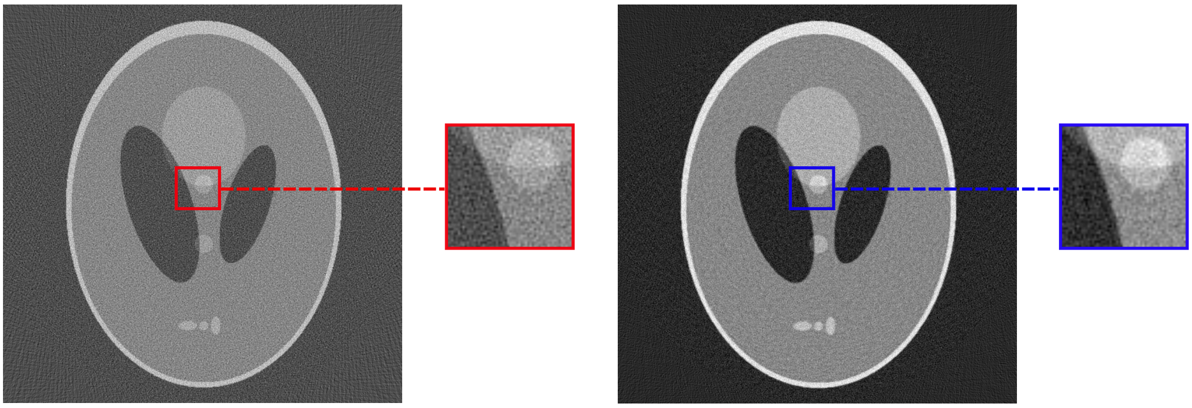
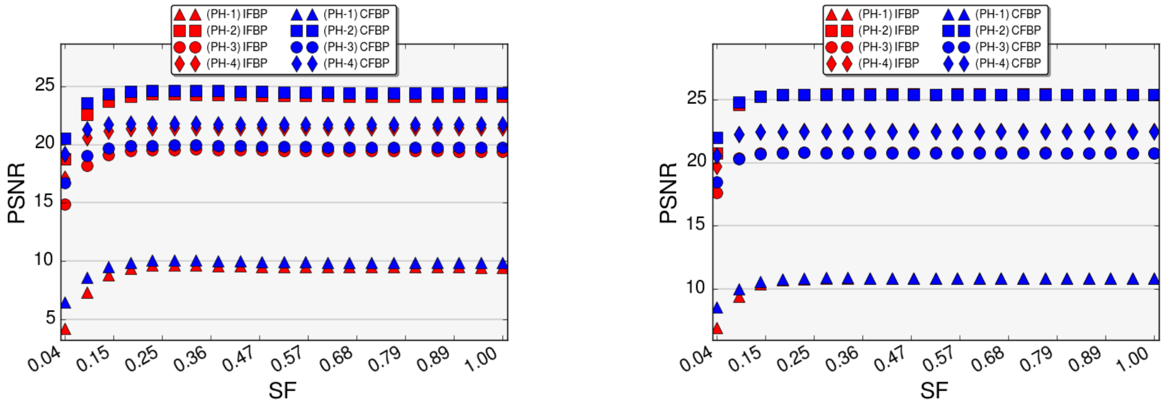

(a) FBP (b) CFBP

Figure 7.11: Reconstructions performed by FBP and CFBP with Ram-Lak filter of a PH-4 sinogram 149 views × 512 pixels (SF $= 18\%$) + Poisson noise with $\sigma = 1.1\%$ of the original sinogram mean value.

struction, with PWC $= 0.50\%$ (PWC $< 1\%$ generally corresponds to sinograms with high consistency). When the sinogram is interleaved with zero lines (Fig.7.16(b)), its consistency becomes very low, PWC $= 70.71\%$, and is restored once the HLCC are imposed (Fig.7.16(c)), reaching PWC $= 0.56\%$. The original sinogram has slightly lower PWC than the HLSF sinogram, but the FBP reconstruction of the HLSF sinogram (Fig.7.16(f)) has substantially higher PSNR than the FBP reconstruction of the original sinogram (Fig.7.16(e)) as a consequence of using double number of projections.
The same conclusion was obtained when comparing the HLSF algorithm with and without final re-assignment of the original projections, this last version indicated with HLSFn. The HLSFn sinogram



(a) Ram-Lak

(b) Hanning

Figure 7.12: PSNR scores as a function of the sampling factor, SF, of the FBP reconstructions performed with Ram-Lak and Hanning filters. The red markers correspond to IFBP (cubic spline interpolation along the view direction + FBP) reconstructions, the blue ones to CFBP reconstructions. The marker shape is related to the phantom used to create the simulated data.

shown in Fig.7.16(d) has slighly smaller PWC than the HLSF sinogram in Fig.7.16(c), but its FBP reconstruction in Fig.7.16(g) has inferior PSNR than the one in Fig.7.16(f). This empirical result explains why the HLSF algorithm includes the reassignment of the original projections at the end of the procedure.

The same mismatch is also observed when other inpainting algorithms are used. The sinogram in Fig.7.16(a) doubled in number of views by means of 1D cubic interpolation has PWC = 2.5% and its FBP reconstruction has PSNR = 19.61. The same sinogram doubled in number of views by the Telea inpainting algorithm has PWC = 4.7% and its FBP reconstruction has PSNR = 21.54.

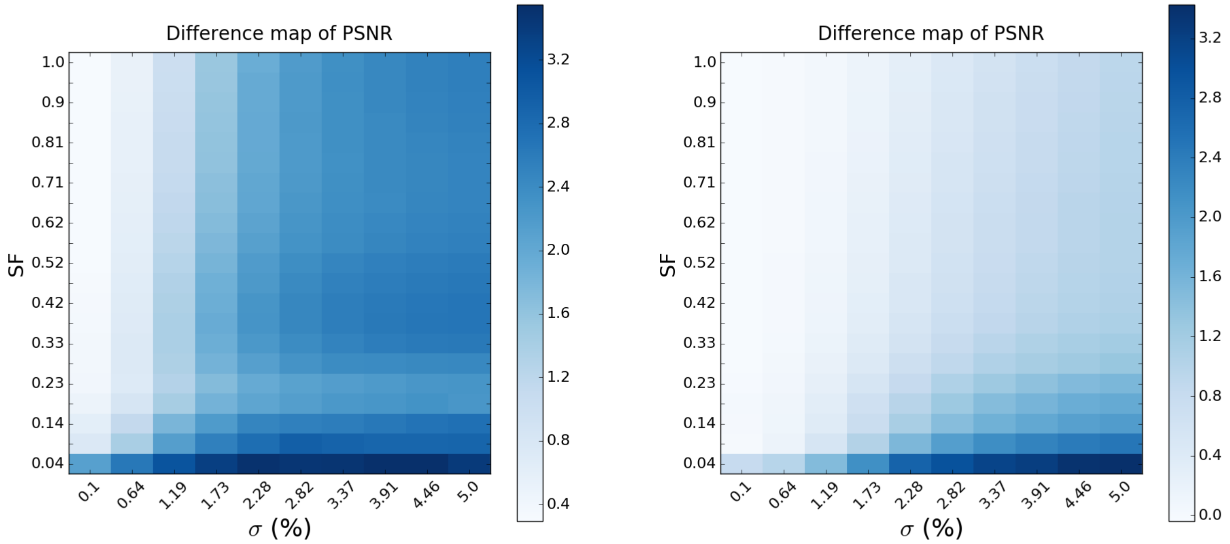
These example show that a more consistent sinogram does not necessarily lead to a FBP reconstruction of higher quality, at least when comparing sinograms with PWC values $\lesssim 10\%$.

7.3 Discussion and conclusion

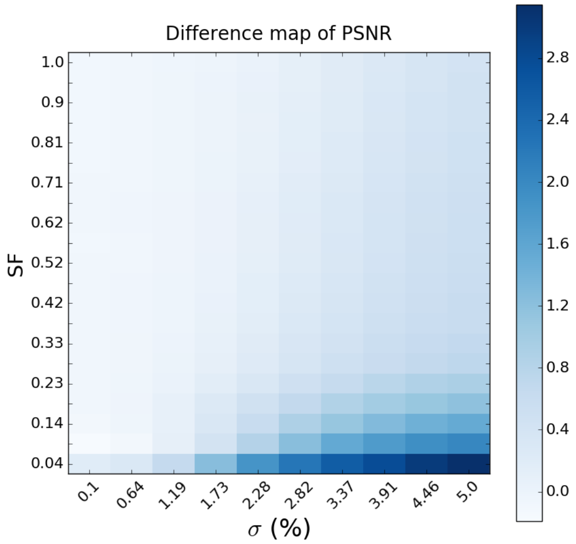
This work presents a fast procedure to improve analytical tomographic reconstructions of undersampled datasets in parallel beam geometry. The proposed method is a filter working in the Radon domain and based on the Helgason-Ludwig consistency conditions. It doubles the number of views of a sinogram homogenously sampled in $[0, \pi)$, by extrapolating projections at intermediate angular positions.

This sinogram filter, abbreviated HLSF, is a non-iterative, parameterless procedure, that can be efficiently implemented with FFTs and only marginally impacts the total computational cost for analytical reconstructions.

Experiments, performed on simulated data of different structural complexity, have shown that FBP reconstructions of sinograms pre-processed with the presented HLSF are charaterized by a higher PSNR compared to FBP reconstructions of standard sinograms. HLSF improves the reconstruction



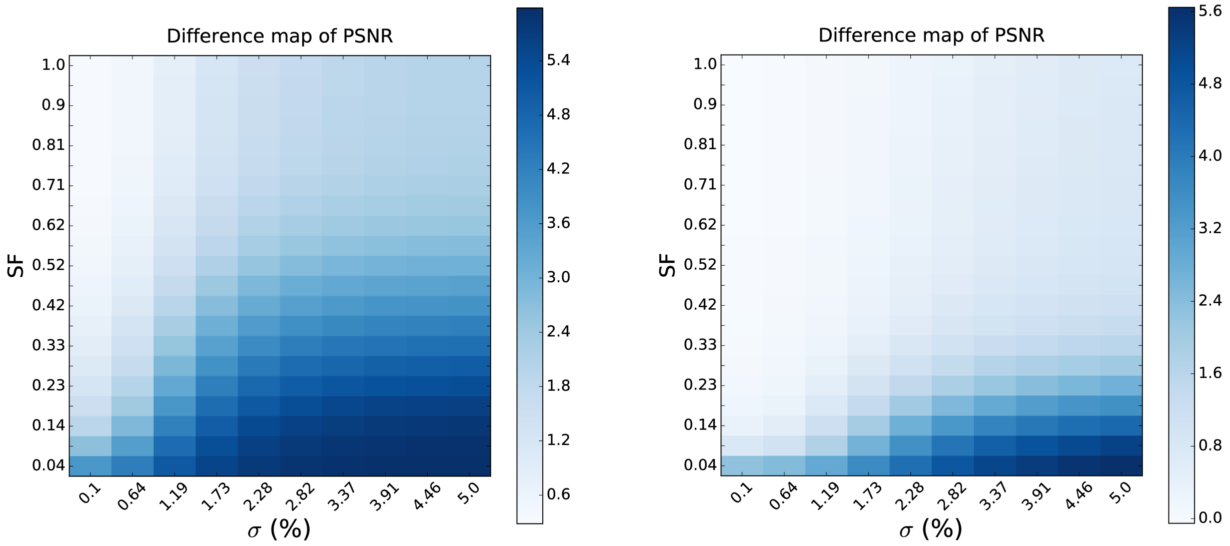
(a) Ram-Lak (PH-4) (b) Hanning (PH-4)



(c) Parzen (PH-4)

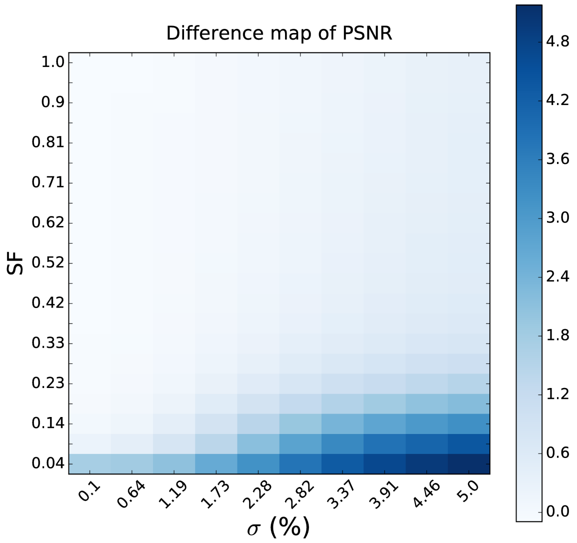
Figure 7.13: Maps of difference between PSNR scores of the CFBP and the IFBP (cubic spline interpolation along the view direction + FBP) reconstructions as a function of the sampling factor, SF, and the variance, σ, of the additional Poisson noise. σ is expressed as percentage of the original sinogram mean value. The caption of each map specifies what filter and phantom were used for the reconstructions.

quality for both noiseless (generally, for SF $\lesssim$ 0.31) and noisy undersampled (especially, for SF $\lesssim$ 0.33 and $\sigma \gtrsim 1.73\%$) sinograms. Moreover, HLSF outperforms 1D cubic spline intepolation along the view direction and the Telea algorithm for image inpainting: improvements in the reconstruc-



(a) Ram-Lak (PH-2)

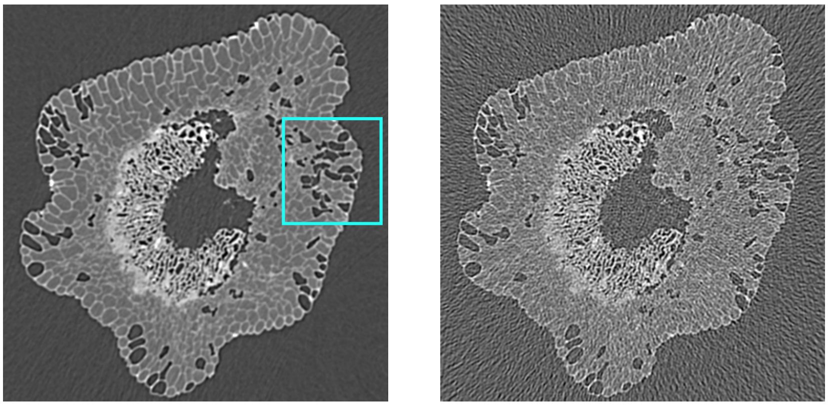
(b) Hanning (PH-2)



(c) Parzen (PH-2)

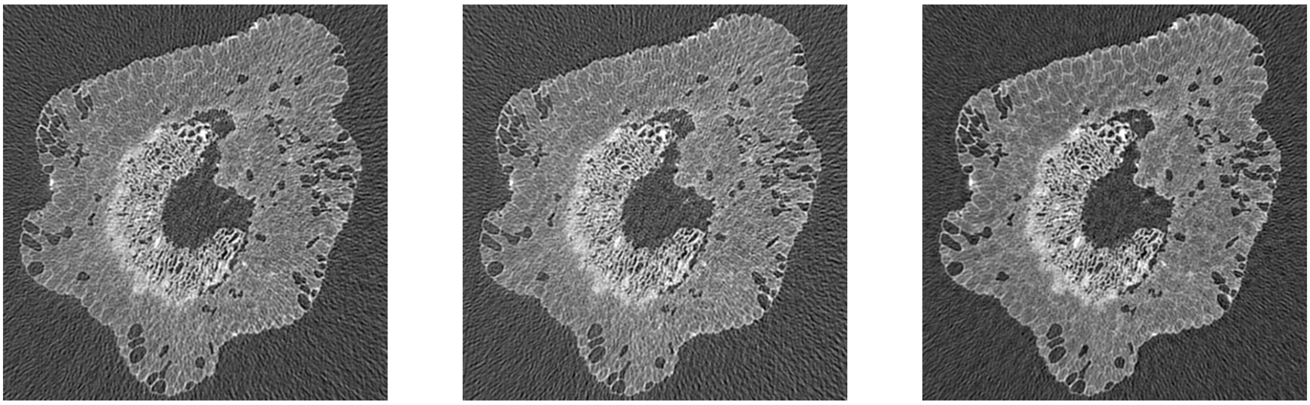
Figure 7.14: Maps of difference between PSNR scores of the CFBP and the TFBP (Telea inpainting algorithm + FBP) reconstructions as a function of the sampling factor, SF, and the variance, σ, of the additional Poisson noise. σ is expressed as percentage of the original sinogram mean value. The caption of each map specifies what filter and phantom were used for the reconstructions.

tion accuracy are substantial when dealing with noisy undersampled datasets (especially, for SF $\lesssim$ 0.33 and $\sigma \gtrsim 1.73\%$). This result has been confirmed by experiments performed on real datasets of synchrotron-radiation-based X-ray tomographic microscopy.



(a) FBP (900 proj.)

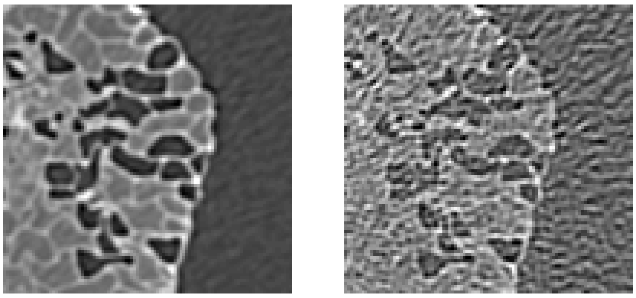
(b) FBP (150 proj.) | PSNR = 24.3



(c) IFBP (150 proj.) | PSNR = 26.3

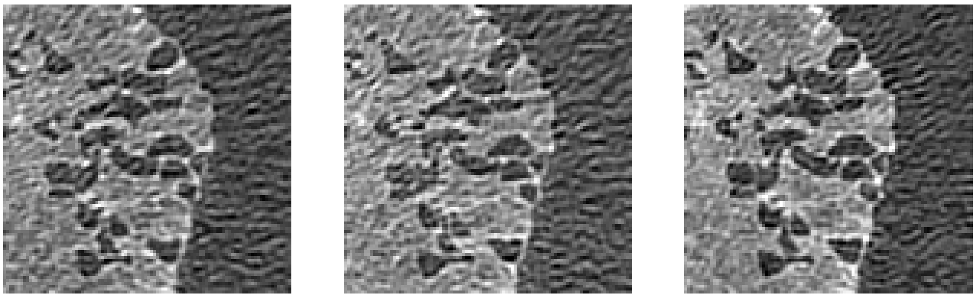
(d) TFBP (150 proj.) | PSNR = 26.8

(e) CFBP (150 proj.) | PSNR = 28.6



(f) Detail - FBP (900 proj.)

(g) Detail - FBP (150 proj.)

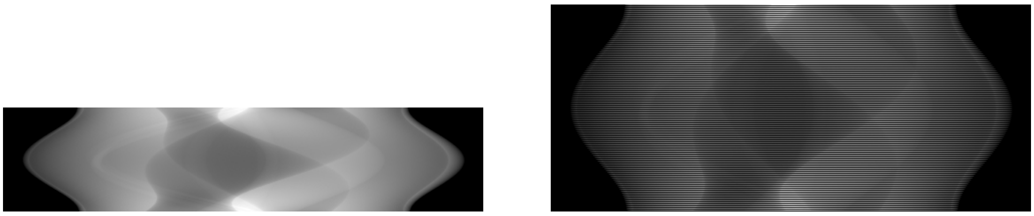


(h) Detail - IFBP (150 proj.)

(i) Detail - TFBP (150 proj.)

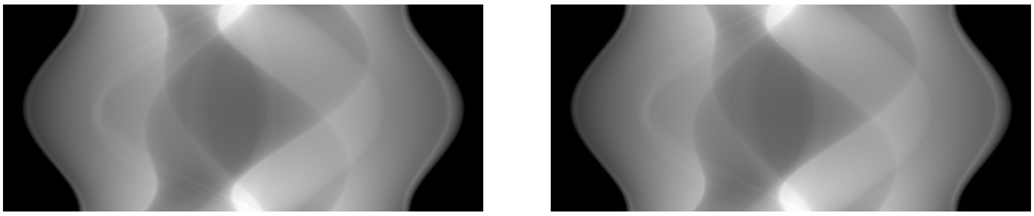
(j) Detail - CFBP (150 proj.)

Figure 7.15: FBP reconstructions with Hanning filter of a dataset of a modern seed specimen. The PSNR values of reconstructions (b-e) are computed with respect to (a), that is used as ground truth. (a) FBP reconstruction of the original sinogram with 900 views × 2560 pixels; the light blue square identifies the regions zoomed in images (f-l). (b) FBP reconstruction of the downsampled sinogram with 150 views × 900 views. (c) IFBP reconstruction (cubic spline interpolation along the view direction + FBP) of the downsampled sinogram with 150 views × 900 views. (d) TFBP reconstruction (Telea inpainting algorithm + FBP). (e) CFBP reconstruction of the downsampled sinogram with 150 views × 900 views. (f-l) Blow-ups of the region indicated in (a) of all the reconstructions. Sample courtesy: S. Smith, University of Michigan.



(a) Undersampled sinogram, PWC=0.50%

(b) Interleaved sinogram, PWC=70.71%



(c) HLSF sinogram, PWC=0.56%

(d) HLSFn sinogram, PWC=0.49%

(e) FBP of sinogram (a), PSNR=20.40

(f) FBP of HLSF sinogram, PSNR=22.56

(g) FBP of HLSFn sinogram, PSNR=22.44

Figure 7.16: Experiment showing the relationship between PWC and PSNR. (a) Original undersampled sinogram with 110 views × 512 pixels. (b) Interleaved sinogram. (c) Result of applying the HLSF to sinogram (a). (d) Result of applying the HLSFn to sinogram (a), where HLSFn corresponds to HLSF without final reassignment of the original projections. (e) FBP reconstruction with Hanning filter of the original undersampled sinogram. (f) FBP reconstruction with Hanning filter of the HLSF sinogram. (g) FBP reconstruction with Hanning filter of the HLSFn sinogram.

Chapter 8

Summary & Outlook

8.1 Overview of the results

The challenge embraced by this doctoral project was designing reconstruction algorithms for underconstrained datasets of SRXTM low-dose scans, acquired in absorption-contrast, in-line phase-contrast and differential phase-contrast, in full-FOV or full interior tomography. A good trade-off between accuracy and computational efficiency, versatility in tackling datasets acquired with different imaging modalities and no need for supervised a-priori knowledge were pivotal requirements for the investigated reconstruction methods.

In view of the stated scientific aim, the following can be regarded as the main results achieved by this work.

- Design and optimization of the gridding projector with minimal oversampling (FGP-KB). FGP-KB opens the way for fast iterative tomographic reconstruction on CPU clusters. When used inside an iterative scheme, FGP-KB has comparable accuracy as more sophisticated forward projectors working in real domain and its speed is similar to that of (real-domain) tomographic operators implemented on GPUs. This projector can be used for different types of datasets, namely absorption-contrast, inline phase-contrast differential phase contrast and full interior tomography with minor changes in the gridding kernel.

- Identification of the alternate direction method of multipliers plug-and-play (ADMP) as iterative reconstruction method fulfilling the requirements defined in the scientific aim of this project. The ADMP features fast convergence, an excellent CNR-resolution trade-off and the versatility of switching forward projector without changing the structure of the algorithm. In addition, the ADMP offers a framework where sophisticated image denoising methods can be "plugged in" and "played" as regularization for the iterative reconstruction. Advanced[1] and computationally-feasible regularization schemes, preliminarily tested within the ADMP, are the split-Bregman

---

[1] *"Advanced"* is meant with respect to the well-established regularization schemes listed 2.5.2 (e.g. Tikhonov, Huber).

total variation, the non-local split-Bregman total variation and the total generalized variation. All these denoising methods do not require supervised a-priori knowledge.

- Experimental characterization of the role played by the tomographic projectors on the performance of an iterative reconstruction algorithm. In this regard, two results must be highlighted. The first result is that the accuracy of the projectors alone has limited impact on the reconstruction accuracy of the iterative method: in other words, projectors with different "standalone" accuracies (i.e. creating a single sinogram), used by the same iterative algorithm, yield almost identical reconstructions at visual inspection and in terms of standard metrics for image quality. The second result is that a high degree of coupling between forward and backprojector is of vital importance to ensure the convergence and accuracy of an iterative algorithm. The overall take-home message is that computational efficiency and coupling degree are the only aspects to consider for tomographic projectors when setting up an iterative reconstruction algorithm.

- Introduction of the virtual strategy as an efficient way to perform iterative reconstruction of underconstrained datasets in full interior tomography (FINT) without knowledge of the object support. The virtual strategy is effective in removing the typical FINT artifacts and yields a reconstruction quality comparable to what achievable with the edge-padding and differentiated strategies. The virtual strategy has the upside of being faster by a factor of 20 and 50-80 compared to the edge-padding and differentiated strategies, respectively.

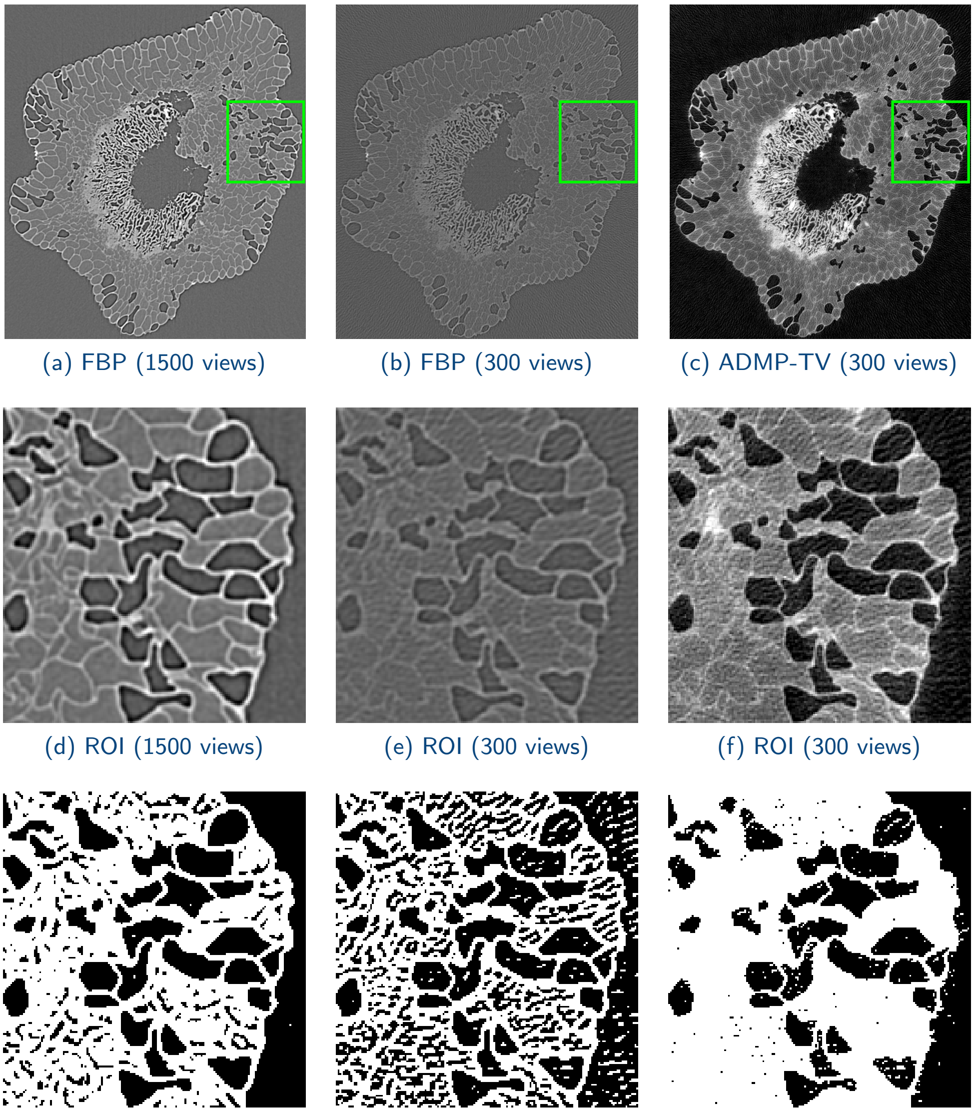
- Design of the Helgason-Ludwig sinogram filter (HLSF) to improve the accuracy of analytical reconstructions of strongly underconstrained absorption-contrast and inline phase-contrast datasets in full FOV. The HLSF has the upside of being parameterless and relatively fast, as it adds a negligible amount of computations with respect to filtered backprojection (FBP). On the other hand, the HLSF cannot compete with the accuracy achievable by finely tuned iterative reconstruction algorithms, thus offering a reconstruction quality halfway between what provided by FBP and an iterative algorithm.

8.2 Implications

Now that we have identified the most suited iterative algorithm (ADMP) and its components (gridding projectors, one of the aforementioned TV regularization schemes, etc ...) to address the reconstruction of underconstrained SRXTM datasets, the still unanswered question is: what advantage can iterative reconstruction offer in terms of dose reduction?

To address this question it is necessary to bear in mind that the achievable dose reduction depends on the characteristics of the selected dataset (structural complexity, intrinsic contrast, etc ...) and on the quantification goal.

Consider the example in Fig.8.1, regarding the tomographic reconstruction of a modern seed sample in absorption-contrast and full-FOV. The original sinogram consists of 1500 views × 2560 pixels. Assume that the final goal is segmenting the inner holes, displayed, e.g., in Fig.8.1(d). When the number of projections of the original sinogram is downsampled by a factor 5 (300 views), the Otsu



(a) FBP (1500 views) (b) FBP (300 views) (c) ADMP-TV (300 views)

(d) ROI (1500 views) (e) ROI (300 views) (f) ROI (300 views)

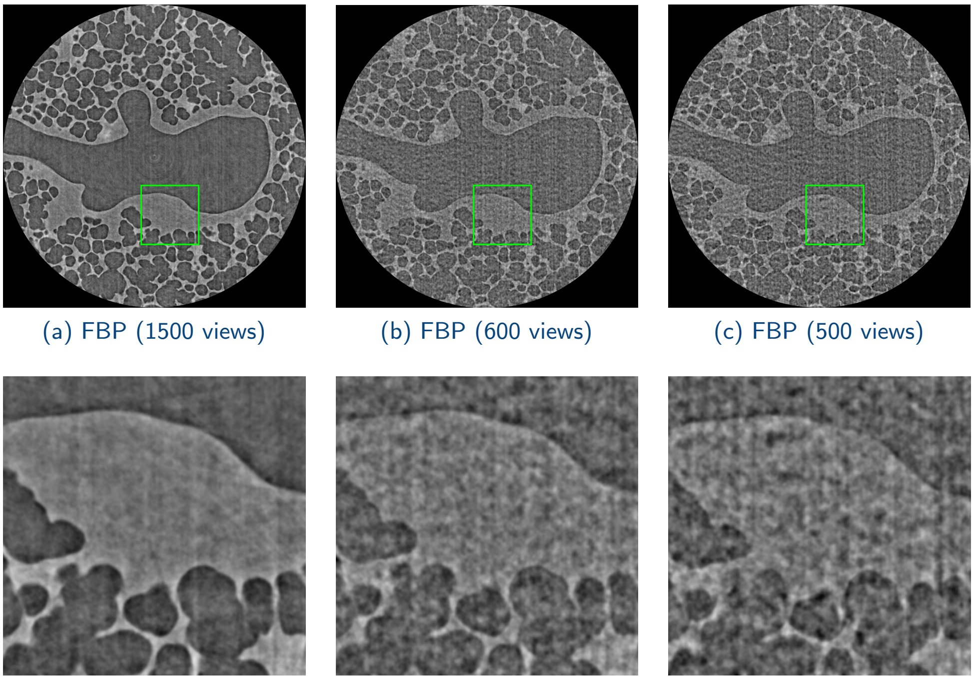
(g) OTSU (1500 views) (h) OTSU (300 views) (i) OTSU (300 views)

Figure 8.1: Reconstructions and segmentations of a SRXTM dataset of a modern seed in absorption-contrast and full-FOV. The top row shows the entire reconstructions. The middle row shows the zoom-ins of the selected ROI. The bottom row shows the Otsu segmentations of the ROI [305]. The first column contains results related to the analytical reconstruction of the full sinogram with 1500 views × 2560 pixels; the second and third columns contain results related to the analytical and iterative reconstruction (ADMP-TV) of the sinogram with 300 views × 2560 pixels. Sample courtesy: S. Smith, University of Michigan.

segmentation [3] [305] (Fig.8.1(h)) of the FBP reconstruction (Fig.8.1(b,e)) has very poor quality, as

---

[3] The Otsu method has been selected because it does not require any input parameter.



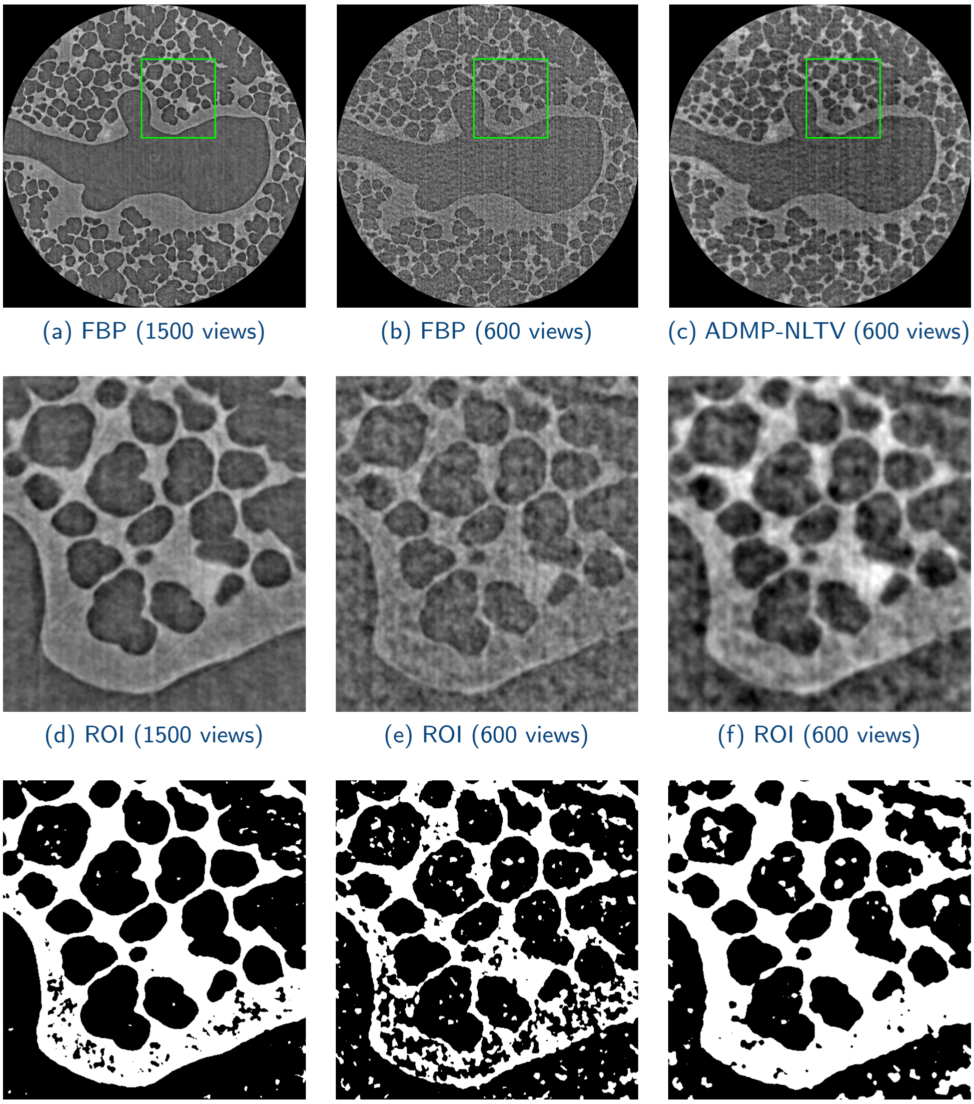
(a) FBP (1500 views) (b) FBP (600 views) (c) FBP (500 views)

(d) ROI (1500 views) (e) ROI (600 views) (f) ROI (500 views)

Figure 8.2: FBP reconstructions of a SRXTM dataset of mouse lung tissue in inline phase-contrast and full interior tomography at 1.1µm resolution. Speckle artifacts affect the analytical reconstructions of the downsampled sinograms with 600 and 500 views. In the FBP reconstruction of the sinogram with 500 views, the airways start to merge and fake tiny structures begin to appear in the pulmonary tissue. This brings to the conclusion that it is not possible to push the number of views below 600 and have reliable tomographic reconstructions at the same time.

it features several clusters of misclassified pixels. The iterative reconstruction with the ADMP-TV of the same undersampled dataset (Fig.8.1(c,f)) offers, instead, a better Otsu segmentation (Fig.8.1(i)), where the misclassified pixels are more likely to be removable with a proper use of morphological operators. In this fictitious example, the conclusion is that using the ADMP-TV can lead to a dose reduction by a factor of 5 for the considered dataset.

The same approach is used to estimate the achievable dose reduction for the acquisition of SRXTM datasets of ex-vivo mouse lung tissue in inline phase-contrast and full interior tomography at a resolution of 1.1 µm [1]. The starting point is always to analyze how the quality of the FBP reconstruction deteriorates for a decreasing number of views. The reconstructed slice consists of two phases: the pulmonary tissue (light gray) and the airways (dark gray). Figure 8.2 shows that decreasing the original 1500 views to 600 and 500 leads to speckle artifacts that severely affect the quality of the FBP reconstruction. In particular, when the number of views is 500 (Fig.8.2(c,f)), the airways start to merge and some fake tiny structures seem to appear in the middle of the pulmonary tissue. With this degree of undersampling (500 views), iterative reconstruction without ad-hoc a-priori knowledge cannot offer any help in improving the accuracy of the results. A good rule of thumb to understand, indeed, when


(a) FBP (1500 views) (b) FBP (600 views) (c) ADMP-NLTV (600 views)

(d) ROI (1500 views) (e) ROI (600 views) (f) ROI (600 views)

(g) OTSU (1500 views) (h) OTSU (600 views) (i) OTSU (600 views)

Figure 8.3: Reconstructions and segmentations of the SRXTM dataset already displayed in Fig.8.2. The top row shows the entire reconstructions. The middle row shows the zoom-ins of the selected ROI. The bottom row shows the Otsu segmentations of the ROI [305]. The first column contains results related to the analytical reconstruction of the full sinogram with 1500 views × 2016 pixels; the second and third columns contain results related to the analytical and iterative reconstruction (ADMP-NLTV) of the sinogram with 600 views × 2016 pixels.

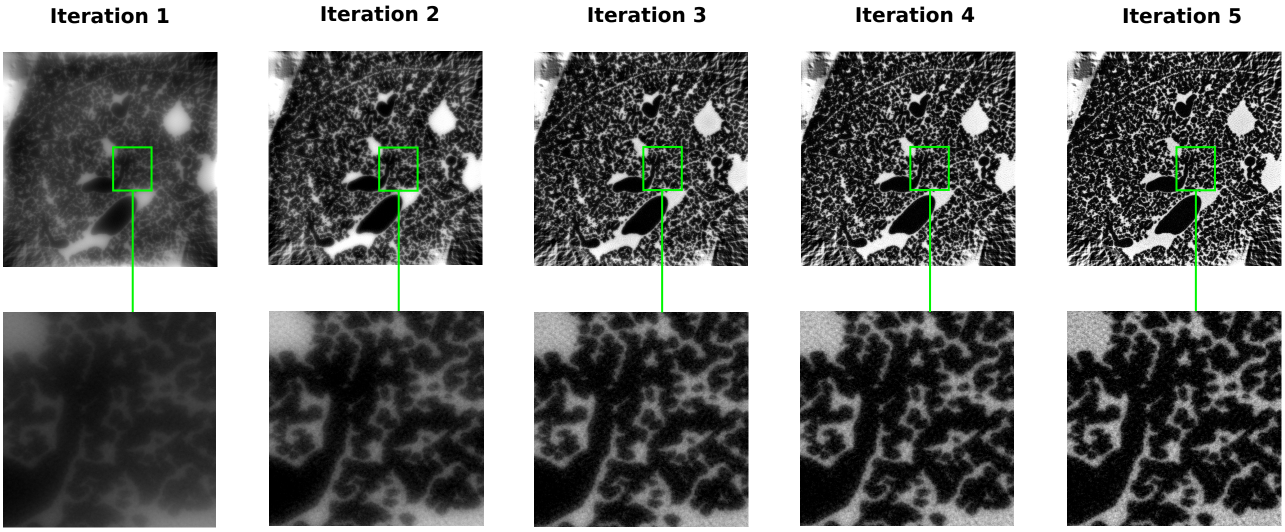
iterative reconstruction can represent an asset or not, is the following: structures that cannot be well identified at visual inspection in the FBP reconstruction of an underconstrained dataset, because of the noise and artifacts, will not be satisfactorily retrieved by an iterative reconstruction algorithm of the same dataset. In the case of the lung dataset, this means that is not possible to work with a number of views below 600 and still hope to be able to properly segment all the airways. Figure 8.3 shows reconstructions and segmentations of the original and the undersampled dataset with 600 views. From the results in Fig.8.3(g,h,i), it appears that iterative reconstruction (performed in this case by the ADMP-NLTV) can help in boosting the contrast-to-noise ratio, thus facilitating the subsequent segmentation, although the result in Fig.8.3(i) still requires additional processing to correctly classify all pixels [306]. In conclusion, for the considered type of SRXTM dataset, it is possible to decrease the irradiated dose by a factor 2.5.

8.3 Perspectives

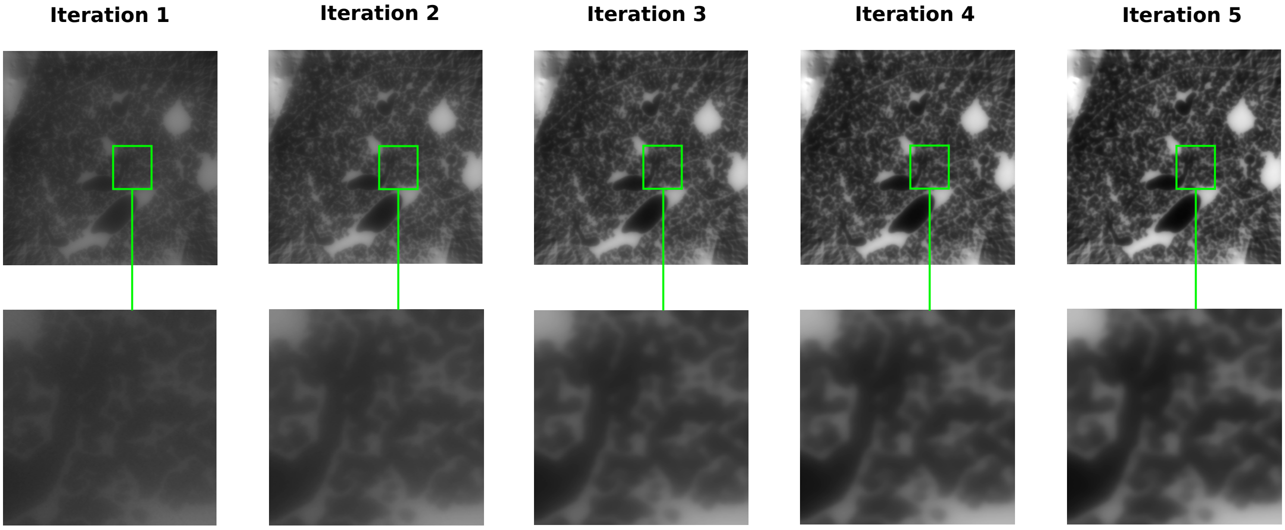
Key ingredients missing in the proposed ADMP are an automatic stopping criterion and a criterion to optimally select the input parameters λ and μ. The ADMP with gridding projectors stands out in terms of computational efficiency and versatility, but it still lacks a solid "user-friendliness". From a practical point of views, in order to know how to stop the ADMP at the right moment, it is necessary to perform various preliminary tests with different number of iterations or values of ϵ for criteria of the form "is $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2 < \epsilon$?", $\mathbf{x}^{(k)}$ and $\mathbf{x}^{(k+1)}$ being consecutive intermediate results of the iterative process. In the same way, the λ-μ space has to be explored in advance to know which setting provides optimal reconstruction quality. This is definitely not appealing from a user perspective.

The solution to these problems could be found by analyzing the working mechanism of the ADMP. Consider Fig.8.4(a), showing the first 5 iterations of the ADMP-TV for the reconstruction of a simulated underconstrained dataset. After the first iteration, the reconstruction looks very blurred, but it becomes increasingly sharper as other iterations of the ADMP-TV follow. Assume now that the trade-off between CNR and spatial resolution can be considered acceptable already at the fourth iteration. Then, signaling the iterative procedure that the improvement in CNR or spatial resolution at iteration 5 is negligible compared to iteration 4 would be an effective way to stop the ADMP. Notice that this is very different from stopping the iterative process when $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2 < \epsilon$. Since the CNR is a supervised metric requiring manually selected ROIs, the spatial resolution can represent a better candidate to serve as stopping criterion. The Fourier ring correlation computed between $\mathbf{x}^{(k)}$ and $\mathbf{x}^{(k+1)}$ could provide the information about the change in spatial resolution necessary to stop the ADMP.

Regarding the optimal selection of λ and μ, consider Fig.8.4(a) and Fig.8.4(b), showing reconstructions of the same simulated underconstrained dataset performed by the ADMP-TV with different parameter settings. In both cases, the reconstruction starts as very blurred and gets sharper after each iteration; the difference is that this "sharpening" process clearly works at different speeds. At the end, the final reconstruction obtained with $\lambda = \mu = 1.0$ will be very sharp, but also noisy, whereas the one obtained with $\lambda = \mu = 10^5$ will be very blurred, but with better SNR. What is, therefore, the best parameter setting in this case? It depends on what visualization/quantification goal has been planned for the reconstructed dataset. The point is that an iterative algorithm like the ADMP should not have an automatic criterion for the selection of the parameters. From a practical perspective, the input parameters (λ, μ) are just a "leverage" to set the target trade-off between SNR and spatial resolution. A good practical solution would be to have a graphical user interface (GUI) connected to



(a) First 5 iterations of ADMP-TV with $\lambda = \mu = 1.0$



(b) First 5 iterations of ADMP-TV with $\lambda = \mu = 10^5$

Figure 8.4: First 5 iterations of the reconstructions performed by ADMP-TV with different parameter settings of a simulated underconstrained dataset. (a) First 5 iterations of ADMP-TV with $\lambda = \mu = 1.0$. (b) First 5 iterations of ADMP-TV with $\lambda = \mu = 10^{-5}$.

the ADMP and reconstructing a single slice with, e.g., 5-6 choices of (λ, μ). All these reconstructions, characterized by different SNR-resolution trade-offs, are, then, displayed side by side, so that the user can select the best result according to his own judgement. Once the optimal set (λ, μ) is chosen by the user, the ADMP reconstructs the entire volume with the same setting.
Considering this hypothetical GUI support for the ADMP and that processing a volume with, e.g., 500 views × 2048 detector pixels × 2048 slices requires between 25-35 minutes on a cluster with 50

cores, the ADMP has really the potential to become a useful tool to address the reconstruction of low-dose tomographic datasets acquired at synchrotron imaging beamlines.

This thesis, as most of the studies in the field of iterative tomographic reconstruction, presents experiments on real datasets with variable sampling factor. To perform low-dose tomographic scans, decreasing the exposure time per projection, i.e. acquiring noisier datasets, is a strategy that can be used in place of reducing the number of views. Since compressed-sensing-based denoising methods are designed to suppress random noise (especially Poissonian and Gaussian), the ADMP working with the discussed regularization schemes is expected to work better with datasets acquired with short exposure time per projection rather than with a small amount of views. As a matter of fact, undersampling leads to tomographic reconstructions affected by aliasing artifacts, that look differently depending on the kind of dataset and object. For example, aliasing artifacts appear in the form of wavy streaks in the reconstruction of the modern seed sample in Fig.8.1(e) and in the form of speckles in the reconstruction of the mouse lung tissue in Fig.8.2(c). Regularization schemes based on compressed sensing do not perform well with structured artifacts and may also run the risk to exchange those artifacts for structures to be preserved. For this reason, substantial reduction of the dose irradiated in SRXTM scans might be achieved by decreasing the exposure time per projection and using iterative reconstruction methods as those proposed by this work.

Looking beyond this thesis, promising future developments for tomographic reconstruction can be envisaged in the incorporation of machine learning techniques. The pioneering and elegant work of Pelt [307] has proved that FBP boosted by a simple multilayer perceptron (MLP) with a relatively small training dataset can offer an appealing alternative to iterative reconstruction. This algorithm, called neural network FBP (NN-FBP), is however limited to learn a set of optimized filters and the final reconstruction can be roughly considered a linear combination of FBP reconstructions.
Some research groups are currently exploring the possibilities offered by deep learning for image reconstruction. "Deep learning" is an expression used to define MLPs with several hidden layers. In particular, for image processing, convolutional and recurrent neural networks (CNN and RNN) [308, 309] have proved to be the most powerful tools so far to perform segmentation [310], pattern recognition [311] and natural language processing [312]. An interesting aspect is that very little of a CNN/RNN for segmentation or pattern recognition has to be changed to transform it into a CNN/RNN for image reconstruction [313], namely only the last layer of the network. The challenge with CNNs and RNNs for many applications is to design a valid training dataset. Despite the lack of mathematical criteria to assess the quality of a training dataset, there are few important rule of thumbs: (1) the dataset should embrace the full experimental variance; (2) the dataset parsing should be non-trivial; (3) the dataset should be of high-quality with respect to the envisaged application. To better understand these points, let us consider a practical example, tomographic reconstruction of SRXTM data of mouse lung tissue. In this case, rule 1 requires the training dataset to contain image patches from lung tissue of different mice, to account for the biological variance, and from different beamtimes, to account for average changes in the experimental setup (X-ray beam, optics or camera). Rule 2 suggests to select image patches with plenty of structures or at the interface between different image regions. According to rule 3, the high-quality part of the training dataset should not be affected by zingers, ring artifacts or

any other sort of spurious signal that the neural network should not learn. How big should the training dataset be to obtain a reliable deep learning system for SRXTM image reconstruction? Since such study has not been conducted yet, we can gain some insights by analyzing the results published in [314]. This work shows that a rather standard CNN, trained on thousands of patches extracted from few tens of reconstructed slides, manage to perform complex tasks like finding the center of rotation or removing typical artifacts. The preliminary results on simulated and experimental datasets of [313] for image reconstruction seem to point in the same direction. A synchrotron imaging beamline offers the perfect framework to test the idea of combining tomographic reconstruction with deep learning: SRXTM experiments produce terabytes of imaging data for different samples and are often repeated in the course of different beamtimes. As a matter of fact, the lack of training data will not represent a problem for SRXTM applications.

Deep learning is surely better suited than compressed sensing techniques to boost tomographic image reconstruction, when dealing with undersampled datasets. Undersampling yields artifacts that are correlated to the object and cannot be completely disentangled by enforcing sparsity in a certain space during iterative reconstruction. A deep learning system is instead more likely to learn to "recognize" and to "filter out" a streak artifact than an iterative method using a compressed sensing based regularization. For this reason, I do not see any potential for studies exploring how neural networks can mimic finely tuned iterative reconstruction methods. By default, no improvement in terms of reconstruction accuracy compared to iterative regularized methods would come from such approach. Only the overall computational efficiency would increase, as a trained network in feed-forward mode would reconstruct faster than any iterative method developed so far.

Chapter 9

Additional contributions

9.1 Contrast-transfer-function phase retrieval based on compressed sensing

The following section represents a reprinted manuscript published as: P. Villanueva-Perez, F. Arcadu, P. Cloetens, and M. Stampanoni, *"Contrast-transfer-function phase retrieval based on compressed sensing"*, submitted to Applied Physics Letters, 2016.
P. Villanueva and F. Arcadu contributed equally to the manuscript.

9.1.1 Abstract

We report on a new contrast-transfer-function (CTF) phase-retrieval method based on the alternating direction method of multipliers (ADMM-CTF), which allows to exploit any compressed sensing regularization scheme reflecting the sparsity of the investigated object. The proposed iterative algorithm retrieves accurate phase maps from highly noisy single-distance projection microscopy data and is characterized by a stable convergence, not bounded to the prior knowledge of the object support or to the initialization strategy. Experiments on simulated and real datasets acquired at ID16A-NI beamline at the European Synchrotron Radiation Facility (ESRF) show that ADMM-CTF yields reconstructions with a substantial lower amount of artifacts and enhanced signal-to-noise ratio compared to the standard analytical inversion.

9.1.2 Introduction

A current challenge in biology is performing high-resolution non-destructive imaging of radiosensitive specimens like cells and bacteria. Hard X-ray synchrotron-based phase-contrast techniques are suited to address this problem [315]. Phase-contrast methods [78, 316] exploit elastic interactions that do not deposit dose and provide enough sensitivity to discriminate different soft tissues. In the hard X-ray regime, the elastic component of the refractive index for soft tissues results three orders of magnitude bigger than the inelastic component which originates the standard X-ray absorption contrast. Resolutions up to tens of nanometers are achieved thanks to the high coherence and brilliance of third generation synchrotron sources [317–319].

Projection microscopy [320], abbreviated as PM and also known as in-line holography, offers a better signal-to-dose ratio compared to other free-propagation phase-contrast techniques, e.g, coherent diffraction imaging [321]. PM records magnified holographic intensity patterns and retrieves the complex phase information by means of a phase retrieval algorithm. It requires a small X-ray source, which can also be secondary [322]. The sample, positioned at a distance z_1 from the source, is irradiated with a coherent divergent beam and the intensity of the Fresnel diffraction pattern (or hologram) is measured at a distance z_2 from the sample, as depicted in Fig. 9.1. The distances z_1 and z_2 are chosen such that the defocusing distance $z = z_1 z_2/(z_1 + z_2)$ is sufficient to provide a measurable near-field diffraction pattern [72].

The phase map of the sample can be retrieved from the hologram by means of non-linear iterative algorithms [323, 324], that require knowledge of the object support and have no guaranteed convergence. Analytical inversion algorithms stem either from the linearization of the propagation distance, as the transport-of-intensity equation (TIE) approaches [325], or from the linearization of the sample transmissivity, leading to the contrast-transfer function (CTF) method [73]. The so-called mixed approach [326], finally, applies for slow varying samples and beyond the small distance limit. From a physical point of view, the interaction between the beam and small soft-tissue samples is well approximated through linearization, making CTF the optimal approach to retrieve the phase map of the sample. Due to the missing frequencies characterizing the transfer function, CTF cannot retrieve a quantitative solution from a single hologram, unless the case of a pure-phase or one-material object [73] is considered. To overcome this limitation and ensure an overall non-zero transfer function, CTF requires the acquisition of holograms at different distances z_1 [54, 327], increasing the dose delivered to the sample. CTF reconstructions also suffer from low frequency artifacts due to the fact that frequencies below $f_{\text{min}} \approx 1/\sqrt{2\lambda z_{\text{max}}}$ [327], where λ is the wavelength and z_{max} the maximum defocusing distance, cannot be recovered. Furthermore, the sensitivity and resolution achievable by PM for radio-sensitive specimens is dictated by the radiation damage [328]. To prevent the alteration or destruction of the sample due to the radiation damage, the exposure time of PM projections has to be limited, resulting in the acquisition of highly noisy datasets. The signal retrieved from such datasets through analytical inversion is overall affected by noise and artifacts, that can impair the

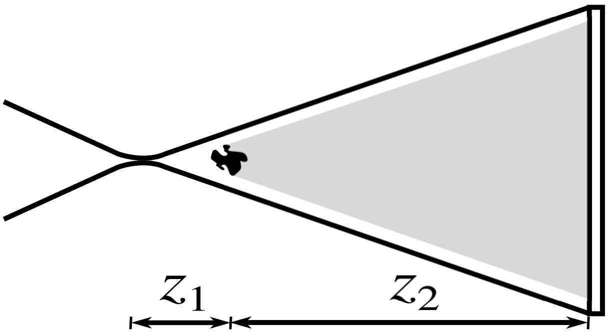


Figure 9.1: Projection microscopy setup. The sample is illuminated at a distance z_1 from the source point and a Fresnel holographic pattern is recorded at a distance z_2 from the sample.

identification of the soft tissues under study.

Here we report on a new iterative algorithm based on the alternating direction method of multipliers [113] and abbreviated ADMM-CTF, that can solve the CTF problem with a single noisy hologram by incorporating any kind of compressed sensing regularization scheme matching the specimen under study. The ADMM-CTF recovers the missing frequencies of the transfer function, providing an accurate reconstruction of the phase map where the noise pattern and distortions are substantially decreased. The algorithm has stable convergence not bounded to the usage of the object support (e.g. setting all labeled pixels of the background to zero at each iteration) and does not require warm initialization, i.e., initializing the solution with an approximated reconstruction. Experiments performed on simulated and real datasets acquired at ID16A-NI at ESRF show that ADMM-CTF working with total variation regularization (TV) [128, 329] improves the quality of the retrieved piecewise constant phase maps at visual inspection and in terms of signal-to-noise ratio and structural fidelity. The proposed algorithm is an effective tool to enable low-dose fast acquisitions of PM with short exposure time, as it can operate in single distance and with datasets affected by a considerable amount of Poisson noise. Nevertheless, this approach can be trivially extended to retrieve a phase map combining different defocusing distances.

Similar compressed sensing approaches applied to phase retrieval have also been studied for Gabor holography [330], single-distance PM tomographic reconstruction [331], TIE with pure-phase objects [332], and Fourier transform holography and CDI [333]. Recently, a regularized Newton method that can address any specific kind of phase retrieval problem has also been introduced [334]. The works of Brady [330], Kostenko [331] and Maretzke [334] deal either with the same forward model or type of regularization as the ADMM-CTF, but use a different iterative scheme. The work of Bostan [332] is based on the ADMM, but it has only been applied to TIE for data acquired in the optical domain.

9.1.3 Forward model for CTF

The forward model for CTF is here derived and discussed. Using the projection approximation [78], the sample transmissivity in the transverse plane can be expressed as

$$t(\mathbf{x}) = e^{j(\phi(\mathbf{x})+jB(\mathbf{x}))} , \tag{9.1}$$

where

$$\phi(\mathbf{x}) = k\int \delta(\mathbf{x};z)dz , \tag{9.2}$$

$$B(\mathbf{x}) = k\int \beta(\mathbf{x};z)dz , \tag{9.3}$$

$k = 2\pi/\lambda$ is the wave number, and δ and β are the real and imaginary part of the complex refractive index, respectively. The wavefield after the sample is given by

$$\psi(\mathbf{x}) = \psi_0(\mathbf{x})t(\mathbf{x}) , \tag{9.4}$$

where $\psi_0(\mathbf{x})$ is the wave field illuminating the sample. To simplify the model, we assume a monochromatic plane wave illumination, as the magnified geometry in Fig. 9.1 is equivalent to a plane wave geometry [72], with propagation distance equal to the defocusing distance z and with demagnified coordinates. The intensity measured in the near-field at a distance z is given by the squared module of the convolution between the incoming wave field and the Fresnel propagator

$$I(\mathbf{x}; z) = |\psi(\mathbf{x}) * h(\mathbf{x}; z)|^2 \ , \tag{9.5}$$

where the Fresnel propagator is given by

$$h(\mathbf{x}; z) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k|x|^2}{2z}} \ . \tag{9.6}$$

In general, the inversion is not a linear problem. As mentioned before, the interaction between hard X-rays and soft tissues is quite small. Therefore, the CTF equation [73] is a good approximation to retrieve the complex transmissivity from the measured intensity at the detector:

$$\widehat{I}(\mathbf{f}; z) = \delta(|\mathbf{f}|) + 2\cos(\pi\lambda z|\mathbf{f}|^2)\widehat{B}(\mathbf{f}) - 2\sin(\pi\lambda z|\mathbf{f}|^2)\widehat{\phi}(\mathbf{f}) \ , \tag{9.7}$$

where the widehat operator indicates the Fourier transform, and $\mathbf{f}$ is the Fourier dual variable of $\mathbf{x}$. For imaging of soft tissues, the absorption component can be neglected, i.e., $\delta \gg \beta$, and the discretized forward model can be written as:

$$\mathbf{b} = \mathbf{A}\boldsymbol{\phi} \ , \tag{9.8}$$

where

$$\mathbf{A} = \mathcal{F}^{-1} \left(-2\sin(\pi\lambda z|\mathbf{f}|^2)\right) \mathcal{F} \tag{9.9}$$

and $\mathbf{b}$ is the flat corrected intensity minus one, and $\mathcal{F}$ is the Fourier transform operator. This forward projector can be easily extended to the single-material scenario, considering that the ratio β/δ is a fundamental property of the sample:

$$\mathbf{A} = \mathcal{F}^{-1} \left(2\frac{\beta}{\delta}\cos(\pi\lambda z|\mathbf{f}|^2) - 2\sin(\pi\lambda z|\mathbf{f}|^2)\right) \mathcal{F} \ . \tag{9.10}$$

Incoherent effects can also be included in the forward model by adding an optical transfer function (OTF):

$$\mathbf{A} = \mathcal{F}^{-1} \left(2\frac{\beta}{\delta}\cos(\pi\lambda z|\mathbf{f}|^2) - 2\sin(\pi\lambda z|\mathbf{f}|^2)\right) \mathrm{OTF}(\mathbf{f})\mathcal{F} \ . \tag{9.11}$$

9.1.4 ADMM for CTF

The compressed-sensing reconstruction of a discrete physical map (x) minimizes a cost function consisting of a quadratic fidelity term, that encodes the forward model, and a regularization term R, that enforces the sparsity featured by the investigated specimen:

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \eta R(\mathbf{x}) \right] \ , \tag{9.12}$$

where η is the regularization strength. If the specimen is piecewise constant, i.e., sparse in the gradient domain, $R(\mathbf{x})$ can be chosen to be the total variation [128] and (9.12) becomes a LASSO problem [329]: $R(\mathbf{x}) = TV(\mathbf{x}) := \sum_{j=1,2} \|(\nabla\mathbf{x})_j\|_1$, where ∇ represents the gradient operator. To use ADMM [113], the auxiliary variable $\mathbf{u}_j = (\nabla\mathbf{x})_j$ $j = 1,2$ is introduced and the LASSO problem is mapped into the minimization of the following Lagrangian:

$$\mathcal{L}_\tau(\mathbf{x},\mathbf{u},\mathbf{m}) = \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \eta\sum_j \|\mathbf{u}_j\|_1 + \mathbf{m}^T(\nabla\mathbf{x}-\mathbf{u}) + \frac{\tau}{2}\|\nabla\mathbf{x}-\mathbf{u}\|_2^2 \quad , \tag{9.13}$$

where $\mathbf{m}$ and τ are Lagrangian multipliers. ADMM iteratively minimizes $\mathcal{L}_\tau$ by sequentially solving smaller problems; each iteration $k+1$ involves two sub-optimizations with respect to $\mathbf{x}$ and to $\mathbf{u}$, followed by the update of $\mathbf{m}$:

1. $\mathbf{x}^{(k+1)} \longleftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}_\tau\left(\mathbf{x},\mathbf{u}^{(k)},\mathbf{m}^{(k)}\right)$
2. $\mathbf{u}^{(k+1)} \longleftarrow \underset{\mathbf{u}}{\operatorname{argmin}}\, \mathcal{L}_\tau\left(\mathbf{x}^{(k+1)},\mathbf{u},\mathbf{m}^{(k)}\right)$
3. $\mathbf{m}^{(k+1)} \longleftarrow \mathbf{m}^{(k)} + \tau\left(\nabla\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)}\right)$

To obtain a close form for Step (1), one sets $\nabla_\mathbf{x}\mathcal{L}_\tau = 0$, which leads to the equation:

$$(\mathbf{A}^\dagger\mathbf{A} + \tau\mathbb{I})\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} + \nabla^\dagger\left(\tau\mathbf{u}^{(k)} - \mathbf{m}^{(k)}\right) \quad , \tag{9.14}$$

where $\mathbb{I}$ is the identity and $\dagger$ indicates the adjoint operator. Equation 9.14 can be efficiently solved for $\mathbf{x}$ through FFT, as both $\mathbf{A}$ and ∇ are block Toeplitz matrices and periodic boundary conditions are assumed [335]. Setting $\nabla_\mathbf{u}\mathcal{L}_\tau = 0$ provides the close form for step (2), which corresponds to a shrinkage operation:

$$\mathbf{u}^{(k+1)} = \max\left\{\left|\nabla\mathbf{x}^{(k+1)} + \frac{\mathbf{m}^{(k)}}{\tau}\right| - \frac{\eta}{\tau}, 0\right\} \operatorname{sgn}\left(\nabla\mathbf{x}^{(k+1)} + \frac{\mathbf{m}^{(k)}}{\tau}\right) \quad . \tag{9.15}$$

9.1.5 Experimental results

To test the capabilities of this framework, we first performed different validations including and not including incoherent effects via a known OTF. A modified Shepp-Logan phantom including a maximum phase shift of 0.4 rad and a ratio $\beta/\delta = 0.5$ was used in transmission, as illustrated in Fig. 9.2(a). A plane wave illumination with a propagation distance 137 μm, pixel size 10 nm, and energy 17.01 keV was simulated. The simulated intensity without any incoherent effect (no OTF) and including Poisson noise for an imaging fluence of $\Phi = 1\cdot 10^{12}$ ph/μm^2 is depicted in Fig. 9.2(b). The phase map is retrieved by analytical inversion and with the proposed ADMM-CTF algorithm. The analytical reconstruction is obtained by inverting $\mathbf{A}$ from Eq. (9.10) with a Tikhonov regularization term $\epsilon = 0.1$ and by imposing the simulated ratio β/δ. The obtained result is shown in Fig. 9.2(c). ADMM-CTF is applied without any incoherent effect and without Tikhonov regularization. For this reconstruction, we set $\tau = 5\cdot 10^{-5}$ and $\eta = 0.02\cdot\tau$. The iterative procedure is stopped when the average difference between two consecutive reconstructed intensities is smaller than the average expected uncertainty

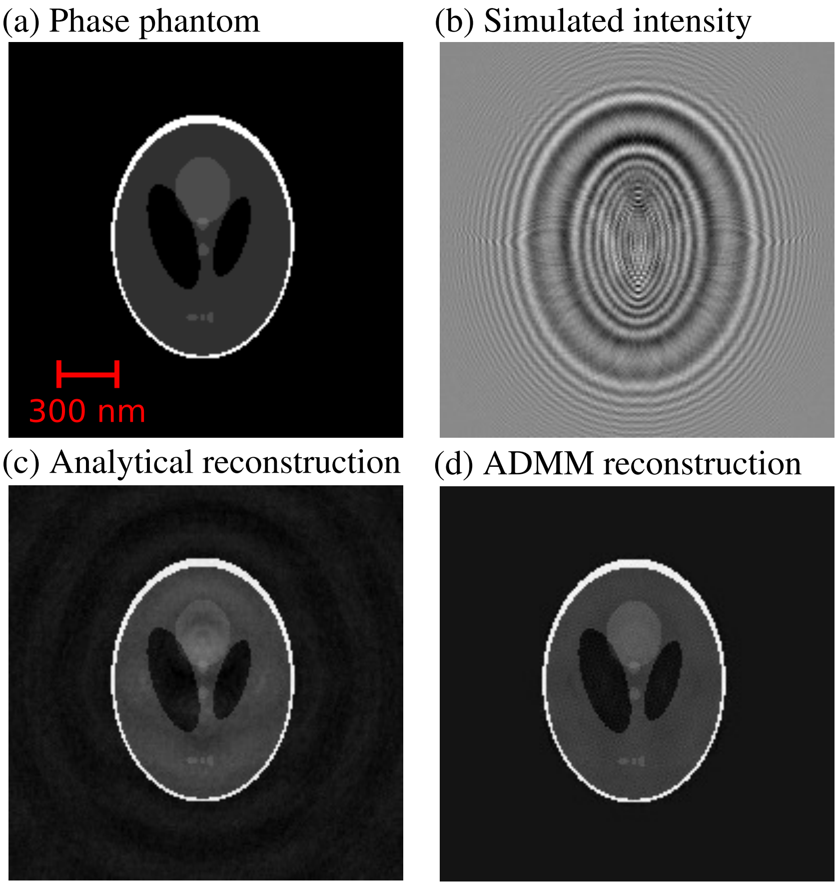


Figure 9.2: (a) A modified Shepp-Logan phantom. (b) Simulated PM intensity with Poisson noise. (c) Best analytical reconstruction with Tikhonov regularization. (d) ADMM-CTF reconstruction with TV regularization.

for the flat corrected intensity. The flat corrected intensity is computed by dividing the intensity with the sample (I_s) by the intensity without sample (I_f):

$$I(\mathbf{x};z) = \frac{I_s(\mathbf{x};z)}{I_f(\mathbf{x};z)} . \tag{9.16}$$

The average intensity measured with the sample is given by $\langle I_s \rangle \approx (1 - 2\langle B \rangle)\langle I_f \rangle$, neglecting second order absorption effects. The expected uncertainty of the flat corrected intensity, used as stopping threshold for the ADMM-CTF, is given by

$$\langle \sigma_I \rangle^2 \approx \frac{2\,(1 - 3\langle B \rangle)}{\langle I_f \rangle} . \tag{9.17}$$

The code [336] was implemented in MATLAB on a Intel(R) Core(TM) i7-3540M CPU at 3.00GHz. Convergence was achieved after seven iterations, amounting to a total time of 3.02 s.
The reconstruction with ADMM-CTF depicted in Fig. 9.2(d) presents a clear improvement, at visual

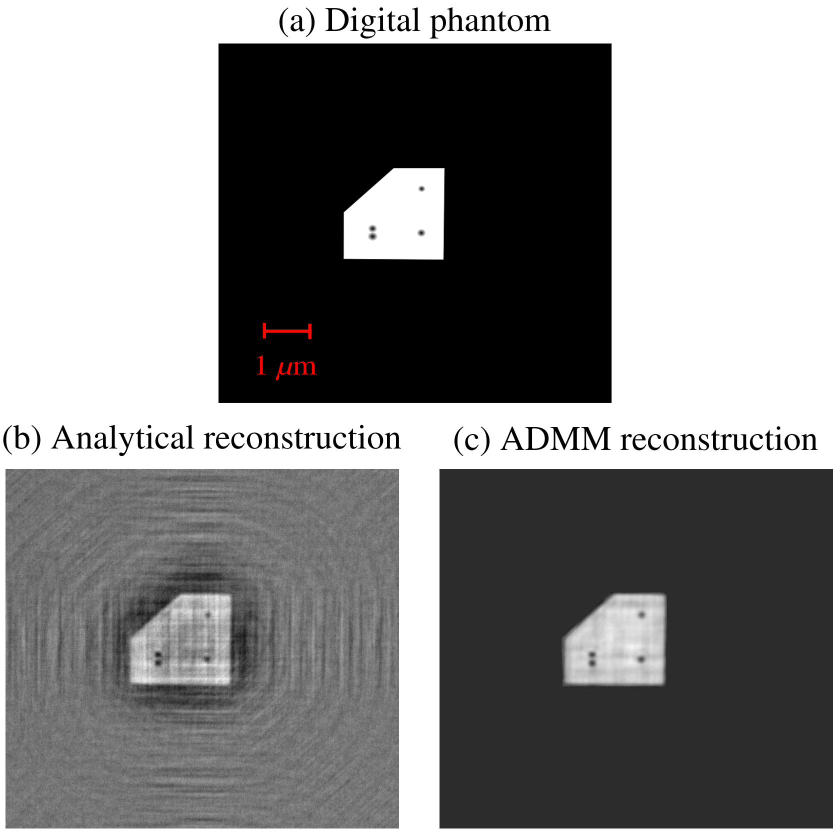


Figure 9.3: (a) Digital version of the sample. (b) Analytical reconstruction. (c) ADMM-CTF reconstruction with TV regularization.

inspection, compared to the analytical result. The accuracy of the reconstructions has also been assessed in terms of peak-signal-to-noise-ratio (PSNR) [144] and mean structural similarity index (MSSIM) [145], computed against the phantom. For the analytical reconstruction, PSNR=20.52 and MSSIM=0.13; for the ADMM-CTF reconstruction, PSNR=38.59 and MSSIM=0.17. Therefore, ADMM-CTF combined with TV regularization succeeds to retrieve missing frequencies and to reduce noise and distortions.

Finally, the proposed algorithm was tested on real single-distance PM datasets acquired at the ID16A-NI beamline of ESRF. The new ID16A-NI Nano-imaging beamline provides nano-focused X-ray beams, exploiting high coherence at 185 m from the insertion device. At ID16A-NI, PM was performed with an effective pixel size of 10 nm and a defocusing distance $z = 4.01$ mm. The samples were illuminated by a 17 keV beam with a focal spot characterized by a half width half maximum of 37×24.5 nm^2 and a divergence of 5.6 mrad. The OTF corresponds to a rectangular aperture with a size of 0.3×0.56 mm^2. The samples are silicon etched wafers with small pillars mimicking biological features that range from 50-100 nm and have low phase contrast between 5-30 mrad. The digital version of one of these samples is shown in Fig. 9.3(a). The analytical result, in Fig. 9.3(b), was obtained with a regularization parameter $\epsilon = 0.5$ and by setting the ratio $\beta/\delta = 5.4 \cdot 10^{-3}$ for silicon at 17.01 keV. The reconstruction computed by the ADMM-CTF approach, depicted in Fig. 9.3(c), was obtained with $\tau = 5$ and $\eta = 2 \cdot 10^{-6} \cdot \tau$ and by applying physical constraints, i.e, setting to zero all negative pixels at each iteration. The convergence was obtained after 29 iterations, corresponding to a run time of 25 s. The wiggling artifacts in the background and the dark grey halo in the area within the object pillars are substantially suppressed in the ADMM-CTF reconstruction compared to the analytical one. At visual inspection, Fig. 9.3(c) is much more similar to the digital phantom than the result in Fig. 9.3(b). Furthermore, for the analytical CTF retrieval PSNR=14.59 and MSSIM=0.10, whereas for the ADMM-CTF retrieval PSNR=24.46 and MSSIM=0.16.

9.1.6 Conclusion

In summary, the proposed ADMM-CTF algorithm allows to accurately retrieve phase maps from highly noisy single-distance PM data by exploiting compressed sensing regularization schemes suitable to the specimen under study. The ADMM-CTF features a stable and fast convergence. Furthermore, it does not require an initialization strategy nor the knowledge of the specific object support. The amount of distortions affecting the analytical CTF inversion are greatly reduced by the ADMM-CTF, leading to more accurate reconstructions. All considered, the ADMM-CTF has the potential to become a very useful tool for projection microscopy experiments because of the following reasons: (i) It provides higher quality reconstructions for low-dose acquisition schemes (single-distance and short exposure time), allowing subsequent quantitative and morphological analysis to be more reliable. (ii) Its convergence is independent from the sample.

9.1.7 Acknowledgements

We are grateful to Rajmund Mokso, Bill Pedrini, and Julio da Silva for their help during the experimental acquisition and fruitful discussions. Part of this work has been supported by the ERC grant ERC-2012-StG 310005-PhaseX.

Bibliography

[1] G. Lovric, S. F. Barré, J. C. Schittny, M. Roth-Kleiner, M. Stampanoni, and R. Mokso, "Dose optimization approach to fast x-ray microtomography of the lung alveoli," *J Appl Cryst*, vol. 46, pp. 856–860, jun 2013.

[2] S. M. Walker, D. A. Schwyn, R. Mokso, M. Wicklein, T. Müller, M. Doube, M. Stampanoni, H. G. Krapp, and G. K. Taylor, "In vivo time-resolved microtomography reveals the mechanics of the blowfly flight motor," *PLoS Biology*, vol. 12, p. e1001823, mar 2014.

[3] B. Hesse, P. Varga, M. Langer, A. Pacureanu, S. Schrof, N. Männicke, H. Suhonen, P. Maurer, P. Cloetens, F. Peyrin, and K. Raum, "Canalicular network morphology is the major determinant of the spatial distribution of mass density in human bone tissue: Evidence by means of synchrotron radiation phase-contrast nano-CT," *J Bone Miner Res*, vol. 30, pp. 346–356, jan 2015.

[4] L. K. Aagesen, A. E. Johnson, J. L. Fife, P. W. Voorhees, M. J. Miksis, S. O. Poulsen, E. M. Lauridsen, F. Marone, and M. Stampanoni, "Universality and self-similarity in pinch-off of rods by bulk diffusion," *Nature Physics*, vol. 6, pp. 796–800, aug 2010.

[5] G. Campi, A. Bianconi, N. Poccia, G. Bianconi, L. Barba, G. Arrighetti, D. Innocenti, J. Karpinski, N. D. Zhigadlo, S. M. Kazakov, M. Burghammer, M. v. Zimmermann, M. Sprung, and A. Ricci, "Inhomogeneity of charge-density-wave order and quenched disorder in a high-tc superconductor," *Nature*, vol. 525, pp. 359–362, sep 2015.

[6] S. Berg, H. Ott, S. A. Klapp, A. Schwing, R. Neiteler, N. Brussee, A. Makurat, L. Leu, F. Enzmann, J.-O. Schwarz, M. Kersten, S. Irvine, and M. Stampanoni, "Real-time 3d imaging of haines jumps in porous media flow," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 3755–3759, feb 2013.

[7] J. Ganne, V. D. Andrade, R. F. Weinberg, O. Vidal, B. Dubacq, N. Kagambega, S. Naba, L. Baratoux, M. Jessell, and J. Allibon, "Modern-style plate subduction preserved in the palaeoproterozoic west african craton," *Nature Geoscience*, vol. 5, pp. 60–65, nov 2011.

[8] M. Ebner, F. Marone, M. Stampanoni, and V. Wood, “Visualization and quantification of electrochemical and mechanical degradation in li ion batteries,” *Science*, vol. 342, pp. 716–720, oct 2013.

[9] J. Lu, Y. J. Lee, X. Luo, K. C. Lau, M. Asadi, H.-H. Wang, S. Brombosz, J. Wen, D. Zhai, Z. Chen, D. J. Miller, Y. S. Jeong, J.-B. Park, Z. Z. Fang, B. Kumar, A. Salehi-Khojin, Y.-K. Sun, L. A. Curtiss, and K. Amine, “A lithium–oxygen battery based on lithium superoxide,” *Nature*, vol. 529, pp. 377–382, jan 2016.

[10] Y. Chaimanee, D. Jolly, M. Benammi, P. Tafforeau, D. Duzer, I. Moussa, and J.-J. Jaeger, “A middle miocene hominoid from thailand and orangutan origins,” *Nature*, vol. 422, pp. 61–65, mar 2003.

[11] E. M. Friis, P. R. Crane, K. R. Pedersen, M. Stampanoni, and F. Marone, “Exceptional preservation of tiny embryos documents seed dormancy in early angiosperms,” *Nature*, vol. 528, pp. 551–554, dec 2015.

[12] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Society for Industrial & Applied Mathematics (SIAM), jan 2001.

[13] G. T. Herman, A. Lent, and S. W. Rowland, “ART: Mathematics and applications,” *Journal of Theoretical Biology*, vol. 42, pp. 1–32, nov 1973.

[14] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections,” *Journal of Theoretical Biology*, vol. 36, pp. 105–117, jul 1972.

[15] A. H. Andersen and A. C. Kak, “Simultaneous algebraic reconstruction technique (SART): A superior implementation of the art algorithm,” *Ultrasonic Imaging*, vol. 6, pp. 81–94, jan 1984.

[16] S. Kaczmarz, “Angenäherte Auflösung von Systemen linearer Gleichungen,” *Bulletin International de l’Académie Polonaise des Sciences et des Lettres*, vol. 35, pp. 355–357, 1937.

[17] H. Erdogan and J. A. Fessler, “Ordered subsets algorithms for transmission tomography,” *Physics in Medicine and Biology*, vol. 44, no. 11, p. 2835, 1999.

[18] I. Elbakri and J. Fessler, “Statistical image reconstruction for polyenergetic x-ray computed tomography,” *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[19] J. Hsieh, “Adaptive statistical iterative reconstruction,” *GE Healthcare white paper*, 2008.

[20] D. Mehta, R. Thompson, T. Morton, A. Dhanantwari, and E. Shefer, “Iterative model reconstruction: simultaneously lowered computed tomography radiation dose and improved image quality,” *Med Phys Int J*, vol. 2, no. 1, pp. 147–55, 2013.

[21] B. Irwan, S. Nakanishi, and A. Blum, “Adaptive statistical iterative reconstruction,” *Toshiba Medical Systems white paper*, 2011.

[22] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, aug 2008.

[23] C. Guang-Hong, T. Jie, and L. Shuai, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical Physics*, vol. 35, no. 2, p. 660, 2008.

[24] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order methoda)," *Medical Physics*, vol. 37, pp. 5113–5125, aug 2010.

[25] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland, "Compressed sensing electron tomography," *Ultramicroscopy*, vol. 131, pp. 70 – 91, 2013.

[26] M. Haltmeier, T. Berer, S. Moon, and P. Burgholzer, "Compressed sensing and sparsity in photoacoustic tomography," *Journal of Optics*, vol. 18, p. 114004, oct 2016.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.

[28] F. Arcadu, M. Nilchian, A. Studer, M. Stampanoni, and F. Marone, "A forward regridding method with minimal oversampling for accurate and efficient iterative tomographic algorithms," *IEEE Transactions on Image Processing*, vol. 25, pp. 1207–1218, mar 2016.

[29] F. Arcadu, M. Stampanoni, and F. Marone, "Fast gridding projectors for analytical and iterative tomographic reconstruction of differential phase contrast data," *Optics Express*, vol. 24, pp. 14748–14764, June 2016.

[30] F. Arcadu, F. Marone, and M. Stampanoni, "An experimental study on the coupling projector-backprojector in iterative tomographic reconstruction," *submitted*, 2016.

[31] F. M. F. Arcadu and M. Stampanoni, "Fast iterative reconstruction of data in interior tomography without a priori knowledge," *Journal of Synchrotron Radiation*, 2016.

[32] M. S. F. Arcadu, J. Vogel and F. Marone, "Improving analytical tomographic reconstructions through consistency conditions," *Fundamenta Informaticae*, 2016.

[33] P. Villanueva-Perez, F. Arcadu, P. Cloetens, and M. Stampanoni, "Contrast-transfer-function phase retrieval based on compressed sensing," *submitted*, 2016.

[34] G. D. Rubin, "Über eine neue art von strahlen," *Aus den Sitzungsberichten der Würzburger Physik-Medizinische Gesellschaft*, pp. S137–S147, 1895.

[35] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications," *J. Appl. Phys.*, vol. 34, no. 9, p. 2722, 1963.

[36] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. description of system," *The British Journal of Radiology*, vol. 46, pp. 1016–1022, dec 1973.

[37] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances, Second Edition (SPIE Press Monograph Vol. PM188)*. SPIE–The International Society for Optical Engineering, 2009.

[38] P. Reimers and J. Goebbels *Materials Evaluation*, vol. 41, no. 3, pp. 732–737, 1983.

[39] T. Kanamori, K. Kotani, and N. Ozaki, "Application of gamma ray computed tomography to non-destructive testing," *Nuclear Engineering and Design*, vol. 94, pp. 421–426, jul 1986.

[40] L. Grodzins, "Optimum energies for x-ray transmission tomography of small samples," *Nuclear Instruments and Methods in Physics Research*, vol. 206, pp. 541–545, mar 1983.

[41] B. Flannery, H. Deckman, W. Roberge, and K. L. D'Amico, "Three-dimensional x-ray microtomography," *Science*, vol. 237, no. 4821, pp. 1439–1444, 1987.

[42] U. Bonse and F. Busch, "X-ray computed microtomography (microct) using synchrotron radiation (sr)," *Progress in Biophysics and Molecular Biology*, vol. 65, no. 1-2, pp. 133–169, 1996.

[43] P. Rüegsegger, B. Koller, and R. Müller, "A microtomographic system for the nondestructive evaluation of bone architecture," *Calcified Tissue International*, vol. 58, no. 1, pp. 24–29, 1996.

[44] M. W. Westneat, O. Betz, R. W. Blob, K. Fezzaa, W. J. Cooper, and W.-K. Lee, "Tracheal Respiration in Insects Visualized with Synchrotron X-ray Imaging," *Science*, vol. 299, no. 5606, pp. 558–560, 2003.

[45] F. Plouraboue, P. Cloetens, C. Fonta, A. Steyer, F. Lauwers, and J. Marc-Vergnes, "X-ray high-resolution vascular network imaging," *Journal of Microscopy*, vol. 215, pp. 139–148, aug 2004.

[46] P. Cloetens, R. Mache, M. Schlenker, and S. Lerbs-Mache, "Quantitative phase tomography of arabidopsis seeds reveals intercellular void network," *Proceedings of the National Academy of Sciences*, vol. 103, no. 39, pp. 14626–14630, 2006.

[47] O. Betz, U. Wegst, D. Weide, M. Heethoff, L. Helfen, W. Lee, and P. Cloetens, "Imaging applications of synchrotron x-ray phase-contrast microtomography in biological morphology and biomaterials science. i. general aspects of the technique and its advantages in the analysis of millimetre-sized arthropod structure," *Journal of Microscopy*, vol. 227, pp. 51–71, jul 2007.

[48] J. Baruchel, J.-Y. Buffiere, P. Cloetens, M. D. Michiel, E. Ferrie, W. Ludwig, E. Maire, and L. Salvo, "Advances in synchrotron radiation microtomography," *Scripta Materialia*, vol. 55, pp. 41–46, jul 2006.

[49] F. Beckmann, "HARWI-II, the new high-energy beamline for materials science at HASYLAB/DESY," in *AIP Conference Proceedings*, AIP Publishing, 2004.

[50] K. Uesugi, “Direct measurement of the resolving power of x-ray CT system in SPring-8,” in *AIP Conference Proceedings*, AIP Publishing, 2004.

[51] H.-R. Lee, B. P. Lai, W. Yun, D. C. Mancini, and Z. Cai, “X-ray microtomography as a fast three-dimensional imaging technology using a CCD camera coupled with a cdwosingle-crystal scintillator,” in *Developments in X-Ray Tomography* (U. Bonse, ed.), SPIE-Intl Soc Optical Eng, oct 1997.

[52] Y. Wang, F. D. Carlo, D. C. Mancini, I. McNulty, B. Tieman, J. Bresnahan, I. Foster, J. Insley, P. Lane, G. von Laszewski, C. Kesselman, M.-H. Su, and M. Thiebaux, “A high-throughput x-ray microtomography system at the advanced photon source,” *Rev. Sci. Instrum.*, vol. 72, no. 4, p. 2062, 2001.

[53] A. Snigirev, I. Snigireva, V. Kohn, S. Kuznetsov, and I. Schelokov, “On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation,” *Rev. Sci. Instrum.*, vol. 66, no. 12, p. 5486, 1995.

[54] P. Cloetens, W. Ludwig, J. Baruchel, J.-P. Guigay, P. Pernot-Rejmánková, M. Salomé-Pateyron, M. Schlenker, J.-Y. Buffière, E. Maire, and G. Peix, “Hard x-ray phase imaging using simple propagation of a coherent synchrotron radiation beam,” *Journal of Physics D: Applied Physics*, vol. 32, no. 10A, p. A145, 1999.

[55] J.-Y. Chen, D. J. Bottjer, E. H. Davidson, S. Q. Dornbos, X. Gao, Y.-H. Yang, C.-W. Li, G. Li, X.-Q. Wang, D.-C. Xian, H.-J. Wu, Y.-K. Hwu, and P. Tafforeau, “Phosphatized Polar Lobe-Forming Embryos from the Precambrian of Southwest China,” *Science*, vol. 312, no. 5780, pp. 1644–1646, 2006.

[56] P. C. J. Donoghue, S. Bengtson, X. ping Dong, N. J. Gostling, T. Huldtgren, J. A. Cunningham, C. Yin, Z. Yue, F. Peng, and M. Stampanoni, “Synchrotron x-ray tomographic microscopy of fossil embryos,” *Nature*, vol. 442, pp. 680–683, aug 2006.

[57] M. Stampanoni, A. Groso, A. Isenegger, G. Mikuljan, Q. Chen, D. Meister, M. Lange, R. Betemps, S. Henein, and R. Abela, “TOMCAT: A beamline for TOmographic microscopy and coherent rAdiology experimenTs,” in *AIP Conference Proceedings*, AIP Publishing, 2007.

[58] “Technical data of the tomcat beamline at the swiss light source.” `https://www.psi.ch/sls/tomcat/tomcat`. Accessed: 10-10-2016.

[59] M. Stampanoni, G. Borchert, P. Wyss, R. Abela, B. Patterson, S. Hunt, D. Vermeulen, and P. Rüegsegger, “High resolution x-ray detector for synchrotron-based microtomography,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 491, pp. 291–301, sep 2002.

[60] P. Bouger, *Essai d’optique sur la gradation de la lumière*. Claude Jombert, France, 1729.

[61] J. Lambert, *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Eberhardt Klett, Germany, 1760.

[62] Beer, "Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten," *Annalen der Physik und Chemie*, vol. 86, pp. 78–88, 1852.

[63] "Beamline optics of the swiss light source." `https://www.psi.ch/sls/tomcat/optics`. Accessed: 10-10-2016.

[64] "Photoelectric effect." `https://en.wikibooks.org/wiki/Basic_Physics_of_Digital_Radiography/The_Patient#Photoelectric_Effect`. Accessed: 10-10-2016.

[65] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*. Wiley, 2011.

[66] F. van der Veen and F. Pfeiffer, "Coherent x-ray scattering," *Journal of Physics: Condensed Matter*, vol. 16, no. 28, p. 5003, 2004.

[67] *Emerging Imaging Technologies in Medicine (Imaging in Medical Diagnosis and Therapy)*. CRC Press, 2012.

[68] S. W. Wilkins, T. E. Gureyev, D. Gao, A. Pogany, and A. W. Stevenson, "Phase-contrast imaging using polychromatic hard x-rays," *Nature*, vol. 384, pp. 335–338, nov 1996.

[69] C. David, B. Noöhammer, H. H. Solak, and E. Ziegler, "Differential x-ray phase contrast imaging using a shearing interferometer," *Appl. Phys. Lett.*, vol. 81, no. 17, p. 3287, 2002.

[70] T. J. Davis, D. Gao, T. E. Gureyev, A. W. Stevenson, and S. W. Wilkins, "Phase-contrast imaging of weakly absorbing materials using hard x-rays," *Nature*, vol. 373, pp. 595–598, feb 1995.

[71] A. Momose, T. Takeda, Y. Itai, and K. Hirano, "Phase–contrast x–ray computed tomography for observing biological soft tissues," *Nature Medicine*, vol. 2, pp. 473–475, apr 1996.

[72] J. Goodman, *Introduction to Fourier Optics*. McGraw-Hill physical and quantum electronics series, W. H. Freeman, 2005.

[73] J. P. Guigay, "Fourier transform analysis of Fresnel diffraction patterns and in-line holograms," *Optik (Stuttgart)*, vol. 49, pp. 121–125, 1977.

[74] M. R. Teague, "Deterministic phase retrieval: a green's function solution," *Journal of the Optical Society of America*, vol. 73, p. 1434, nov 1983.

[75] S. M. Rytov, Y. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics 4: Wave Propagation Through Random Media*. Springer, 2011.

[76] M. Langer, P. Cloetens, J.-P. Guigay, and F. Peyrin, "Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography," *Medical Physics*, vol. 35, no. 10, p. 4556, 2008.

[77] D. Paganin, S. C. Mayo, T. E. Gureyev, P. R. Miller, and S. W. Wilkins, "Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object," *Journal of Microscopy*, vol. 206, pp. 33–40, apr 2002.

[78] D. Paganin, *Coherent X-Ray Optics (Oxford Series on Synchrotron Radiation)*. Oxford University Press, 2006.

[79] A. Momose, S. Kawamoto, I. Koyama, Y. Hamaishi, K. Takai, and Y. Suzuki, "Demonstration of x-ray talbot interferometry," *Japanese Journal of Applied Physics*, vol. 42, no. 7B, p. L866, 2003.

[80] H. Talbot, "LXXVI. facts relating to optical science. no. IV," *Philosophical Magazine Series 3*, vol. 9, pp. 401–407, dec 1836.

[81] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 1999.

[82] A. Momose, W. Yashiro, Y. Takeda, Y. Suzuki, and T. Hattori, "Phase tomography by x-ray talbot interferometry for biological imaging," *Japanese Journal of Applied Physics*, vol. 45, no. 6R, p. 5254, 2006.

[83] L. Tlustos, M. Campbell, E. Heijne, and X. Llopart, "Signal variations in high granularity si pixel detectors," in *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No.03CH37515)*, Institute of Electrical & Electronics Engineers (IEEE), 2003.

[84] V. V. Nieuwenhove, J. D. Beenhouwer, F. D. Carlo, L. Mancini, F. Marone, and J. Sijbers, "Dynamic intensity normalization using eigen flat fields in x-ray imaging," *Opt. Express*, vol. 23, p. 27975, oct 2015.

[85] J. Radon, "On the determination of functions from their integral values along certain manifolds," *IEEE Transactions on Medical Imaging*, vol. 5, pp. 170–176, dec 1986.

[86] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Academic Press, 1996.

[87] P. A. Toft and K. V. HANSEN, "Fast radon transform for detection of seismic reflections," tech. rep., In Signal Processing VII - Theories and Applications, volume I, 1994.

[88] F. John, *Plane Waves and Spherical Means: Applied to Partial Differential Equations*. Springer, 1981.

[89] P. D. Lax and R. S. Phillips, "Scattering theory," *Bull. Amer. Math. Soc.*, vol. 70, pp. 130–142, 01 1964.

[90] H. R. Tizhoosh, "Barcode annotations for medical image retrieval: A preliminary investigation," in *ICIP*, 2015.

[91] N. Aggarwal and W. Karl, "Line detection in images through regularized hough transform," *IEEE Transactions on Image Processing*, vol. 15, pp. 582–591, mar 2006.

[92] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*. Society for Industrial & Applied Mathematics (SIAM), jan 2001.

[93] P. Toft and J. SÃ¸rensen, *The Radon Transform - Theory and Implementation*. PhD thesis, 11 1996.

[94] B. D. Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology*, vol. 49, no. 11, p. 2463, 2004.

[95] G. T. Herman, *Fundamentals of Computerized Tomography*. Springer London, 2009.

[96] G. Zeng and G. Gullberg, "A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms," in *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, Institute of Electrical & Electronics Engineers (IEEE).

[97] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Transactions on Medical Imaging*, vol. 1, pp. 192–196, nov 1982.

[98] B. D. Man and S. Basu, "Distance-driven projection and backprojection," in *2002 IEEE Nuclear Science Symposium Conference Record*, Institute of Electrical & Electronics Engineers (IEEE).

[99] R. M. Lewitt, "Multidimensional digital image representations using generalized kaiser–bessel window functions," *J. Opt. Soc. Am. A*, vol. 7, pp. 1834–1846, Oct 1990.

[100] A. K. Louis, "Nonuniqueness in inverse radon problems: The frequency distribution of the ghosts," *Mathematische Zeitschrift*, vol. 185, pp. 429–440, sep 1984.

[101] F. Kuo and J. Kaiser, *System analysis by digital computer*. Wiley, 1966.

[102] S. Horbelt, M. Liebling, and M. Unser, "Discretization of the radon transform and of its inverse by spline convolutions," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 363–376, apr 2002.

[103] M. Nilchian, C. Vonesch, P. Modregger, M. Stampanoni, and M. Unser, "Fast iterative reconstruction of differential phase contrast x-ray tomograms," *Opt. Express*, vol. 21, pp. 5511–5528, Mar 2013.

[104] J. Whittaker, *Interpolation function theory*. University Press, Cambridge, England.

[105] H. Murrell, "Computer-aided tomography," *The Mathematica Journal*, vol. 6, no. 2, pp. 60–65, 1996.

[106] M. Lyra and A. Ploussi, "Filtering in SPECT image reconstruction," *International Journal of Biomedical Imaging*, vol. 2011, pp. 1–14, 2011.

[107] S. Pedemonte, A. Bousse, K. Erlandsson, M. Modat, S. Arridge, B. F. Hutton, and S. Ourselin, "GPU accelerated rotation-based emission tomography reconstruction," in *IEEE Nuclear Science Symposuim & Medical Imaging Conference*, Institute of Electrical & Electronics Engineers (IEEE), oct 2010.

[108] W. Palenstijn, K. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," *Journal of Structural Biology*, vol. 176, pp. 250–253, nov 2011.

[109] S. Basu and Y. Bresler, "O(n/sup 2/log/sub 2/n) filtered backprojection reconstruction algorithm for tomography," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1760–1773, 2000.

[110] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, pp. 113–122, oct 1982.

[111] J. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 13, pp. 290–300, jun 1994.

[112] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, pp. 323–343, jan 2009.

[113] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *FNT in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[114] J. Wang, J. Ma, B. Han, and Q. Li, "Split bregman iterative algorithm for sparse reconstruction of electrical impedance tomography," *Signal Processing*, vol. 92, pp. 2952–2961, dec 2012.

[115] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 677–688, mar 2012.

[116] S. Y. Chun, Y. K. Dewaraja, and J. A. Fessler, "Alternating direction method of multiplier for tomography with nonlocal regularizers," *IEEE Transactions on Medical Imaging*, vol. 33, pp. 1960–1968, oct 2014.

[117] E. R. L.G. Gubin, B.T. Polyak, "The method of projections for finding the common point of convex sets," *U.S.S.R Computational Mathematics and Mathematical Physics*, vol. 7, pp. 1–24, 1967.

[118] M. Defrise, F. Noo, R. Clackdoyle, and H. Kudo, "Truncated hilbert transform and image reconstruction from limited tomographic data," *Inverse Problems*, vol. 22, no. 3, p. 1037, 2006.

[119] G. T. Herman and A. Kuba, eds., *Discrete Tomography*. Birkhäuser Boston, 1999.

[120] T. Schüle, C. Schnörr, S. Weber, and J. Hornegger, "Discrete tomography by convex–concave regularization and d.c. programming," *Discrete Applied Mathematics*, vol. 151, pp. 229–243, oct 2005.

[121] K. J. Batenburg, "A network flow algorithm for reconstructing binary images from continuous x-rays," *Journal of Mathematical Imaging and Vision*, vol. 30, pp. 231–248, dec 2007.

[122] A. Alpers, H. F. Poulsen, E. Knudsen, and G. T. Herman, "A discrete tomography algorithm for improving the quality of three-dimensional x-ray diffraction grain maps," *J Appl Cryst*, vol. 39, pp. 582–588, jul 2006.

[123] H. Y. Liao and G. T. Herman, "A coordinate ascent approach to tomographic reconstruction of label images from a few projections," *Discrete Applied Mathematics*, vol. 151, pp. 184–197, oct 2005.

[124] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Transactions on Image Processing*, vol. 20, pp. 2542–2553, sep 2011.

[125] A.-I. N. Tikhonov, *Solutions of Ill Posed Problems (Scripta series in mathematics)*. Vh Winston, 1977.

[126] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, mar 1964.

[127] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, apr 2006.

[128] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, nov 1992.

[129] L. A. Shepp and B. F. Logan, "The fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21–43, jun 1974.

[130] H. Yu and S. Winkler, "Image complexity and spatial information," in *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, Institute of Electrical & Electronics Engineers (IEEE), jul 2013.

[131] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," *Medical Physics*, vol. 25, no. 11, pp. 2139–2147, 1998.

[132] B. R. Whiting, "Signal statistics in x-ray computed tomography," in *Medical Imaging 2002*, pp. 53–60, International Society for Optics and Photonics, 2002.

[133] I. A. Elbakri and J. A. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography," in *Medical Imaging 2003*, pp. 1839–1850, International Society for Optics and Photonics, 2003.

[134] J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen, and J. Manzione, "An experimental study on the noise properties of x-ray CT sinogram data in radon space," *Physics in Medicine and Biology*, vol. 53, pp. 3327–3341, jun 2008.

[135] A. V. Bronnikov, "Theory of quantitative phase-contrast computed tomography," *Journal of the Optical Society of America A*, vol. 19, p. 472, mar 2002.

[136] S. W. Smith, *The Scientist & Engineer's Guide to Digital Signal Processing*. California Technical Pub, 1997.

[137] W. O. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," *Journal of Microscopy*, vol. 127, pp. 127–138, aug 1982.

[138] N. Banterle, K. H. Bui, E. A. Lemke, and M. Beck, "Fourier ring correlation as a resolution criterion for super-resolution microscopy," *Journal of Structural Biology*, vol. 183, pp. 363–367, sep 2013.

[139] G. Harauz and M. van Heel, "Exact filters for general geometry three dimensional reconstruction," *Optik*, vol. 78, no. 4, pp. 146–156, 1986.

[140] R. Beckmann, "Alignment of conduits for the nascent polypeptide chain in the ribosome-sec61 complex," *Science*, vol. 278, pp. 2123–2126, dec 1997.

[141] M. van Heel and M. Schatz, "Fourier shell correlation threshold criteria," *Journal of Structural Biology*, vol. 151, pp. 250–262, sep 2005.

[142] G. Cardone, K. Grünewald, and A. C. Steven, "A resolution criterion for electron tomography based on cross-validation," *Journal of Structural Biology*, vol. 151, pp. 117–129, aug 2005.

[143] J. Miao, F. Förster, and O. Levi, "Equally sloped tomography with oversampling reconstruction," *Phys. Rev. B*, vol. 72, aug 2005.

[144] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, p. 800, 2008.

[145] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, apr 2004.

[146] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, pp. 98–117, jan 2009.

[147] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, pp. 121–132, feb 2004.

[148] D. M. Rouse and S. S. Hemami, "Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM," in *Human Vision and Electronic Imaging XIII* (B. E. Rogowitz and T. N. Pappas, eds.), SPIE-Intl Soc Optical Eng, feb 2008.

[149] S. Vorren, *Subjective quality evaluation of the effect of packet loss in High-Definition Video*. PhD thesis, June 2006.

[150] R. Dosselmann and X. D. Yang, "A comprehensive assessment of the structural similarity index," *Signal, Image and Video Processing*, vol. 5, pp. 81–91, nov 2009.

[151] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, pp. 35–51, mar 1996.

[152] N. D. Cahill, "Normalized measures of mutual information with general definitions of entropy for multimodal image registration," in *Biomedical Image Registration*, pp. 258–268, Springer Science and Business Media, 2010.

[153] J. Mason, *Computed Tomography with Limited Data*. PhD thesis, 2014.

[154] H. Schomberg and J. Timmer, "The gridding method for image reconstruction by fourier transformation," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 596–607, 1995.

[155] R. N. Bracewell and A. C. Riddle, "Inversion of fan-beam scans in radio astronomy," *ApJ*, vol. 150, p. 427, nov 1967.

[156] R. N. Bracewell and A. R. Thompson, "The main beam and ring lobes of an east-west rotation-synthesis array," *ApJ*, vol. 182, p. 77, may 1973.

[157] W. N. Brouw, *Aperture Synthesis*, pp. 301–307. Dordrecht: Springer Netherlands, 1975.

[158] J. Jackson, C. Meyer, D. Nishimura, and A. Macovski, "Selection of a convolution function for fourier inversion using gridding (computerised tomography application)," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 473–478, 1991.

[159] C. H. Meyer, B. S. Hu, D. G. Nishimura, and A. Macovski, "Fast spiral coronary artery imaging," *Magnetic Resonance in Medicine*, vol. 28, pp. 202–213, dec 1992.

[160] V. Rasche, R. Proksa, R. Sinkus, P. Bornert, and H. Eggers, "Resampling of data between arbitrary grids using convolution interpolation," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 385–392, may 1999.

[161] J. G. Pipe and P. Menon, "Sampling density compensation in mri: Rationale and an iterative numerical solution," *Magnetic Resonance in Medicine*, vol. 41, no. 1, pp. 179–186, 1999.

[162] R. Lewitt, "Reconstruction algorithms: Transform methods," *Proceedings of the IEEE*, vol. 71, no. 3, pp. 390–408, 1983.

[163] J. D. O'Sullivan, "A fast sinc function gridding algorithm for fourier inversion in computer tomography," *IEEE Transactions on Medical Imaging*, vol. 4, pp. 200–207, dec 1985.

[164] S.-R. Zhao and H. Halling, "A new fourier method for fan beam reconstruction," in *1995 IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, Institute of Electrical & Electronics Engineers (IEEE).

[165] S. Schaller, T. Flohr, and P. Steffen, "An efficient fourier method for 3-d radon inversion in exact cone-beam CT reconstruction," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 244–250, apr 1998.

[166] J. Walden, "Analysis of the direct fourier method for computer tomography," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 211–222, mar 2000.

[167] D. Gottleib, B. Gustafsson, and P. Forssen, "On the direct fourier method for computer tomography," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 223–232, mar 2000.

[168] P. Beatty, D. Nishimura, and J. Pauly, "Rapid gridding reconstruction with a minimal oversampling ratio," *IEEE Transactions on Medical Imaging*, vol. 24, pp. 799–808, jun 2005.

[169] F. Marone and M. Stampanoni, "Regridding reconstruction algorithm for real-time tomographic imaging," *J Synchrotron Radiat*, vol. 19, pp. 1029–1037, sep 2012.

[170] S. Matej, J. Fessler, and I. Kazantsev, "Iterative tomographic image reconstruction using fourier-based forward and back-projectors," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 401–412, apr 2004.

[171] Y. O'Connor and J. Fessler, "Fourier-based forward and back-projectors in iterative fan-beam tomographic image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 582–589, may 2006.

[172] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra from the Point of View of Communications Engineering*. Dover Publications, 1959.

[173] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty - i," *Bell System Technical Journal*, vol. 40, pp. 43–63, jan 1961.

[174] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty - II," *Bell System Technical Journal*, vol. 40, pp. 65–84, jan 1961.

[175] V. Rokhlin and H. Xiao, "Approximate formulae for certain prolate spheroidal wave functions valid for large values of both order and band-limit," *Applied and Computational Harmonic Analysis*, vol. 22, pp. 105–123, jan 2007.

[176] G. N. Ramachandran and A. V. Lakshminarayanan, "Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of fourier transforms," *Proceedings of the National Academy of Sciences*, vol. 68, no. 9, pp. 2236–2240, 1971.

[177] B. R. Hunt, "Image reconstruction from projections: Implementation and applications (g. t. herman, editor)," *SIAM Review*, vol. 23, pp. 399–400, jul 1981.

[178] E. Frey, Z.-W. Ju, and B. Tsui, "A fast projector-backprojector pair modeling the asymmetric, spatially varying scatter response function for scatter compensation in SPECT imaging," *IEEE Transactions on Nuclear Science*, vol. 40, no. 4, pp. 1192–1197, 1993.

[179] G. Zeng, Y.-L. Hsieh, and G. Gullberg, "A rotating and warping projector/backprojector for fan-beam and cone-beam iterative algorithm," *IEEE Transactions on Nuclear Science*, vol. 41, pp. 2807–2811, dec 1994.

[180] D.-C. Yu and S.-C. Huang, "Study of reprojection methods in terms of their resolution loss and sampling errors," *IEEE Transactions on Nuclear Science*, vol. 40, no. 4, pp. 1174–1178, 1993.

[181] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, no. 2, p. 252, 1985.

[182] L. Popescu and R. Lewitt, "Ray tracing through a grid of blobs," in *IEEE Symposium Conference Record Nuclear Science 2004.*, Institute of Electrical and Electronics Engineers (IEEE).

[183] C. R. Crawford, J. G. Colsher, N. J. Pelc, and A. H. R. Lonn, “High speed reprojection and its applications,” in *Medical Imaging II* (R. H. Schneider and S. J. D. III, eds.), SPIE-Intl Soc Optical Eng, jun 1988.

[184] M. L. Brady and W. Yong, “Fast parallel discrete approximation algorithms for the radon transform,” in *Proceedings of the fourth annual ACM symposium on Parallel algorithms and architectures - SPAA '92*, Association for Computing Machinery (ACM), 1992.

[185] M. L. Brady, “A fast discrete approximation algorithm for the radon transform,” *SIAM Journal on Computing*, vol. 27, pp. 107–119, feb 1998.

[186] W. A. Gotz, *Ein Schnelle Diskrete Radon Transformation Basierend auf Rekursiv Definierten Digitalen Geraden*. PhD thesis, 1993.

[187] W. Götz and H. Druckmüller, “A fast digital radon transform: An efficient means for evaluating the hough transform,” *Pattern Recognition*, vol. 28, pp. 1985–1992, dec 1995.

[188] P. E. Danielsson, *Iterative techniques for projection and back-projection*. PhD thesis, 06 1997.

[189] A. Brandt and J. Dym, “Fast calculation of multiple line integrals,” *SIAM J. Sci. Comput.*, vol. 20, pp. 1417–1429, jan 1999.

[190] A. Boag, Y. Bresler, and E. Michielssen, “A multilevel domain decomposition algorithm for fast o(n/sup 2/logN) reprojection of tomographic images,” *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1573–1582, 2000.

[191] T. Pipatsrisawat, A. Gacic, F. Franchetti, M. Pueschel, and J. Moura, “Performance analysis of the filtered backprojection image reconstruction algorithms,” in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Institute of Electrical and Electronics Engineers (IEEE).

[192] J. Fessler and B. Sutton, “Nonuniform fast fourier transforms using min-max interpolation,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 560–574, feb 2003.

[193] H. Sedarat and D. Nishimura, “On the optimality of the gridding reconstruction algorithm,” *IEEE Transactions on Medical Imaging*, vol. 19, pp. 306–317, apr 2000.

[194] M. Frigo and S. G. Johnson, “Fftw: An adaptive software architecture for the fft,” pp. 1381–1384, IEEE, 1998.

[195] A. Dutt and V. Rokhlin, “Fast fourier transforms for nonequispaced data,” *SIAM Journal on Scientific Computing*, vol. 14, pp. 1368–1393, nov 1993.

[196] G. Steidl *Advances in Computational Mathematics*, vol. 9, no. 3/4, pp. 337–352, 1998.

[197] N. Nguyen and Q. H. Liu, “The regular fourier matrices and nonuniform fast fourier transforms,” *SIAM Journal on Scientific Computing*, vol. 21, pp. 283–293, jan 1999.

[198] D. Potts, G. Steidl, and M. Tasche, *Fast Fourier Transforms for Nonequispaced Data: A Tutorial*, pp. 247–270. Boston, MA: Birkhäuser Boston, 2001.

[199] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[200] K. Lange, "Convergence of EM image reconstruction algorithms with gibbs smoothing," *IEEE Transactions on Medical Imaging*, vol. 9, no. 4, pp. 439–446, 1990.

[201] J. A. Scales, "Tomographic inversion via the conjugate gradient method," *GEOPHYSICS*, vol. 52, pp. 179–185, feb 1987.

[202] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci.*, vol. 3, pp. 253–276, jan 2010.

[203] T. Weitkamp, A. Diaz, C. David, F. Pfeiffer, M. Stampanoni, P. Cloetens, and E. Ziegler, "X-ray phase imaging with a grating interferometer," *Opt. Express*, vol. 13, p. 6296, aug 2005.

[204] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," *Nature Physics*, vol. 2, pp. 258–261, mar 2006.

[205] F. Pfeiffer, O. Bunk, C. David, M. Bech, G. L. Duc, A. Bravin, and P. Cloetens, "High-resolution brain tumor visualization using three-dimensional x-ray phase contrast tomography," *Physics in Medicine and Biology*, vol. 52, pp. 6923–6930, nov 2007.

[206] T. Köhler, B. Brendel, and E. Roessl, "Iterative reconstruction for differential phase contrast imaging using spherically symmetric basis functions," *Med. Phys.*, vol. 38, no. 8, p. 4542, 2011.

[207] Z. Qi, J. Zambelli, N. Bevins, and G.-H. Chen, "A novel method to reduce data acquisition time in differential phase contrast: computed tomography using compressed sensing," in *Medical Imaging 2009: Physics of Medical Imaging* (E. Samei and J. Hsieh, eds.), SPIE-Intl Soc Optical Eng, feb 2009.

[208] D. Hahn, P. Thibault, A. Fehringer, M. Bech, T. Koehler, F. Pfeiffer, and P. B. NoÃ«l, "Statistical iterative reconstruction algorithm for x-ray phase-contrast CT," *Sci. Rep.*, vol. 5, p. 10452, jun 2015.

[209] H. B. Dwight, *Tables of Integrals and Other Mathematical Data*. The Macmillan Company, 1961.

[210] K. Mueller and R. Yagel, "On the use of graphics hardware to accelerate algebraic reconstruction methods," in *In Proceedings of SPIE Medical Imaging Conference 1999, number 3659-62*, 1999.

[211] N. GAC, S. Mancini, M. Desvignes, and D. Houzet, "High speed 3d tomography on CPU, GPU, and FPGA," *EURASIP Journal on Embedded Systems*, vol. 2008, no. 1, p. 930250, 2008.

[212] P. B. NoÃ«l, A. M. Walczak, J. Xu, J. J. Corso, K. R. Hoffmann, and S. Schafer, "GPU-based cone beam computed tomography," *Computer Methods and Programs in Biomedicine*, vol. 98, pp. 271–277, jun 2010.

[213] L. A. Flores, V. Vidal, P. Mayo, F. Rodenas, and G. Verdú, "CT image reconstruction based on GPUs," *Procedia Computer Science*, vol. 18, pp. 1412–1420, 2013.

[214] M. Birk, R. Dapp, N. Ruiter, and J. Becker, "GPU-based iterative transmission reconstruction in 3d ultrasound computer tomography," *Journal of Parallel and Distributed Computing*, vol. 74, pp. 1730–1743, jan 2014.

[215] F. Andersson, M. Carlsson, and V. V. Nikitin, "Fast algorithms and efficient GPU implementations for the radon transform and the back-projection operator represented as convolution operators," *SIAM J. Imaging Sci.*, vol. 9, pp. 637–664, jan 2016.

[216] T. M. Peters, "Algorithms for fast back- and re-projection in computed tomography," *IEEE Trans. Nucl. Sci.*, vol. 28, no. 4, pp. 3641–3647, 1981.

[217] W. Zhuang, S. Gopal, and T. Hebert, "Numerical evaluation of methods for computing tomographic projections," *IEEE Trans. Nucl. Sci.*, vol. 41, no. 4, pp. 1660–1665, 1994.

[218] C. H. Chapman, "Generalized radon transforms and slant stacks," *Geophysical Journal International*, vol. 66, pp. 445–453, aug 1981.

[219] Y. Long, J. A. Fessler, and J. M. Balter, "3d forward and back-projection for x-ray CT using separable footprints," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 1839–1850, nov 2010.

[220] E. Papenhausen, Z. Zheng, and K. Mueller, "Creating optimal code for GPU-accelerated CT reconstruction using ant colony optimization," *Med. Phys.*, vol. 40, no. 3, p. 031110, 2013.

[221] R. T. Rockafellar, *Convex Analysis (Princeton Landmarks in Mathematics and Physics)*. Princeton University Press, 2015.

[222] J. Eckstein and M. C. Ferris, "Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control," *INFORMS Journal on Computing*, vol. 10, pp. 218–235, may 1998.

[223] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Transactions on Automatic Control*, vol. 60, pp. 644–658, mar 2015.

[224] M. R. Hestenes and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems," *Journal of Research of the National Bureau of Standards*, vol. 49, pp. 409–436, Dec. 1952.

[225] J. Fessler and S. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Transactions on Image Processing*, vol. 8, pp. 688–699, may 1999.

[226] R. J. Santos, "Preconditioning conjugate gradient with symmetric algebraic reconstruction technique (ART) in computerized tomography," *Applied Numerical Mathematics*, vol. 47, pp. 255–263, nov 2003.

[227] Q. Yang, C. R. Vogel, and B. L. Ellerbroek, "Fourier domain preconditioned conjugate gradient algorithm for atmospheric tomography," *Appl. Opt.*, vol. 45, p. 5281, jul 2006.

[228] N. Clinthorne, T.-S. Pan, P.-C. Chiao, W. Rogers, and J. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Transactions on Medical Imaging*, vol. 12, pp. 78–83, mar 1993.

[229] J. Fessler and S. Booth, "Conjugate-gradient preconditioning methods for shift-variant pet image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 688–699, May 1999.

[230] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, Institute of Electrical & Electronics Engineers (IEEE), dec 2013.

[231] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, pp. 118–121, jul 2007.

[232] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, mar 2008.

[233] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, 2004.

[234] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[235] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, pp. 183–202, jan 2009.

[236] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[237] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 22, pp. 503–516, jan 2000.

[238] M. Lysaker and X.-C. Tai, "Iterative image restoration combining total variation minimization and a second-order functional," *Int J Comput Vision*, vol. 66, pp. 5–18, jan 2006.

[239] P. Rodriguez and B. Wohlberg, "Efficient minimization method for a generalized total variation functional," *IEEE Transactions on Image Processing*, vol. 18, pp. 322–332, feb 2009.

[240] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imaging Sci.*, vol. 3, pp. 492–526, jan 2010.

[241] K. Shirai and M. Okuda, "FFT based solution for multivariable equations using KKT system via FFT and efficient pixel-wise inverse calculation," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Institute of Electrical & Electronics Engineers (IEEE), may 2014.

[242] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[243] O. Guleryuz, "Weighted overcomplete denoising," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Institute of Electrical & Electronics Engineers (IEEE).

[244] O. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part i: theory," *IEEE Transactions on Image Processing*, vol. 15, pp. 539–554, mar 2006.

[245] O. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II: adaptive algorithms," *IEEE Transactions on Image Processing*, vol. 15, pp. 555–571, mar 2006.

[246] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, pp. 3736–3745, dec 2006.

[247] W.-H. Chen, C. Smith, and S. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Transactions on Communications*, vol. 25, pp. 1004–1009, sep 1977.

[248] T. Korah and C. Rasmussen, *PCA-Based Recognition for Efficient Inpainting*, pp. 206–215. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[249] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, Institute of Electrical & Electronics Engineers (IEEE).

[250] A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, sep 2011.

[251] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Modeling & Simulation*, vol. 6, pp. 595–630, jan 2007.

[252] A. Elmoataz, O. Lezoray, and S. Bougleux, "Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing," *IEEE Transactions on Image Processing*, vol. 17, pp. 1047–1060, jul 2008.

[253] D. Liang, H. Wang, and L. Ying, "SENSE reconstruction with nonlocal TV regularization," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Institute of Electrical & Electronics Engineers (IEEE), sep 2009.

[254] V. P. Gopi, P. Palanisamy, K. A. Wahid, and P. Babyn, "MR image reconstruction based on framelets and nonlocal total variation using split bregman method," *Int J CARS*, vol. 9, pp. 459–472, sep 2013.

[255] H. Zhang, B. Yan, L. Wang, L. Li, X. Xi, and G. Hu, "Sparse-view image reconstruction with nonlocal total variation," in *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Institute of Electrical & Electronics Engineers (IEEE), nov 2014.

[256] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[257] 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*, vol. 14, pp. 1487–1503, nov 1993.

[258] O. Scherzer, "The use of morozov's discrepancy principle for tikhonov regularization for solving nonlinear ill-posed problems," *Computing*, vol. 51, pp. 45–60, mar 1993.

[259] Y.-W. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle," *IEEE Transactions on Image Processing*, vol. 21, pp. 1770–1781, apr 2012.

[260] M. Sugiyama and H. Ogawa, "Optimal design of regularization term and regularization parameter by subspace information criterion," *Neural Networks*, vol. 15, pp. 349–361, apr 2002.

[261] Y. Fang, L. Chen, J. Wu, and B. Huang, "GPU implementation of orthogonal matching pursuit for compressive sensing," in *2011 IEEE 17th International Conference on Parallel and Distributed Systems*, Institute of Electrical and Electronics Engineers (IEEE), dec 2011.

[262] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

[263] X. Xiao, F. D. Carlo, and S. Stock, "Practical error estimation in zoom-in and truncated tomography reconstructions," *Rev. Sci. Instrum.*, vol. 78, no. 6, p. 063705, 2007.

[264] R. N. Chityala, K. R. Hoffmann, D. R. Bednarek, and S. Rudin, "Region of interest (ROI) computed tomography," in *Medical Imaging 2004: Physics of Medical Imaging* (M. J. Yaffe and M. J. Flynn, eds.), SPIE-Intl Soc Optical Eng, may 2004.

[265] G. Wang and H. Yu, "The meaning of interior tomography," *Physics in Medicine and Biology*, vol. 58, pp. R161–R186, aug 2013.

[266] R. M. Lewitt, "Processing of incomplete measurement data in computed tomography," *Med. Phys.*, vol. 6, no. 5, p. 412, 1979.

[267] M. Müller and G. R. Arce, "Truncation artifacts in tomographic reconstructions from projections," *Appl. Opt.*, vol. 35, p. 3902, jul 1996.

[268] M. M. Seger, "Rampfilter implementation on truncated projection data. application to 3d linear tomography for logs," in *Proceedings SSAB 2002*, 2002.

[269] F. Noo, R. Clackdoyle, and J. D. Pack, "A two-step hilbert transform method for 2d image reconstruction," *Physics in Medicine and Biology*, vol. 49, pp. 3903–3923, aug 2004.

[270] A. Faridani, D. Finch, E. Ritman, and K. Smith, "Local tomography ii," *SIAM Journal on Applied Mathematics*, vol. 57, pp. 1095–1127, 8 1997.

[271] Y. Ye, H. Yu, Y. Wei, and G. Wang, "A general local reconstruction approach based on a truncated hilbert transform," *International Journal of Biomedical Imaging*, vol. 2007, pp. 1–8, 2007.

[272] H. Kudo, M. Courdurier, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," *Physics in Medicine and Biology*, vol. 53, no. 9, p. 2207, 2008.

[273] M. Courdurier, F. Noo, M. Defrise, and H. Kudo, "Solving the interior problem of computed tomography using a priori knowledge," *Inverse Problems*, vol. 24, no. 6, p. 065001, 2008.

[274] K. Ogawa, M. Nakajima, and S. Yuta, "A reconstruction algorithm from truncated projections," *IEEE Transactions on Medical Imaging*, vol. 3, pp. 34–40, mar 1984.

[275] F. Marone, B. Münch, and M. Stampanoni, "Fast reconstruction algorithm dealing with tomography artifacts," in *Developments in X-Ray Tomography VII* (S. R. Stock, ed.), SPIE-Intl Soc Optical Eng, aug 2010.

[276] A. KYRIELEIS, V. TITARENKO, M. IBISON, T. CONNOLLEY, and P. WITHERS, "Region-of-interest tomography using filtered backprojection: assessing the practical limits," *Journal of Microscopy*, vol. 241, pp. 69–82, dec 2010.

[277] Q. Xu, H. Yu, X. Mou, and G. Wang, "Statistical interior tomography," in *Developments in X-Ray Tomography VII* (S. R. Stock, ed.), SPIE-Intl Soc Optical Eng, aug 2010.

[278] E. A. Rashed and H. Kudo, "Towards high-resolution synchrotron radiation imaging with statistical iterative reconstruction," *J Synchrotron Radiat*, vol. 20, pp. 116–124, nov 2012.

[279] H. Zhang, J. Huang, J. Ma, Z. Bian, Q. Feng, H. Lu, Z. Liang, and W. Chen, "Iterative reconstruction for x-ray computed tomography using prior-image induced nonlocal regularization," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 2367–2378, sep 2014.

[280] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Physics in Medicine and Biology*, vol. 54, pp. 2791–2805, apr 2009.

[281] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical Chemistry*, vol. 36, pp. 1627–1639, jul 1964.

[282] M. Bertalmio, G. Shapiro, and C. Ballester, "Image inpainting," SIGGRAPH, 2000.

[283] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits Syst.*, vol. 22, pp. 735–742, sep 1975.

[284] R. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta: International Journal of Optics*, vol. 21, pp. 709–720, sep 1974.

[285] A. Messaoudi, "Matrix extrapolation algorithms," *Linear Algebra and its Applications*, vol. 256, pp. 49–73, apr 1997.

[286] J. Koloda, J. Seiler, A. Kaup, V. Sanchez, and A. M. Peinado, "Frequency selective extrapolation with residual filtering for image error concealment," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Institute of Electrical & Electronics Engineers (IEEE), may 2014.

[287] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, pp. 1200–1212, sep 2004.

[288] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR*, pp. 355–362, 2001.

[289] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, pp. 23–34, jan 2004.

[290] D. Zhou, W. Dong, and X. Shen, "Image zooming using directional cubic convolution interpolation," *IET Image Processing*, vol. 6, pp. 627–634, aug 2012.

[291] J. Radon, "Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten," *Akad. Wiss.*, vol. 69, pp. 262–277, 1917.

[292] D. Ludwig, "The radon transform on euclidean space," *Communications on Pure and Applied Mathematics*, vol. 19, pp. 49–81, sep 2010.

[293] E. Heyn, "Helgason, s., the radon transform. progress in mathematics 5. boston-basel-stuttgart, birkhäuser verlag 1980. VII, 192 s., sFr. 15.–. ISBN 3-7643-3006-6," *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 61, no. 8, pp. 411–411, 1981.

[294] J. Prince and A. Willsky, "Constrained sinogram restoration for limited-angle tomography," *Optical Engineering*, vol. 29, pp. 535–544, 5 1990.

[295] H. Kudo and T. Saito, "Sinogram recovery with the method of convex projections for limited-data reconstruction in computed tomography," *Journal of the Optical Society of America A*, vol. 8, p. 1148, jul 1991.

[296] G. V. Gompel, M. Defrise, and D. V. Dyck, "Elliptical extrapolation of truncated 2d CT projections using helgason-ludwig consistency conditions," in *Medical Imaging 2006: Physics of Medical Imaging* (M. J. Flynn and J. Hsieh, eds.), SPIE-Intl Soc Optical Eng, mar 2006.

[297] H. Yu and G. Wang, "Data consistency based rigid motion artifact reduction in fan-beam CT," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 249–260, feb 2007.

[298] Z. Liu and S. F. Yau, "A sinogram restoration technique for the hollow projection problem in computer tomography," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, Institute of Electrical and Electronics Engineers (IEEE).

[299] A. M. Alessio, P. E. Kinahan, K. M. Champley, and J. H. Caldwell, “Attenuation-emission alignment in cardiac PET/CT based on consistency conditions,” *Med. Phys.*, vol. 37, no. 3, p. 1191, 2010.

[300] G. Herman, *Image Reconstruction from Projections: Implementation and Applications (Topics in Applied Physics)*. Springer, 1979.

[301] G.-H. Chen and S. Leng, “A new data consistency condition for fan-beam projection data,” *Med. Phys.*, vol. 32, no. 4, p. 961, 2005.

[302] R. Clackdoyle and L. Desbat, “Full data consistency conditions for cone-beam projections with sources on a plane,” *Physics in Medicine and Biology*, vol. 58, pp. 8437–8456, nov 2013.

[303] T. Bortfeld and U. Oelfke, “Fast and exact 2d image reconstruction by means of chebyshev decomposition and backprojection,” *Physics in Medicine and Biology*, vol. 44, no. 4, p. 1105, 1999.

[304] A. Poularikas, *The Transforms and Applications Handbook, Second Edition (Electrical Engineering Handbook)*. CRC Press, 2000.

[305] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, jan 1979.

[306] I. Vogiatzis Oikonomidis, G. Lovric, T. Cremona, F. Arcadu, A. Patera, J. Schittny, and M. Stampanoni, “Imaging and segmentation of entire fresh non-fixed acini at micrometer resolution,” *Submitted as proceeding for the X-ray Microscopy Conference (XRM)*, 2016.

[307] D. M. Pelt and K. J. Batenburg, “Fast tomographic reconstruction from limited data using artificial neural networks,” *IEEE Transactions on Image Processing*, vol. 22, pp. 5238–5251, dec 2013.

[308] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[309] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained handwriting recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 855–868, may 2009.

[310] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *CoRR*, vol. abs/1505.04597, 2015.

[311] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, jun 2015.

[312] T. Mikolov, S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur, "Extensions of recurrent neural network language model," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, may 2011.

[313] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *CoRR*, vol. abs/1611.03679, 2016.

[314] X. Yang, F. De Carlo, C. Phatak, and D. Gürsoy, "A convolutional neural network approach to calibrating the rotation axis for X-ray computed tomography," *Journal of Synchrotron Radiation*, vol. 24, pp. 469–475, Mar 2017.

[315] D. Shapiro, D. Shapiro, P. Thibault, P. Thibault, T. Beetz, T. Beetz, V. Elser, V. Elser, M. Howells, M. Howells, C. Jacobsen, C. Jacobsen, J. Kirz, J. Kirz, E. Lima, E. Lima, H. Miao, H. Miao, A. M. Neiman, A. M. Neiman, D. Sayre, and D. Sayre, "Biological imaging by soft x-ray diffraction microscopy," *Pnas*, vol. 102, no. 43, pp. 15343–15346, 2005.

[316] A. Bravin, P. Coan, and P. Suortti, "X-ray phase-contrast imaging: from pre-clinical applications towards clinics.," *Physics in medicine and biology*, vol. 58, no. 1, pp. R1–35, 2013.

[317] J. A. Rodriguez, R. Xu, C.-C. Chen, Z. Huang, H. Jiang, A. L. Chen, K. S. Raines, A. Pryor Jr, D. Nam, L. Wiegart, C. Song, A. Madsen, Y. Chushkin, F. Zontone, P. J. Bradley, and J. Miao, "Three-dimensional coherent X-ray diffractive imaging of whole frozen-hydrated cells," *IUCrJ*, vol. 2, pp. 575–583, Sep 2015.

[318] A. Diaz, B. Malkova, M. Holler, M. Guizar-Sicairos, E. Lima, V. Panneels, G. Pigino, A. G. Bittermann, L. Wettstein, T. Tomizaki, O. Bunk, G. Schertler, T. Ishikawa, R. Wepf, and A. Menzel, "Three-dimensional mass density mapping of cellular ultrastructure by ptychographic X-ray nanotomography," *Journal of Structural Biology*, vol. 192, no. 3, pp. 461–469, 2015.

[319] M. Bartels, M. Krenkel, J. Haber, R. N. Wilke, and T. Salditt, "X-ray holographic imaging of hydrated biological cells in solution," *Phys. Rev. Lett.*, vol. 114, p. 048103, Jan 2015.

[320] S. C. Mayo, P. R. Miller, S. W. Wilkins, T. J. Davis, D. Gao, T. E. Gureyev, D. Paganin, D. J. Parry, a. Pogany, and a. W. Stevenson, "Quantitative X-ray projection microscopy: phase-contrast and multi-spectral imaging," *Journal of Microscopyoxford*, vol. 207, no. Part 2, pp. 79–96, 2002.

[321] P. Villanueva-Perez, B. Pedrini, R. Mokso, M. Guizar-Sicairos, F. Arcadu, and M. Stampanoni, "Signal-to-noise criterion for free-propagation imaging techniques at free-electron lasers and synchrotrons," *Opt. Express*, vol. 24, pp. 3189–3201, Feb 2016.

[322] O. Hignette, P. Cloetens, G. Rostaing, P. Bernard, and C. Morawe, "Efficient sub 100 nm focusing of hard x rays," *REVIEW OF SCIENTIFIC INSTRUMENTS*, vol. 76, jun 2005.

[323] Z. Zalevsky, D. Mendlovic, and R. G. Dorsch, "Gerchberg-Saxton algorithm applied in the fractional Fourier or the Fresnel domain.," *Optics Letters*, vol. 21, no. 12, pp. 842–844, 1996.

[324] K. Giewekemeyer, S. P. Krüger, S. Kalbfleisch, M. Bartels, C. Beta, and T. Salditt, "X-ray propagation microscopy of biological cells using waveguides as a quasipoint source," *Physical Review A*, vol. 83, p. 023804, feb 2011.

[325] M. Reed Teague, "Deterministic phase retrieval: a Green's function solution," *Journal of the Optical Society of America*, vol. 73, no. 11, p. 1434, 1983.

[326] J. P. Guigay, M. Langer, R. Boistel, and P. Cloetens, "Mixed transfer function and transport of intensity approach for phase retrieval in the Fresnel region," *Optics Letters*, vol. 32, p. 1617, jun 2007.

[327] S. Zabler, P. Cloetens, J. P. Guigay, J. Baruchel, and M. Schlenker, "Optimization of phase contrast imaging using hard x rays," *Review of Scientific Instruments*, vol. 76, no. 7, p. 073705, 2005.

[328] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. M. Holton, C. J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. A. Shapiro, J. C. H. Spence, and D. Starodub, "An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy.," *Journal of electron spectroscopy and related phenomena*, vol. 170, pp. 4–12, mar 2009.

[329] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, pp. 273–282, apr 2011.

[330] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive holography," *Opt. Express*, vol. 17, p. 13040, jul 2009.

[331] A. Kostenko, K. J. Batenburg, A. King, S. E. Offerman, and L. J. van Vliet, "Total variation minimization approach in in-line x-ray phase-contrast tomography," *Opt. Express*, vol. 21, p. 12185, may 2013.

[332] E. Bostan, E. Froustey, B. Rappaz, E. Shaffer, D. Sage, and M. Unser, "Phase retrieval by using transport-of-intensity equation and differential interference contrast microscopy," in *2014 IEEE International Conference on Image Processing (ICIP)*, Institute of Electrical & Electronics Engineers (IEEE), oct 2014.

[333] K. He, M. K. Sharma, and O. Cossairt, "High dynamic range coherent imaging using compressed sensing," *Opt. Express*, vol. 23, p. 30904, nov 2015.

[334] S. Maretzke, M. Bartels, M. Krenkel, T. Salditt, and T. Hohage, "Regularized newton methods for x-ray phase contrast and general imaging problems," *Opt. Express*, vol. 24, p. 6490, mar 2016.

[335] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering (Fundamentals of Algorithms 3)*. Society for Industrial and Applied Mathematic, 2006.

[336] P. Villanueva-Perez, "ADMM-CTF code." `http://www.bitbucket.com;.....`, 2016. [Online; accessed 30-September-2016].